

Геоинформатика. 2022. № 4. С. 8–13.
Geoinformatika. 2022;(4):8–13.

Материалы VI Всероссийской конференции ITES-2022

Научная статья

УДК 004.550

<https://doi.org/10.47148/1609-364X-2022-4-8-13>

Разработка единого цифрового пространства геологических научных знаний

© 2022 г. — Вера Викторовна Наумова^{a)}, Виталий Сергеевич Ерёмченко^{b)}, Михаил Иванович Патук^{c)}

Государственный геологический музей им. В.И. Вернадского РАН; Россия, Москва

^{a)}naumova_new@mail.ru, ^{b)}vitaer@gmail.com, ^{c)}patuk@mail.ru

Аннотация: Обсуждается разработка Информационной системы для создания, развития и поддержки научных исследований в геологии: до стадии Единого цифрового пространства геологических научных знаний. Современная разрабатываемая стадия Проекта — это шаг вперед от традиционной территориально распределенной геологической системы к цифровой среде будущего.

Ключевые слова: *цифровое пространство научных знаний, цифровая среда, интеллектуальное рабочее пространство*

Для цитирования: Наумова В.В., Ерёмченко В.С., Патук М.И. Разработка единого цифрового пространства геологических научных знаний // Геоинформатика. — 2022. — № 4. — С. 8–13. <https://doi.org/10.47148/1609-364X-2022-4-8-13>.

VI All-Russian Conference ITES-2022

Original article

Development of a single digital space geological scientific knowledge

© 2022 — Vera V. Naumova^{a)}, Vitaliy S. Eremenko^{b)}, Michail I. Patuk^{c)}

V.I. Vernadsky State geological museum of RAS; Moscow, Russia

^{a)}naumova_new@mail.ru, ^{b)}vitaer@gmail.com, ^{c)}patuk@mail.ru

Abstract: The development of an Information system for the creation, development and support of scientific research in geology is discussed: up to the stage of a Single digital space of geological scientific knowledge. The current stage of the Project under development is a step forward from the traditional geographically distributed geological system to the digital environment of the future.

Key words: *digital space of scientific knowledge, digital environment, intelligent workspace*

For citation: Naumova V.V., Eremenko V.S., Patuk M.I. Development of a single digital space geological scientific knowledge. *Geoinformatika*. 2022;(4):8–13. <https://doi.org/10.47148/1609-364X-2022-4-8-13>. In Russ.

В ряде публикаций последних лет появилось понятие единого цифрового пространства знаний, которое рассматривается как некоторая перспектива создания, развития и интеграции информационных ресурсов и сервисов в определенной области науки в современной цифровой среде.

Под единым цифровым пространством знаний мы будем понимать компьютерную среду, обращаясь к которой пользователь должен получить ответы на вопросы, касающиеся тех или иных областей науки. Эта среда должна содержать достоверную фундаментальную и научно-популярную информацию; состоять из ряда подпространств, вообще говоря, связанных между собой, относящихся к отдельным научным направлениям; каждое подпространство должно иметь фундаментальную (статическую) основу, включающую базовые постулаты и результа-

ты данного научного направления, и динамическую часть, включающую информацию о новейших достижениях в данной научной области [1, 2].

Современные научные информационные ресурсы и сервисы

Активно формируются и развиваются научные электронные библиотеки, которые предоставляют широкому исследовательскому сообществу научную информацию и разнообразные сервисы.

Одним из примеров современной научной электронной библиотеки может служить Национальная цифровая библиотека по науке США (The National Science Digital Library — NSDL, <https://nsdl.oercommons.org>), которая основана Национальным научным фондом (NSF) в 2000 г. с целью обеспечить доступ информации и сервисам высокого качества

по многим областям науки, технологий и техники. NSDL агрегирует ресурсы из цифровых библиотек и проектов, поддерживаемых NSF.

Научная электронная библиотека «КиберЛенинка» — крупнейшее в России собрание научных статей, в основном на русском языке.

Библиотека eLibrary. На этом сайте выкладываются статьи и научные публикации, входящие в РИНЦ (российский индекс научного цитирования).

Системы открытого доступа к данным и системам обработки

Основная цель U.S. Geoscience Information Network (<http://usgin.org>) — облегчить открытый доступ к совместимым цифровым данным и программному обеспечению в науках о Земле. USGIN стандарты, протоколы и задачи — наследие National Geothermal Data System (NGDS), системы совместного использования данных, которая обеспечивает доступ к информации о геотермических ресурсах.

Специализированные платформы интеграции и анализа геологических данных

An Open Source Geospatial Data Management & Analysis Platform. Open Data Cube (ODC) — это некоммерческий проект с открытым исходным кодом, который был создан для обеспечения доступа, управления и анализа больших объемов данных ГИС по мониторингу Земли. Он представляет собой общий аналитический фреймворк, содержащий наборы структурированных данных и инструментов, которые позволяют проводить анализ больших коллекций пространственных данных.

Macrostrat: A Platform for Geological Data Integration and Deep-Time Earth Crust Research [3]. Макрострат — это платформа для агрегирования и распределения геологических данных, относящихся к пространственному и временному распределению осадочных, магматических и метаморфических пород, а также данных, извлеченных из них. Он связан с цифровой библиотекой GeoDeepDive и системой машинного чтения и призван стать общественным ресурсом для добавления, редактирования и распространения новых стратиграфических, литологических, экологических и экономических данных.

PAIRS: A scalable geo-spatial data analytics platform — это геопространственный сервис больших данных [4]. Пары содержат огромное количество кураторских геопространственных (или, точнее, пространственно-временных) данных из большого количества публичных и частных информационных ресурсов, а также поддерживают пользовательские слои данных. PAIRS предлагает простую в использовании платформу как для быстрой сборки и извлечения геопространственных наборов данных, так и для выполнения сложной аналитики, что значительно сокращает время обнаружения за счет снижения нагрузки на куратора данных и управление ими.

Бурное развитие информационных технологий, в том числе организация хранения больших данных в облаках, организация их обработки и анализа на территориально распределенных интерактивных вычислительных ресурсах стимулируют разработку территориально-распределенных тематических информационных систем.

Проект ГГМ РАН «Разработка территориально распределенной информационно-аналитической геологической среды для поддержки и сопровождения научных исследований»

Основная цель данного Проекта Государственного геологического музея им В.И. Вернадского РАН (ГГМ РАН) «Разработка Информационно-аналитической геологической среды для поддержки научных исследований <http://geologyscience.ru>» заключается в организации единой точки доступа к геологическим данным на территорию России и системам их обработки с использованием возможностей поиска данных в территориально распределенных разнородных источниках, а также с использованием территориально-распределенных вычислительно-аналитических узлов для обработки данных, взаимодействие с которыми осуществляется с использованием технологии web-сервисов [5, 6, 7].

Интеграция разнотипных геологических данных и сервисов обработки в единую информационно-аналитическую среду на основе единых политик обеспечивает возможность комплексного анализа информации и позволит получать качественно новые знания о геологических объектах.

Предполагается, что создаваемое решение будет сочетать в себе программное обеспечение (ПО) интеграции разнотипных данных из разнородных источников, технологию стандартизации, ПО организации структурированного хранения, поиска и визуализации геологических данных и набор сервисов и инструментов по анализу и обработке различных типов геологических данных. Все это вместе составляет унифицированную платформу геологических данных.

Интеграция геологических данных по территории РФ

- *Пространственные данные* — Доступ к Государственным геологическим картам по территории РФ на основе данных Всероссийского научно-исследовательского геологического института им. А.П. Карпинского. Для поиска и получения доступны геологические карты масштаба 1:1 000 000 и метаданные геологических карт масштаба 1:200 000.

- *Количественные данные* — Удаленный доступ к данным из мировых центров данных, например GEOROCK.

- *Музейные данные* — Доступ к геологическим данным мировых музеев естественной истории по

территории России, включая данные Государственного геологического музея им. В.И. Вернадского РАН.

- *Спутниковые данные* — Унифицированный доступ к данным спутниковых платформ Aqua, Terra, Landsat, Orbview-3 и другим данным среднего и высокого пространственного разрешения. Основные источники спутниковых данных: информационные системы ИАПУ ДВО РАН, USGS и NASA.

- *Научные публикации* — Доступ к научным публикациям по наукам о Земле (геология, геохимия, петрология, минералогия, тектоника, геоморфология, вулканология, палеонтология и т.д.).

- *Экспертные данные* — Энциклопедия по геологическим объектам РФ (месторождения полезных ископаемых).

- *Базы данных* — Доступ к метаданным по месторождениям полезных ископаемых и геологическим отчётам Росгеолфонда.

Вычислительно-аналитический блок — облачный инструментальный пользовательский интерфейс для обработки различных типов геологических данных. Предложенный при построении подход предполагает использование внешних вычислительных узлов для обработки данных.

Реализованная платформа выступает в роли посредника между пользователем и внешними системами обработки, предоставляя единый интерфейс доступа ко всем алгоритмам обработки, имеющимся во внешних системах обработки (узлах системы). Описанная архитектура предполагает возможность использования данных не только из доступных в системе открытых источников, но и загрузку данных для обработки самим пользователем. В разрабатываемой Среде вычислительно — аналитические блоки обработки и анализа геологической информации организованы в виде наборов служебных и аналитических функций с возможностью пользовательского доступа к выбору метода обработки; цепочек обработки, включающих загрузку данных, трансформацию форматов, методов анализа и визуализации результатов; тематических цепочек, осуществляющих последовательность методов анализа. Доступ к сервисам обработки и анализа осуществляется через платформу управления распределёнными сервисами обработки данных.

Некоторые вычислительные сервисы среды: предварительные методы обработки количественных данных, многомерные методы анализа данных, обработка петролого-геохимических данных, структурных анализ публикаций, обработка естественного языка и др.

Разработка единого цифрового пространства геологических научных знаний

Предполагается, что единое цифровое пространство научных геологических знаний поможет сформировать новые процессы генерации данных и знаний. В последнее время нами осуществля-

лась разработка методических подходов и технологических решений для проектирования единого цифрового пространства геологических научных знаний на базовой основе созданной ранее Информационно-аналитической среды для поддержки и сопровождения научных исследований в геологии, осуществляющей интеграцию территориально распределённой геологической информации с использованием специализированных служб её анализа и обработки (<http://geologyscience.ru/>).

Использованные подходы и модульная архитектура при проектировании Информационно-аналитической геологической среды позволяют трансформировать её в единое цифровое пространство научных геологических знаний для решения широкого круга задач для наук о Земле.

А простота использования и гибкость персональной настройки под определённые задачи позволит привлечь к её использованию широкий круг учёных-исследователей. Для этого предполагается разработать систему тематических рабочих пространств, сочетающих в себе необходимые инструменты, алгоритмы, источники данных для решения тех или иных прикладных задач.

В статье Наумовой и др. [8] описываются все стадии развития Информационной системы для создания, развития и поддержки научных исследований в геологии: от Информационно-аналитической среды до Единого цифрового пространства геологических научных знаний.

Интеллектуальное цифровое рабочее пространство

Интеллектуальное цифровое рабочее пространство — это пространство, которое адаптировано под нужды и интересы конкретного пользователя. Сегодня интеллектуальное цифровое рабочее пространство сводит к минимуму операции пользователя, не связанные с выполнением целевых функций. «Умные» системы и приложения, сокращают время, затрачиваемое пользователем на рутинные операции, будь то поиск контента, ответы на вопросы или решение стандартных задач, а также делают повседневную работу физически более комфортной.

Учёному-исследователю в области геологии необходимо на основании его текстового запроса предоставить набор данных из территориально распределённых источников научных данных. Данный набор необходимо структурировать и установить взаимосвязи между данными, если это возможно, исходя из контекста задачи.

Предложены основные компоненты для ядра будущей системы, включающей в себя рабочие пространства для научных сотрудников в геологии.

А именно:

а) Компонент, отвечающий за интеграцию различных видов данных (файлы, таблицы, ссылки на внешние ресурсы, текстовые данные, географиче-

ски привязанные данные, мультимедиа — информация и т.д.) в рабочее пространство;

б) Компонент, отвечающий за загрузку данных с внешних источников;

в) Компонент, позволяющий осуществлять связывание данных между собой либо автоматически, либо на основе выбранного поля данных или поля метаданных;

г) Компонент, отвечающий за визуализацию данных в рабочем пространстве.

В настоящее время авторами осуществляется разработка Единого цифрового пространства научных геологических знаний, а рабочая версия реализуется на Портале <http://geologyscience.ru>

Одним из примеров текущей реализации цифровой системы управления геологическими знаниями является «Wiki-Геология России», которая интегрирует геологическую информацию из других частей системы и из внешних данных [9, 10].

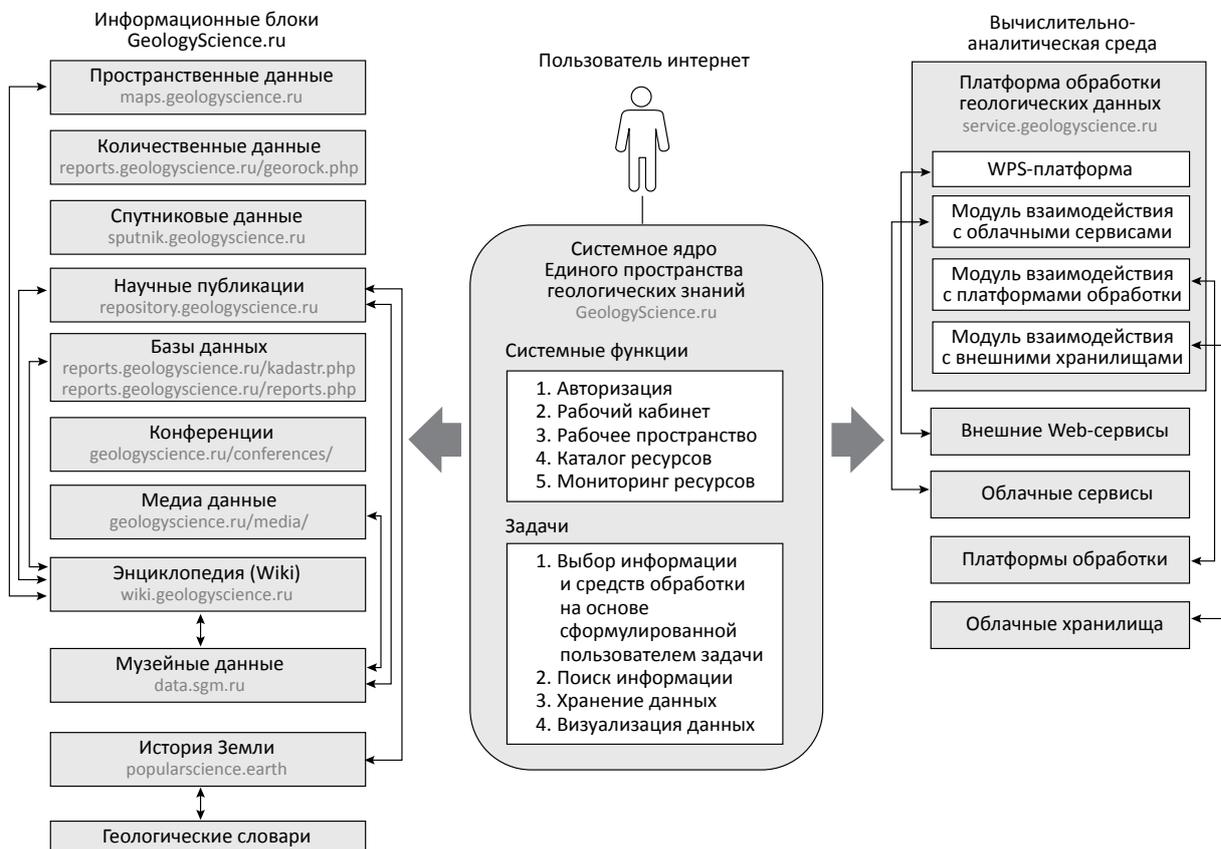
«Wiki-Геология России» предоставляет пользователям не только информацию о геологических объектах, но и ряд сервисов по интеграции информации и ее анализу. Предполагается, что система

будет интегрировать не только дополнительную информацию из научных публикаций, государственных геологических карт, но и спутниковые данные, музейные образцы и другую медиа информацию. Авторами разрабатываются и сервисы анализа информации, в том числе и тематические сервисы. В настоящее время пользователям Wiki-педии доступны следующие сервисы: построение географической карты района месторождения с указанием точки, где расположено месторождение; визуализация геологической карты района месторождения; определение возраста месторождения. В дальнейшем предполагается развитие и других сервисов например, связанных с металлогеническими построениями и др.

Предполагается, что единое цифровое пространство научных геологических знаний поможет сформировать новые процессы генерации данных и знаний.

Работы выполняются в рамках Государственного задания ГТМ РАН по Теме № 1021061009468-8-1.5.1 «Цифровая платформа интеграции и анализа геологических и музейных данных».

Рис. 1. Концептуальная схема Единого цифрового пространства научных геологических знаний, созданная на базовой основе Информационно-вычислительной геологической среды по поддержке научных исследований
 Fig. 1. Conceptual scheme of the Unified digital space of scientific geological knowledge, created on the basic basis of the Information and computing geological environment to support scientific research



Список источников

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестник Российской Академии наук. – 2019. – Т. 89. – № 7. – С. 728–735. DOI: 10.31857/S0869-5873897728-735.
2. Единое цифровое пространство научных знаний: проблемы и решения : сб. науч. трудов / под ред. Н.Е. Каленова, А.Н. Сотникова. – Москва ; Берлин : Директмедиа Паблишинг, 2021. – 464 с. DOI: 10.23681/610687.
3. Peters S.E., Husson J.M., Czaplewski J. Macrostrat: a platform for geological data integration and deep-time Earth crust research [Электронный ресурс] // *Geochemistry, Geophysics, Geosystems*. – 27.01.2018. – Режим доступа: <https://eartharxiv.org/repository/view/1461> (дата обращения 10.11.2022). DOI:10.17605/OSF.IO/YNAXW.
4. Klein L.J., Marianno F.J., Albrecht C.M., Freitag M., Lu S., Ninds N., Shao X., Bermudez Rodriguez S., Hamann H.F. PAIRS: A scalable geo-spatial data analytics platform // 2015 IEEE International Conference on Big Data (Big Data). – 2015. – pp. 1290–1298. DOI: 10.1109/BigData.2015.7363884.
5. Naumova V.V., Platonov K.A., Eremenko V.S., Dyakov S.E. Information and Analytical Environment to Support Scientific Research in Geology: Current Status and Development Perspectives // Proceedings of the V International Conference “Information Technologies in Earth Sciences and Applications for Geology, Mining and Economy (ITES&MP-2019)” (Moscow, 14-18 October 2019) (CEUR Workshop Proceedings). – 2019. – Vol. 2527. – pp. 36–41.
6. Наумова В.В., Платонов К.А., Еременко В.С., Патук М.И., Дьяков С.Е. Информационно-аналитическая среда для поддержки научных исследований в геологии: текущее состояние и перспективы развития // Распределенные информационно-вычислительные ресурсы. Цифровые двойники и большие данные (DICR-2019) : труды XVII Междунар. конф. (Новосибирск, 3-6 декабря 2019 г.). – Новосибирск, ИВТ СО РАН, 2019. – С. 139–147. DOI: 10.25743/ICT.2019.70.61.021.
7. Naumova V.V., Eremenko V.S., Platonov K.A., Dyakov S.V., Patuk M.I., Eremenko A.S. Development of geographically distributed information-analytical geological environment // *Russian Journal of Earth Sciences*. – 2019. – Vol. 19. – No. 6. – ES6012. DOI: 10.2205/2019ES000696.
8. Наумова В.В., Ерёменко В.С., Еременко А.С., Загумённов А.А., Патук М.И. От информационно-аналитической среды для поддержки научных исследований в геологии к единому цифровому пространству геологических научных знаний // Электронные библиотеки. – 2022. – Т. 21. – № 1. – С. 1–28.
9. Патук М.И., Наумова В.В. Построение цифровой системы управления геологическими знаниями для поддержки научных исследований // Электронные библиотеки. – 2022 – Т. 25. – № 1. – С. 1–11.
10. Патук М.И., Наумова В.В. Построение цифровой системы управления геологическими знаниями для поддержки научных исследований // Информационные технологии для наук о Земле и цифровизация в геологии и горнодобывающей промышленности. ITES-2022 : материалы VI Всерос. конф. (Владивосток, 3–8 октября 2022 г.) / сост. А.С. Еременко. – Владивосток : ДВФУ, 2022. – С.24–25. DOI: 10.24866/7444-5341-1.

References

1. Antopolskii A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N. Common digital space of scientific knowledge. *Vestnik Rossijskoj Akademii nauk*. 2019;89(7);728–735. DOI: 10.31857/S0869-5873897728-735.
2. Kalenov N.E., Sotnikov A.N. (eds.) *Edinoe tsifrovoe prostranstvo nauchnykh znaniy: problemy i resheniya : sb. nauch. trudov* [A single digital space of scientific knowledge: problems and solutions: collection of scientific works]. Moscow; Berlin: Direktmedia Publishing; 2021. 464 p. DOI: 10.23681/610687.
3. Peters S.E., Husson J.M., Czaplewski J. Macrostrat: a platform for geological data integration and deep-time Earth crust research. *Geochemistry, Geophysics, Geosystems*. 27.01.2018. Available at: <https://eartharxiv.org/repository/view/1461> (accessed 10.11.2022). DOI:10.17605/OSF.IO/YNAXW.
4. Klein L.J., Marianno F.J., Albrecht C.M., Freitag M., Lu S., Ninds N., Shao X., Bermudez Rodriguez S., Hamann H.F. PAIRS: A scalable geo-spatial data analytics platform. In: 2015 IEEE International Conference on Big Data (Big Data). 2015. pp. 1290-1298. DOI: 10.1109/BigData.2015.7363884.
5. Naumova V.V., Platonov K.A., Eremenko V.S., Dyakov S.E. Information and Analytical Environment to Support Scientific Research in Geology: Current Status and Development Perspectives. In: Proceedings of the V International Conference “Information Technologies in Earth Sciences and Applications for Geology, Mining and Economy (ITES&MP-2019)” (Moscow, 14-18 October 2019) (CEUR Workshop Proceedings). 2019;2527: 36–41.
6. Naumova V.V., Platonov K.A., Eremenko V.S., Patuk M.I., Dyakov S.E. Information and analytical environment to support scientific research in geology: current status and prospects for development. In: *Raspredelennye informatsionno-vychislitel'nye resursy. Tsifrovye dvoyniki i bol'shie dannye (DICR-2019): trudy XVII Mezhdunar. konf. (Novosibirsk, 3—6 December 2019)*. Novosibirsk; IVT SO RAN, 2019. p. 139–147. DOI: 10.25743/ICT.2019.70.61.021.
7. Naumova V.V., Eremenko V.S., Platonov K.A., Dyakov S.V., Patuk M.I., Eremenko A.S. Development of geographically distributed information-analytical geological environment. *Russian Journal of Earth Sciences*. 2019;19(6):ES6012. DOI: 10.2205/2019ES000696.
8. Naumova V.V., Eremenko V.S., Eremenko A.S., Zagumennov A.A., Patuk M.I. Ot informatsionno-analiticheskoi sredy dlya podderzhki nauchnykh issledovaniy v geologii k edinomu tsifrovomu prostranstvu geologicheskikh nauchnykh znaniy [From an information and analytical environment supporting scientific research in geology to a single digital space of geological scientific knowledge]. *Russian digital libraries journal*. 2022;21(1): 1–28.
9. Patuk M.I., Naumova V.V. Postroenie tsifrovoi sistemy upravleniya geologicheskimi znaniyami dlya podderzhki nauchnykh issledovaniy [Building a digital geological knowledge management system to support scientific research]. *Russian digital libraries journal*. 2022;25(1):1–11.
10. Patuk M.I., Naumova V.V. Building a digital geological knowledge management system to support scientific research. In: *Informatsionnye tekhnologii dlya nauk o Zemle i tsifrovizatsiya v geologii i gornodobyvayushchei promyshlennosti. ITES-2022 : materialy VI Vseros. konf. (Vladivostok, 3–8 October 2022)*. Eremenko A.S., ed. Vladivostok: DVFU; 2022. p. 24–25. DOI: 10.24866/7444-5341-1.

Статья поступила в редакцию 25.10.2022, одобрена после рецензирования 27.10.2022, принята к публикации 15.12.2022.
The article was submitted 25.10.2022; approved after reviewing 27.10.2022; accepted for publication 15.12.2022.

Информация об авторах

Наумова Вера Викторовна

Доктор геолого-минералогических наук
Главный научный сотрудник, заведующий Научным отделом
Государственного геологического музея им. В.И. Вернадского
Российской Академии Наук
125009 Москва, Моховая ул., д. 11, стр. 11
e-mail: naumova_new@mail.ru
ORCID: 0000-0002-3001-1638

Ерёменко Виталий Сергеевич

Младший научный сотрудник Научного отдела
Государственного геологического музея им. В.И. Вернадского
Российской Академии Наук
125009 Москва, Моховая ул., д. 11, стр. 11
e-mail: vitaer@gmail.com
ORCID: 0000-0002-5250-5743

Патук Михаил Иванович

Кандидат геолого-минералогических наук,
и. о. научного сотрудника Научного отдела Государственного
геологического музея им. В.И. Вернадского РАН
125009 Москва, Моховая ул., д. 11, стр. 11
e-mail: patuk@mail.ru
ORCID: 0000-0003-3036-2275

Information about authors

Vera V. Naumova

Doctor of Geological and Mineralogical Sciences,
Chief Researcher
Head of Scientific Department of Vernadsky State Geological
Museum RAS (SGM RAS)
11 build. 11, Mokhovaya str., Moscow, 125009, Russia
e-mail: naumova_new@mail.ru
ORCID: 0000-0002-3001-1638

Vitaliy S. Eremenko

Junior Researcher
Scientific Department of Vernadsky State Geological Museum RAS
(SGM RAS)
11 build. 11, Mokhovaya str., Moscow, 125009, Russia
e-mail: vitaer@gmail.com
ORCID: 0000-0002-5250-5743

Michail I. Patuk

Candidate of Geological and Mineralogical Sciences
Scientific Department of Vernadsky State Geological Museum RAS
(SGM RAS)
11 build. 11, Mokhovaya str., Moscow, 125009, Russia
e-mail: patuk@mail.ru
ORCID: 0000-0003-3036-2275