Digital repository "Geology of the Russian Far East" – an open access to the spatially distributed online scientific publications

V. V. Naumova[1] and A. V. Belousov[1]

The digital repository "Geology of the Russian Far East" is developed in Laboratory of Information Technologies of the Far East Geological Institute of the Far East Branch of Russian Academy of Sciences, within the framework of the development of the infrastructure of the spatially distributed different-type scientific data on geology of the Russian Far East. DSpace is the basic program environment of the digital repository. The System is adapted for geology by introduction the thematic block of geological and geographical thesauruses in it. *KEYWORDS: Digital repository; open access; distributed heterogeneous data integration; geology of the Russian Far East.*

Citation: Naumova, V. V. and A. V. Belousov (2014), Digital repository "Geology of the Russian Far East" – an open access to the spatially distributed online scientific publications, *Russ. J. Earth. Sci., 14,* ES1004, doi:10.2205/2014ES000538.

Open Access

Open access is a free access of users to the online scientific publications with the right to read, to charge, to copy, to distribute, to publish, to search, to refer to the full-text articles, to index, and so on, i.e. to use them with any legal purpose without financial, juridical, or technical obstacles.

According to the Berlin Declaration of 2003 (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/), publication of full texts of editions to the access through the Internet must answer two conditions:

1. An author (authors) and right-possessors must concede to all users a free, inalienable, and all possible access to text and license the right to copy, to use, to distribute, to transfer or to show publicly any text, as well as to do and distribute the derivative works in any digital format for any legal purpose with the corresponding attribution of the authorship and to make a limited number of printed copies for personal use.

2. A full version of a given edition and all auxiliary materials to it, including a copy of the author's permission, are deposited in a corresponding electron version (and thus published), at least in one net repository with the use of the necessary technical standards that are maintained and ensured by an institution of a higher school (university), scientific society, government agency, or another distinctly defined organization, which keeps the principles of the open access, unlimited distribution, compatibility, and long-term safety of documents.

Two main technological directions are distinguished: journals of open access and repositories of open access. Both directions are the ways of scientific intercourse. Journals of open access publish the articles been reviewed, and repositories collect the documents not necessarily been reviewed and not necessarily articles. Journals of open access and repositories are not incompatible – they supplement each other.

Primarily, the standards of bibliographic descriptions were developed for the traditional publications in public libraries: GOST 7.1 – 2003, ISO 690, ISBD, and others. To operate with this information in the Internet the following standards were developed: ISO 2709, ISO 15836-2009, NISO Standard Z39.85, etc., and on the basis of them the keeping formats of bibliographic descriptions were developed: MARC (MARC 21, UNIMARC), Dublin core (SDC, QDC), MODS, and others. The process of exploitation of the emerged library catalogues arose a question of their combined use. For these purposes the Library of the USA Congress elaborated the portal of the distributed search Z39.50 that has been developing since 1970s, and new versions were put out in 1988, 1992, 1995, and 2003. It made it possible to dispose the distributed search, which didn't depend on the final systems, data base type, keeping formats, etc. In succeeding years on its basis there were developed the SRW portals that used modern technologies SOAP, HTML, XML, and SRU – an alternative of SRW based on the URL, which have a lower entry threshold than Z39.50 [*Zhizhimov and Mazov*, 2004].

---

[1]Far East Geological Institute FEBRAS, Vladivostok, Russia

Along with the advancement of the initiative of the Open Access they began to develop the digital repositories for keeping and distribution of the digital material of any type. To provide their interoperability the portal of the OAI-PMH metadata collection was elaborated.

Many technological decisions are available for the integration of the library scientific data based on these portals:

- Distributed catalogues, locks Z39.50: Zoopark, YAZ, etc.;

- Digital repositories: Dspace, Eprints, Greenstone, etc.;

- Portals of metadata: OHS (open harvesting system), PANFMP, etc.

Among such systems available today in the world the following ones may be pointed out: Common to the points of access to the heterogeneous resources of the USGS (U.S. Geological Survey, http://www.usgs.gov/pubprod/): to maps, publications, satellite photos, aero photographs, and to the accompanying data.

NBII Metadata Clearing House (National Biological Information Infrastructure) is the initiative of the USGS (U.S. Geological Survey) on the creation of the distributed information system that contains the metadata describing biological data and information products and is based on the submultitude of the CSDGM-NBII Biological Profile.

In Europe, the European Library Project (http://www.theeuropeanlibrary.org/tel4) is being realized that must integrate all national libraries of Europe and leading European research libraries. Within the framework of the Project there has been created the united portal of metadata, which harvests the metadata according to the OAI-PMH Protocol. The creators elaborated the algorithm that with some expenditures and under certain conditions makes it possible to harvest the metadata according to the Z39.50 Protocol. This library provides a quick and easy access to the collections of 48 National Libraries of Europe and leading European Scientific Libraries. Users of the Library can find and use more than 18,644,265 digital sources and 119,246,208 bibliographic references. To facilitate a further search the references to other web-sites of the European group have been constructed.

In Russia a number of the OAI-compatible scientific repositories is rather modest. These are either an addition of the OAI-module to the own system, as it is, for example, in the context of the Scientific-Educational Social Net "Socionet" (http://www.socionet.ru) [*Parinov et al.*, 2003], or the use of the specialized program media. An example of this direction is the Electron Library of the Siberian Branch, RAS (SBRAS) (http://db3.nsc.ru:8080/jspui/) [*Zhizhimov and Mazov*, 2004; *Zhizhimov et al.*, 2011]. The library is developed in DSpace. The main divisions of the Library are reports on scientific research work, virtual and real museums, dissertations, integration projects of the SBRAS, materials on the Program "Telecommunication and Multimedia Resources of the SBRAS", Library of the Siberian Branch of RAS, personalities, subject collections, technical and normative documentation, and workers' proceedings. The program environment is modernized by Z39.50 server addition. The

constructed System gives the chance to obtain data from repositories and from catalogs of scientific libraries of the Siberian Branch of RAS also.

Development of Digital Repository "Geology of the Russian Far East"

The digital repository "Geology of the Russian Far East" is developed in laboratory of information technologies of the Far East Geological Institute of the Far East Branch of Russian Academy of Sciences (FEBRAS) within the framework of the development of the infrastructure of the spatially distributed heterogeneous scientific data on geology of the Russian Far East [*Belousov*, 2013; *Naumova et al.*, 2011] and this System is a separate block of the common infrastructure (http://www.fareastgeology.ru).

The chosen decision is based on the metadata portal. A metadata portal is a system providing a simple and intelligible access to the distributed information resources [*Schindler and Diepenbroek*, 2008].

We chose the DSpace as a basic program, which possesses the functionality sufficient for our purposes: a handy catalogue system, the availability of the server for collection of the OAI-PMH metadata, full-text search based on the search instrument Apache Lucene or Apache Solr, the differentiation of rights and maintenance of the LDAP access protocol, and the possibility to manage and keep the digital material of any type. The open code and a great community of users and elaborators all over the world should be also noted.

Portal Realization

The functional scheme of the Portal being developed is given in Figure 1.

Description of Data

The Portal integrates the scientific publications that refer only to the study of geology of the Russian Far East. We define the Far East as the territory including Amurskaya Oblast', the Jewish Autonomous Region, Kamchatsky Krai, Magadanskaya Oblast', Primorsky Krai, Sakha Republic (Yakutia), Sakhalinskaya Oblast', Khabarovsky Krai, and Chukotsky Autonomous Region. The area of the Far East of Russia is 6,169,329 km^2 that makes 36.08% of the country total area.

Harvesting of Data for the Portal

The publications we need are in the digital repositories of scientific institutes and universities; in electron libraries, in-
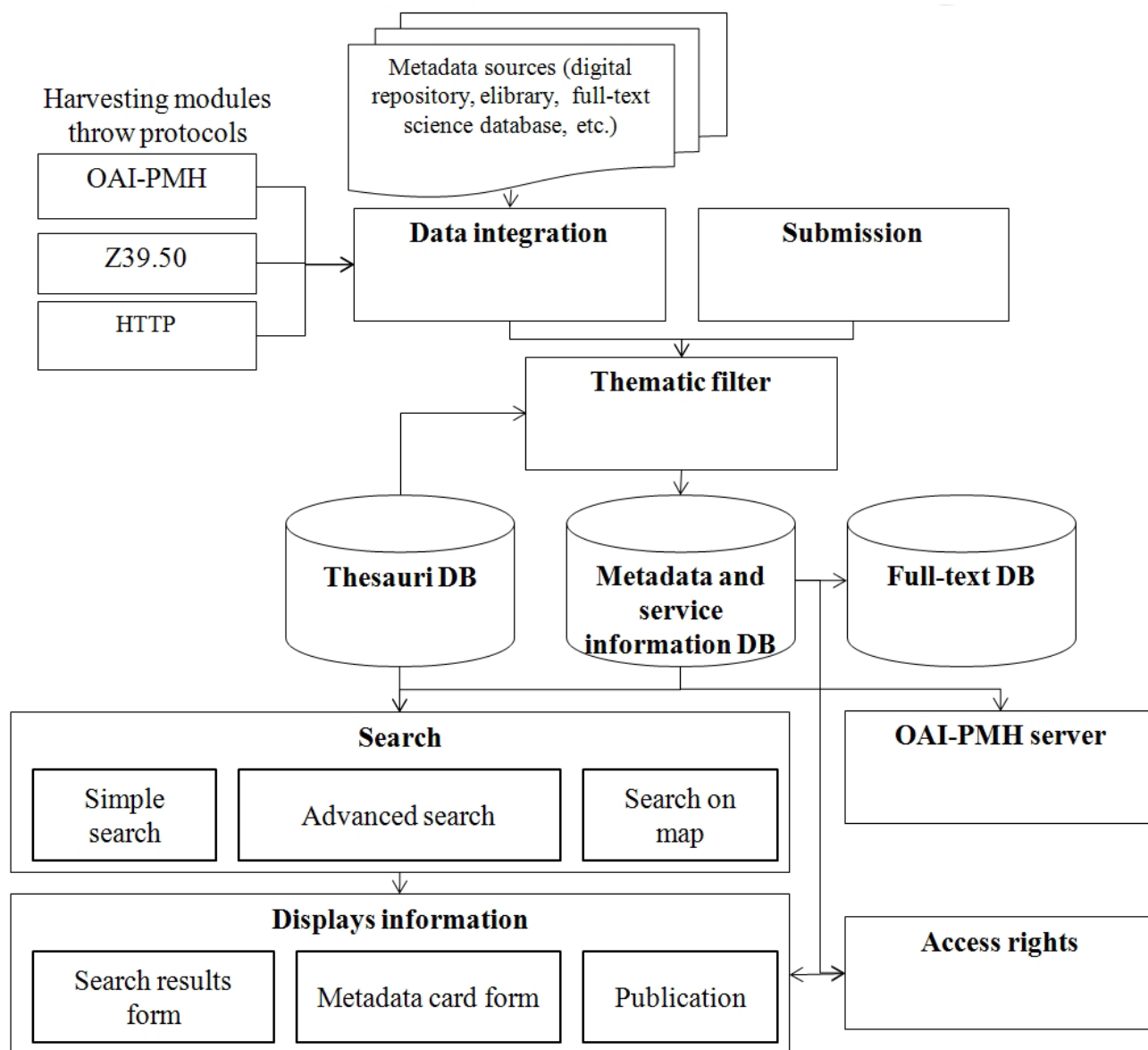
Figure 1. Functional scheme of the Digital Repository "Geology of the Russian Far East".

cluding the Scientific Electron Library (http://elibrary.ru); in the full-text scientific data-bases; in the catalogues of scientific libraries, including the catalogue of the Central Scientific Library of the Far East Branch of RAS, and on other resources.

Publications are harvested with the use of two modes: manual recording of data and automatic integration of data with the use of the collection program modules on different protocols and with the use of the subject filter.

Scientific publications are classified into following divisions:

- Dissertations and authors' abstracts;
- Materials of conferences and report theses;
- Monographs and collections of articles;
- Scientific-popular editions and articles;
- Field excursions: guide-books and virtual excursions;
- Manuscripts;
- Articles;
- Educational supplies.

Program Modules of Data Harvesting

Collections from archives and library catalogues may be accessible on the Portal trough three communication protocols: OAI-PMH (the Open Archives Initiative Protocol for

Metadata Harvesting), Z39.50 or SRU (Search/Retrieve via URL).

Choice of a communication protocol influences greatly the functionality that can be allowed by the portal to a final user. Although all three protocols allow the standard for the communication between a portal and library systems, the communication paradigm, providing the foundation, is significantly different. Whereas the OAI-PMH allows a portal to harvest all metadata records from libraries into the Central Archives, the Z39.50 and SRU were developed for the remote access and assignment, so the metadata records remain at the data provider.

Harvesting of metadata on the Z39.50/SRU protocols involves significant difficulty, because initially they were not designed for the metadata harvest, so some functionality necessary to provide the efficiency and reliability of the harvesting process was not include in the protocol project.

The works of the authors of the Library of the University of Illinois and the European Library (Guidelines for preparing a Z39.50/SRU target to enable metadata harvesting/TELplus. The European Library: http://cyberdoc.univ-lemans.fr/PUB/ CfU/Journee_UNIMARC_Lyon/TELplus-D2.3_v1.0[1].pdf) on the possibility of the metadata harvest on the Z39.50 Protocol showed that the metadata harvest is allowed if the Z39.50 server satisfies some conditions.

In contrast to the OAI-PMH the Z39.50 servers are accessible for a significant number of the systems of the library management, and they are used widely. Many libraries of Russia use Irbis as a system of the library management. Web-catalogue of the Web Irbis system maintains the Z39.50 server that allows us to harvest the bibliographic metadata from the library catalogues. If the Z39.50 server is not adjusted, we use the function of the Web Irbis export.

Thus, the Portal harvests the bibliographic descriptions from other repositories (on the OAI-PMH Protocol) and from library catalogues (on the Z39.50 Protocol) [*Kaczmarek and Naun*, 2005] or uses the function of export of the library management system. The Portal realizes the technological possibility of the metadata harvesting from the full-text scientific databases, such as the Scientific Electron Library and Science Direct.

Data Filtration

Relevancy of the harvested data is provided by filtration on the basis of the morphological search (stemming) in the metadata records of the terms of the chosen thesaurus. The metadata received are added to the database in the Dublin Core format.

Subject Thesauruses

The system is adapted to the geology of the Russian Far East by the entry of some subject thesauruses into it.

Thesauruses perform the following functions [*Kubik*, 2011]:

- give the meaning for transformation of the natural language of author, indexer, and user into the dictionary that is used for indexing and search;

- promote the uniformity of terms;

- point to the semantic relations of terms;

- give the successive and clear hierarchy in the navigation system to help the user in finding the needed object.

We have constructed the following thesauruses: "Geographic Unit of the Russian Far East", "Geologic Unit of the Russian Far East" [*Dal'nauka*, 2006a, 2006b] and "Geologic Time Scale" (Geologic Time Scale. http://geology.com/time.htm). The thesauruses represent the fixed lists of meanings of parameters. Below, some fragments of the Portal thesauruses are given in this paper.

A fragment of thesaurus "Geologic Unit of the Russian Far East":

1. Alazeiskaya island arc;

2. Alazeiskaya-Oloiskaya island arc;

3. Alchansky basin;

4. Amkhotenskaya bed;

5. Anavgaisky volcanic complex;

6. Anavetkinskoe deposit;

7. Anadyrsky metallogenic belt;

8. Anivo-Gomonsky terrane;

9. Anmandykansky massif;

10. Anmandykansky ore node.

. . .

862. Yaurinskaya suite.

A fragment of thesaurus "Geographic Unit of the Russian Far East":

1. Avachinsky volcano;

2. Aion Island;

3. Algansky Ridge;

4. Aldansky Plateau;

5. Al'sky Ridge;

6. Amgun' River;

7. Amguema River;

8. Amur River;

9. Amur-Zeiskaya depression;

10. Anadyrskaya depression.

. . .

296. Sea of Japan.

A fragment of thesaurus "Geologic Time Scale":

1. Hadean;

2. Archean;

3. Proterozoic;

4. Cambrian;

5. Lower Ordovician;

6. Middle Ordovician;

7. Upper Ordovician;

8. Llandovery;

9. Wenlock;

10. Ludlow;

...

34. Holocene.

Collections

At present, the Repository contains the following collections and communities:

1. Electron Library of the Far East Geological Institute, FEB of RAS:
   - author abstracts of dissertations;
   - articles and monographs of workers of the Institute;
   - manuscripts.

2. Electron libraries of universities of the world:
   - DLynx, the Rhodes College Archives Digital Collection;
   - The Texas A&M Digital Repository:
       - Geologic Atlas of the United States.
   - MIT's institutional repository (Massachusetts Institute of Technology);
   - Department of Earth, Atmosphere, and Planetary Sciences;
   - Earth Resources Laboratory.

3. Electron libraries of universities of Russia:
   - Archives of electron resources of the Siberian Federal University;
   - Repository of Tver' State Technical University;
   - Repository of Tver' State University;
   - Udmurt scientific-educational electron Library;
   - Electron Archives of Belgorod State University;
   - Electron Archives of the National Research South-Ural State University;
   - Electron Archives of the P. G. Demidov Yaroslavl' State University.

4. Electron catalogues of scientific libraries:
   - Electron catalogue of the Central Scientific Library of the Far East Branch of RAS.

5. Electron resources of RAS:
   - DSpace repository of the Institute of the Computer Technologies of SB, RAS;
   - Electron Archives of the M. V. Keldysh Institute of the Applied Mathematics of RAS;
   - Electron Archives of the Central Economy-Mathematical Institute of RAS.

Communities

1. Geoarchaeology;

2. Gemology;

3. Geochemistry;

4. Geodynamics and tectonics;

5. Hydrogeology;

6. Geoecology;

7. Glaciology, engineering geology;

8. Informatics, geoinformatics;

9. Lithology and sedimentology;

10. Mathematical methods;

11. Metallogeny, ore deposits;

12. Mineralogy;

13. Marine geology;

14. General and regional geology;

15. Paleoecology and paleoclimatology;

16. Petrology and volcanology;

17. Stratigraphy and paleontology;

18. Physicochemical methods of investigations.

Data Keeping

The System keeps the following objects: Metadata (Table 1), Collections, Communities, Files, Sources, Users, and official information (Figure 2). The PostgreSQL relation database is used for keeping. Metadata are kept in format Qualified Dublin Core.

A sample of metadata records is given in Figure 3.

Search of Metadata Records

The Repository uses the Jakarta Lucene search mechanism maintaining the following functions: full-text search, stop-words, word truncation, morphological search, search by phrases, and others. The widened search makes it possible to indicate what document fields participate in organization of search, which can be matched by logic operators: "AND", "OR", and "NOT" (Figure 4). The search field can be restricted by a community or a collection.

Table 1. Main parameters and description types of the parameters

Parameter name	Description types
Edition type	Text
Title	Text
Authors	Text
Edition date	Date
Publisher	Text
Journal	Text
Volume, number, pages	Text
Number of series/report	Text
Identifiers (DOI, ISBN, etc.)	Text
Language	Fixed meanings (Russian, English)
Geology Part	Fixed meanings (See Communications)
Key words or classifiers	Text
Annotation	Memo
Geological objects	Fixed meanings (see thesaurus "Geologic Unit of the Russian Far East")
Geographical objects	Fixed meanings (see thesaurus "Geographic Unit of the Russian Far East")
Stratigraphic age	Fixed meanings (see thesaurus "Geologic Time Scale")
URL	Text

System Interaction With the User

The user registered on the Portal gets a chance:

- to subscribe to the collection and receive a message to the e-mail about new receipts;
- to pass to page "My Library" where the subscription is traced and other kinds of interaction with DSpace are performed that require the authorization (for example, if a user passes documents to the collection);
- to edit a profile.

Repository "Geology of the Russian Far East" (http://repository.fareastgeology.ru) is the united point of the open access to the spatially distributed scientific publications on geology of the Russian Far East.
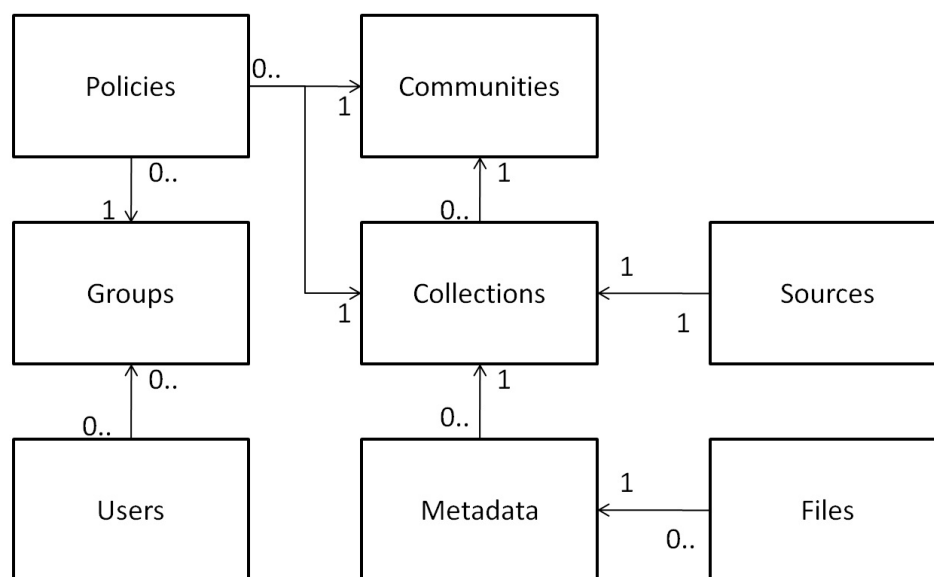

Figure 2. Generalized scheme of DB Repository.

fareastgeology.ru Resources Login

Digital repository "Geology of the Russian Far East"

Contact Us	Send Feedback	Help

Home → Electron libraries of universities of the world → MIT's institutional repository (Massachusetts Institute of Technology) → Department of Earth, Atmospheric, and Planetary Sciences → View Item

Search

[_____] [Go]

⦿ Search
○ This Collection

Advanced Search

Browse

All
  Communities & Collections
  By Issue Date
  Authors
  Titles
  Subjects
This Collection
  By Issue Date
  Authors
  Titles
  Subjects

My Account

Login
Register

Hydrocarbon biomarkers for biotic and environmental evolution through the Neoproterozoic-Cambrian transition

Kelly, Amy E. (Amy Elizabeth), 1980-

Type:
Thesis
Publisher:
Massachusetts Institute of Technology
Date: 2013-03-23
Keywords:
Earth, Atmospheric, and Planetary Sciences.
URI: http://dspace.fegi.ru:80/handle/1721

Description:

The sequence of events over the Neoproterozoic - Cambrian transition that led to the radiation of multicellular organisms has been an issue of debate for over a century. It is a critical interval in the history of life on Earth because it marks the first appearance of all extant animal phyla in the fossil record. We set out to improve understanding of environmental transitions during this key interval of Earth's history by studying chemical fossils (biomarkers) in Neoproterozoic to Cambrian aged sedimentary rocks and oils from Australia, Eastern Siberia and Oman. This thesis presents the distributions of steranes and other hydrocarbons through these various strata and the characterization of novel age and paleostratification biomarkers. Compound specific carbon isotopic data of n-alkanes and isoprenoids were also acquired and evaluated in the context of existing datasets with a focus on elucidating the processes responsible for anomalous trends. Consistent with current theory, our results indicate that there was a significant shift in the redox state the oceans and that this took place on a global scale. The biomarker and isotopic proxies we have measured help us further constrain the timing of this redox shift, and suggest a concomitant switch in the composition of marine photosynthetic communities, at termination of the Neoproterozoic Era.

by Amy Elizabeth Kelly.

Thesis (Ph. D.)--Massachusetts Institute of Technology, Dept. of Earth, Atmospheric, and Planetary Sciences, 2009.

This electronic version was submitted by the student author. The certified thesis is available in the Institute Archives and Special Collections.

Includes bibliographical references.

Files in this item

	Name: 502990922.pdf	View/Open
	Size: 1.367Mb	
	Format: PDF	

	Name: 502990922-MIT.pdf	View/Open
	Size: 8.396Kb	
	Format: PDF	

This item appears in the following Collection(s)

- Department of Earth, Atmospheric, and Planetary Sciences [1138]

Figure 3. A sample of metadata records of the Digital Repository "Geology of the Russian Far East".

Figure 4. The program interface of the widened search of the System with the use of thesaurus "Geologic Unit of the Russian Far East".

References

Belousov, A. V. (2013), Digital repository of the Far East Geological Institute, FEB of RAS: open access to scientific data on geology of the Far East of Russia, *Internet and Modern Society: Collection of abstracts. Proceedings of the XVI All-Russia United Conference "Internet and Modern Society" (IMS-2013), Sankt-Petersburg, October 9–11, 2013*, NIU ITMO, Sankt-Petersburg.

Dal'nauka (2006a), *Geodynamics, Magmatism, and Metallogeny of East Russia. Book 1*, 1–572, Vladivostok.

Dal'nauka (2006b), *Geodynamics, magmatism, and metallogeny of East Russia. Book 2*, 573–981, Vladivostok.

Kaczmarek, J., and C. C. Naun (2005), A statewide metasearch service using OAI, *Library Hi Tech, 23*, 576–586, doi:10.1108/07378830510636355.

Kubik, T. (2011), Role of thesauri in the information management in the web-based services and systems, *Lecture Notes in Computer Science, 6560*, 25–49, doi:10.1007/978-3-642-19968-4_2.

Naumova, V. V., I. N. Goryachev, K. A. Platonov (2011), Web-integration of heterogeneous scientific data and services on geology of the Far East of Russia on the basis of portal decision, *Geoinformatics*, 1, 56–62.

Parinov, S. I., V. M. Lyapunov, R. L. Pusyrev (2003), System Socionet as a platform for development of scientific information resources and online services, *Electron Libraries, 5*, 1.

Schindler, U., and M. Diepenbroek (2008), Generic XML-based framework for metadata portals, *Computers & Geosciences, 34*, 12, 1947–1955, doi:10.1016/j.cageo.2008.02.023.

Zhizhimov, O. L., and N. A. Mazov (2004), *Principle of construction of the distributed information systems on the basis of Protocol Z39.50*, 361 OIGGM, SB of RAS, Novosibirsk.

Zhizhimov, O. L., Yu. I. Molorodov, I. A. Pestunov, V. V. Smirnov, A. M. Fedotov (2011), Integration of heterogeneous data in solving the tasks of investigation of natural ecosystems, *Bulletin of Novosibirsk State University, 9*, 3, 67–74.

A. V. Belousov and V. V. Naumova, Far East Geological Institute FEBRAS, Vladivostok, Russia. (naumova@fegi.ru)