

УДК 004.550

ОТ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ СРЕДЫ ДЛЯ ПОДДЕРЖКИ НАУЧНЫХ ИССЛЕДОВАНИЙ В ГЕОЛОГИИ К ЕДИНОМУ ЦИФРОВОМУ ПРОСТРАНСТВУ ГЕОЛОГИЧЕСКИХ НАУЧНЫХ ЗНАНИЙ

В. В. Наумова¹ [0000-0002-3001-1638], **В. С. Ерёменко**² [0000-0002-5250-5743],
А. С. Еременко³ [0000-0003-1923-8417], **А. А. Загумённов**⁴ [0000-0002-0501-5362],
М. И. Патук⁵ [0000-0003-3036-2275]

^{1, 2, 5} ФГБУН Государственный геологический музей им. В.И. Вернадского РАН
Москва, Россия

^{3, 4} ФГБУН Институт автоматизации и процессов управления ДВО РАН
Владивосток, Россия

¹naumova_new@mail.ru, ²vitaer@gmail.com, ³academy21@gmail.com,

⁴trueepikvic@gmail.com, ⁵patuk@mail.ru

Аннотация

Описаны 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. Современная разрабатываемая стадия Проекта – это шаг вперед от традиционной территориально распределенной геологической системы к цифровой среде будущего.

Ключевые слова: Веб-сервисы, облачные сервисы, обработка геологических данных, геологическая программная платформа, программная экосистема для геологических исследований, единое цифровое пространство геологических научных знаний.

ВВЕДЕНИЕ

Сегодня ИТ-технологии сосредоточены на разработке передовых технологий и рабочих процессов для поддержки успешных открытий всех геологов, геофизиков и геохимиков, чтобы сделать их работу более простой, естественной, мощной. Меньшие затраты времени на обработку данных дают больше свободы в исследованиях и помогают идти в ногу с современными требованиями и возможностями науки.

1. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ГЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ И СПЕЦИАЛИЗИРОВАННЫЕ ВЕБ-СЕРВИСЫ

Задача обработки и анализа данных в геологии требует применения большого количества различных алгоритмов, процедур обработки и соответствующих им программных решений. Зачастую при исследовании конкретного объекта в геологии приходится использовать разнотипные данные, такие как геологические карты, количественные данные, спутниковые данные, музейные данные, текстовые данные и т. д.

Каждый из типов геологических данных имеет собственный набор алгоритмов анализа и обработки с соответствующим программным обеспечением, реализующим эти процедуры обработки и анализа. Комплексный же анализ различных типов данных в рамках решения одной задачи представляет серьезные трудности. Поэтому необходимо предоставить исследователям возможность проводить анализ и обработку разнотипных геологических данных в рамках единой системы с возможностью интеграции полученных результатов в качественно новый результат. В силу непрерывного роста количества геологических данных и инструментов их обработки более перспективным подходом является использование удаленных вычислительных систем, предоставляющих сервисы по обработке отдельных типов геологических данных.

В настоящее время разработчики программных решений по обработке данных активно внедряют в процесс обработки облачные вычисления. Использование внешних сервисов вместо пользовательских приложений позволяет проводить обработку данных на оборудовании, наиболее приспособленном под соответствующие задачи. Тем самым обработка данных происходит более эффективно, а пользователь получает возможность обрабатывать данные наиболее актуальными версиями алгоритмов, используя для этого веб-интерфейс, без необходимости установки, настройки и поддержки программного обеспечения для обработки у себя на персональном компьютере.

Построение информационных систем по обработке данных, основанных на взаимодействии с пользователем через веб-интерфейс, активно ведётся в различ-

ных предметных областях. Так, в области обработки спутниковых данных к подобным системам можно отнести проект NASA Giovanni¹, систему Google Earth Engine², платформы ESA G-POD³, а также информационные системы See the Sea [1], Vega-Science⁴ и т. д., созданные в ИКИ РАН на основе собственной технологии GEOSMIS [2]. В области обработки и анализа пространственных данных примером такой информационной системы является ArcGIS Online⁵.

2. СПЕЦИАЛИЗИРОВАННЫЕ ПЛАТФОРМЫ ИНТЕГРАЦИИ И АНАЛИЗА ГЕОЛОГИЧЕСКИХ ДАННЫХ

ИТ-платформа определяется нами как технологическая база, на которой могут взаимодействовать дополняющие друг друга надстройки, соответствующие стандартам и позволяющие осуществлять операции между заинтересованными сторонами в рамках экосистемы, ориентированной на платформу.

AN OPEN-SOURCE GEOSPATIAL DATA MANAGEMENT & ANALYSIS PLATFORM. Open Data Cube (ODC)⁶ — 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.

ODC был разработан для анализа пространственных данных на больших временных промежутках, однако его можно использовать на любых наборах пространственных данных. Данные могут представлять собой модели высот, геофизические сетки, интерполированные поверхности и т. д.

Ключевой особенностью ODC является возможность сохранения каждого уникального элемента набора пространственных данных, в отличие от многих других методов работы с большими коллекциями пространственных данных.

¹ <https://giovanni.gsfc.nasa.gov/giovanni>

² <https://earthengine.google.com>

³ <https://gpod.eo.esa.int>

⁴ <http://sci-vega.ru>

⁵ <https://www.arcgis.com>

⁶ <https://www.opendatacube.org>

Основные преимущества ODC:

- 1) Гибкая структура;
- 2) Пользовательский контроль над своими данными;
- 3) Смещение парадигмы от анализа сцен к пиксельному анализу;
- 4) Простой доступ к анализу данных дистанционного зондирования.

Система Open Data Cube предназначена для:

- 1) Каталогизации больших объёмов данных ДЗЗ;
- 2) Предоставления программного интерфейса приложения (API) на языке программирования Python для высокопроизводительных запросов и доступа к данным;
- 3) Предоставления учёным и другим пользователям простого доступа к исследовательскому анализу данных;
- 4) Обеспечения масштабируемой обработки имеющихся данных;
- 5) Контроля за источниками всех имеющихся данных для обеспечения качественного менеджмента и обновлений.

ODC состоит из следующих основных компонентов:

- 1) Command Line Tools - набор инструментов, обеспечивающих программный интерфейс к ODC;
 - 2) Open Data Cube Explorer – веб-приложение, предоставляющее визуальный и интерактивный интерфейс для управления доступными данными;
 - 3) Open Data Cube Stats – оптимизированные средства для определения и выполнения расширенного анализа в ODC. Данный инструмент ориентирован на научные исследования;
 - 4) Web User Interface (UI) – веб-приложение, предоставляющее разработчикам демонстрировать результаты работы алгоритмов анализа в интерактивном режиме;
 - 5) Jupyter Notebooks – документы для исследований с использованием инструментов пространственного анализа. Эти документы содержат выполняемый код с подробными примерами по работе с данными для научных исследований;
 - 6) Open Geospatial Consortium (OGC) Web Services – инструменты для подключения сторонних сервисов в ODC.
-

MACROSTRAT: A Platform for Geological Data Integration and Deep-Time Earth Crust Research [3]. Это платформа для агрегирования и распределения геологических данных, относящихся к пространственному и временному распределениям осадочных, магматических и метаморфических пород, а также данных, извлеченных из них. Она связана с цифровой библиотекой GeoDeepDive и системой машинного чтения и призвана стать общественным ресурсом для добавления, редактирования и распространения новых стратиграфических, литологических, экологических и экономических данных. Интерактивные приложения, построенные на базе Macrostrat, предназначены для образовательных и исследовательских целей. Все данные предоставляются свободно по лицензии CC-BY-4.0.

PAIRS: A scalable geo-spatial data analytics platform — это геопро пространственный сервис больших данных [4]. Пары содержат огромное количество кураторских геопро пространственных (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 геопро пространственных наборов данных, так и для выполнения сложной аналитики, что значительно сокращает время обнаружения за счет снижения нагрузки на курацию данных и управление ими. В этой статье мы рассмотрим недавний прогресс с парами и продемонстрируем несколько образцовых аналитических приложений, которые авторы могут построить с относительной легкостью, используя эту технологию.

ИНТЕГРИРОВАННАЯ ПЛАТФОРМА OASIS MONTAJ⁷ компании Geosoft обеспечивает доступ ко всем заданным геофизическим, геохимическим и геологическим данным в высокопроизводительных базах, позволяет быстро строить профессиональные карты и 3D-визуализации среды, динамически связанные между собой. Данные быстро и легко оцениваются и обрабатываются с помощью многочисленных процессов и фильтров. Для построения сетей больших объемов данных используются алгоритмы минимальной кривизны, двунаправленного и прямого гридирования, кригинга, обратно взвешенных расстояний и триангуляции. В си-

⁷ <https://www.seequent.com/products-solutions/geosoft-oasis-montaj/>

стему встроена технология Esri, обеспечивающая создание, просмотр и совместное использование карт ArcGIS MDX. Поддерживается работа практически со всеми типами форматов данных, включая форматы CAD, GIS, планирования и моделирования горных работ. Более 20 доступных программных модулей включают углубленную обработку геофизических и геохимических данных, проведение анализа и контроля качества данных, трехмерную визуализацию скважин, фильтрацию, уравнивание, интерпретацию данных и моделирование. Специализированные модули позволяют обрабатывать полный спектр гамма-спектрометрических съемок.

Компания «Совзонд» выпустила новую версию российской геоинформационной платформы GEOANALITIKA PLATFORM v.1.1⁸. Эта версия характеризуется рядом новых функциональных возможностей в части управления и анализа данных, подключения внешних веб-сервисов, а также общей оптимизации производительности ГИС-сервера. Функционально Geoanalitika Platform включает 5 подсистем: управления данными; публикации веб-сервисов; визуализации; администрирования; интеграции.

3. УНИФИЦИРОВАННАЯ ПЛАТФОРМА ГГМ РАН ДЛЯ ОРГАНИЗАЦИИ ДОСТУПА К ГЕОЛОГИЧЕСКИМ ДАННЫМ И ВЫЧИСЛИТЕЛЬНЫМ СЕРВИСАМ ПО ЕДИНОМУ УНИВЕРСАЛЬНОМУ ПРОТОКОЛУ

Основная цель Проекта Государственного геологического музея им В.И. Вернадского РАН (ГГМ РАН) «Разработка Информационно-аналитической геологической среды для поддержки научных исследований GeologyScience.ru» заключается в организации единой точки доступа к геологическим данным на территории 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 веб-сервисов [5, 6, 7]. Интеграция разнотипных геологических данных и сервисов обработки в единую информационно-аналитическую среду на основе единых политик обеспечивает возможность комплексного

⁸ <https://geoanalitika.com>

анализа информации и позволит получать качественно новые знания о геологических объектах.

Разрабатываемая в ГГМ РАН платформа WPS сервисов⁹ [8] является частью информационно-аналитической среды для поддержки научных исследований в геологии GeologyScience.ru¹⁰. Она предоставляет пользователям доступ к разнообразным территориально распределенным веб-сервисам обработки и анализа геологической информации. Для организации единого механизма взаимодействия с веб-сервисами используется интерфейс OGC Web Processing Service. Для каждого внешнего веб-сервиса создан соответствующий ему WPS-процесс, реализующий интерфейс взаимодействия с внешним вычислительным узлом.

Платформа предоставляет доступ к следующим вычислительным узлам:

- Вычислительный узел «Многомерные методы анализа данных», разработанный в ГГМ РАН [9, 10], который позволяет выполнять обработку табличных данных различными методами анализа данных с настройкой их параметров и визуализацией результатов. Вычислительный узел включает в себя такие группы методов, как статистический анализ, регрессионный анализ, факторный анализ, кластеризация, машинное обучение, методы визуализации. Вычисления производятся в среде Python с использованием известных пакетов обработки данных: Scikit-learn, Pandas, Matplotlib и других. Обработка поступающих запросов на обработку производится фреймворком Flask посредством REST API с использованием очереди заданий, реализованной на основе NoSQL базы данных Redis. Такая архитектура позволяет разделить обработку запросов и «тяжелые вычисления» больших объёмов данных, что обеспечивает отказоустойчивость и масштабируемость узла;

- Обработка петролого-геохимических данных [11]. В Институте Физики Земли РАН разработана интерактивная база методов обработки петролого-геохимических данных. Эта система предоставляет сервисы построения спайдерграмм, гистограмм и классификационных диаграмм; сервис идентификации минералов

⁹ <https://service.geologyscience.ru>

¹⁰ <http://geologyscience.ru>

по их химическому составу; сервис интерпретации состава минерала и разложение на минералы и т. д. Интерфейс взаимодействия с сервисами построен на основе REST архитектуры;

- Структурный анализ публикаций [12]. В междисциплинарном центре математического и вычислительного моделирования (Университет Варшавы, Польша) разработан сервис для извлечения метаданных из научных публикаций. Метаданные включают в себя авторов, их аффилиации, абстракт, ключевые слова, название журнала, объем, год выпуска, разобранные библиографические ссылки, структуру разделов документа, заголовки разделов и абзацы. Интерфейс взаимодействия с сервисами построен на основе REST архитектуры;

- Обработка естественного языка [13]. В Университете Шеффилда в рамках проекта GATE (General Architecture for Text Engineering) разработан ряд сервисов по обработке текстовых данных для различных языков. Для обработки текстовых данных на русском языке предоставляются сервисы по определению частей речи слов, а также выделению именованных сущностей, таких как имена и фамилии, названия организаций, географические названия, даты, денежные единицы и т. д. Интерфейс взаимодействия с сервисами построен на основе REST архитектуры.

В результате проведенных работ нами разработана цифровая платформа по обработке и анализу геологических данных DPAP (рис. 1).

Разрабатываемая платформа по обработке и анализу геологических данных [14] представляет собой программный продукт на языке программирования Java в формате веб-приложения с использованием фреймворка Spring Boot. Платформа реализована с использованием принципов слабосвязанной архитектуры. Для взаимодействия с пользователем реализован веб-интерфейс с использованием JavaScript фреймворка Angularjs. Также разработан программный интерфейс с использованием архитектурного стиля REST для взаимодействия с платформой в формате система–система.

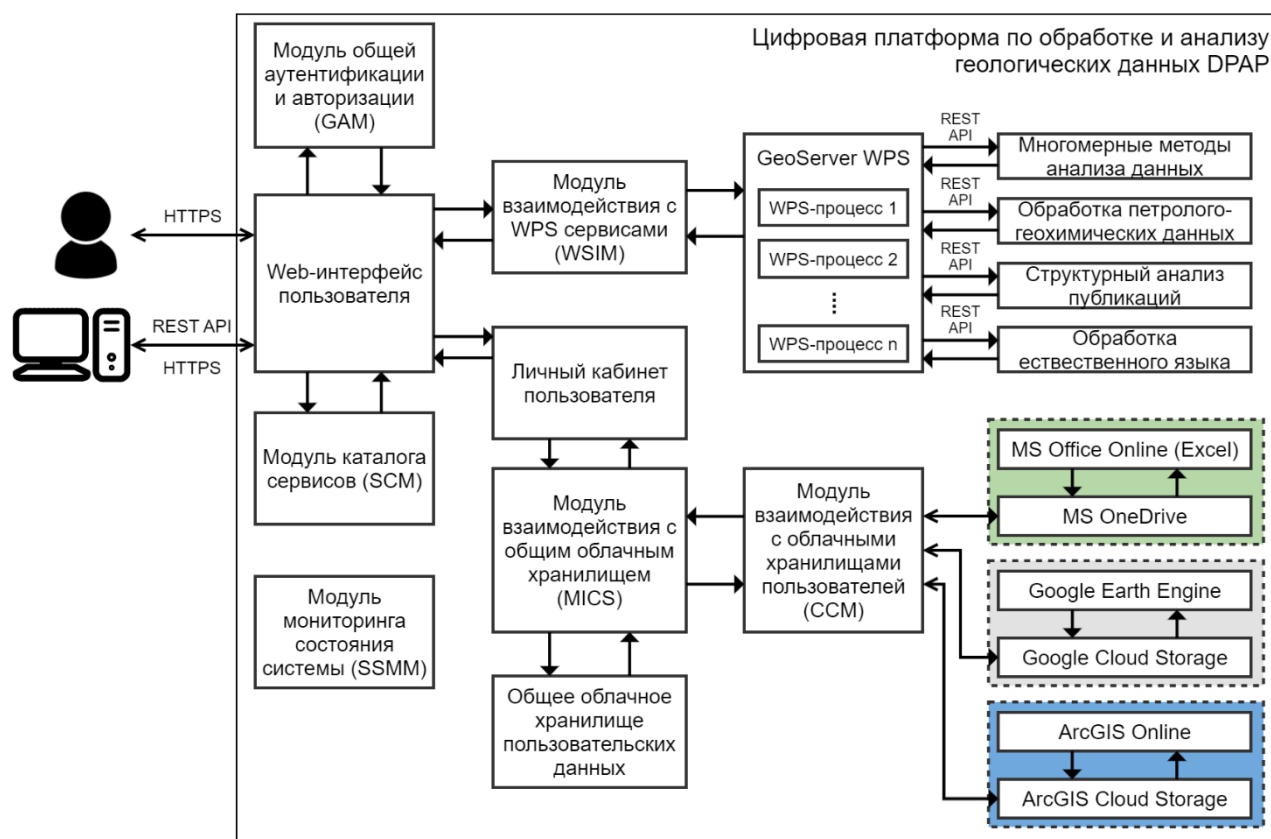


Рис. 1. Цифровая платформа по обработке и анализу геологических данных

Для выполнения поставленных задач разработан ряд авторских модулей:

1. WSIM – модуль взаимодействия с внешними веб-сервисами. Основная функция модуля: обеспечение единого интерфейса доступа к внешним сервисам обработки и анализа, работающих по принципу запрос–ответ. В качестве интерфейса доступа используется OGC Web Processing Service 1.0. Каждый используемый Средой метод публикуется в виде отдельного процесса внутри WPS сервиса. Для реализации WPS сервиса используется программный пакет с открытым исходным кодом GeoServer. GeoServer содержит встроенные компоненты для реализации очереди заданий, разграничения доступа к различным процессам и возможность как синхронного, так и асинхронного выполнения процесса. Для реализации модуля используются программные библиотеки GeoTools и 52 North WPS. Взаимодействие платформы с данным модулем осуществляется через WPS интерфейс.

2. MICS – модуль взаимодействия с общим облачным хранилищем пользователя. Основная функция модуля: обеспечение интерфейса доступа к облачному хранилищу пользователя, используемому в информационно-аналитической системе GeologyScience.ru. Пользователь имеет возможность сохранить данные, выбранные в информационной системе GeologyScience.ru, в облачном хранилище для дальнейшего их использования в других подсистемах, в том числе для обработки и анализа. Результаты обработки также могут быть сохранены пользователем в облачном хранилище для их последующего использования в других сервисах. Взаимодействие с платформой осуществляется через REST интерфейс.

3. GAM – модуль общей авторизации. Основная функция модуля: обеспечение единой точки аутентификации и авторизации для пользователей всех блоков информационно-аналитической системы GeologyScience.ru. Модуль обращается к сервису авторизации на базе подхода Single Sign-On (SSO) с использованием протокола OAuth. Для реализации взаимодействия модуля с сервисом аутентификации и авторизации используется фреймворк Spring Security.

4. CCM – модуль взаимодействия с внешними облачными хранилищами пользователей. Основная функция модуля: организация единого пространства данных на основе имеющихся у пользователей аккаунтов поставщиков услуг облачного хранения для дальнейшего использования данных при работе с разнородными интерактивными облачными сервисами. Поставщики облачных сервисов часто требуют предварительной загрузки данных для обработки в их собственные облачные хранилища. Настоящий модуль позволит исследователю при переходе от одного сервиса к другому в автоматическом режиме перемещать данные в требуемое хранилище. Модуль является основным инструментом, позволяющим интегрировать разнородные интерактивные сервисы в рамках разрабатываемой платформы. Взаимодействие с платформой осуществляется через REST интерфейс.

5. SCM – модуль каталога внешних сервисов [15]. Основные функции модуля: обеспечение пользователя информацией о сервисах, имеющихся в Среде, с возможностью получения информации о каждом зарегистрированном сервисе, включая общее описание, данные о поставщике сервиса и техническую информа-

цию для организации взаимодействия. Информация из данного модуля используется в модуле мониторинга для отслеживания состояния каждого внешнего сервиса, зарегистрированного в Среде. Взаимодействие с платформой осуществляется с использованием REST интерфейса.

6. SSMM – Модуль мониторинга состояния системы и сервисов [15]. Основная функция модуля: отслеживание статуса каждого внешнего поставщика сервиса, зарегистрированного в Среде, а также всех блоков системы GeologyScience.ru. Мониторинг состояния сервисов производится по нескольким параметрам, включая доступность узла, на котором размещён сервис, доступность сервиса по указанному протоколу и проверку сервиса на изменения путём отправки тестовых запросов к сервису. Взаимодействие с платформой осуществляется через REST интерфейс.

4. ЦИФРОВАЯ ЭКОСИСТЕМА ПОДДЕРЖКИ И СОПРОВОЖДЕНИЯ НАУЧНЫХ ГЕОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Активное развитие технологических решений в области обработки и анализа данных требует использования более современных подходов для организации доступа пользователей к таким решениям. Помимо привычных настольных приложений, веб-сервисов и веб-приложений, появляется большое количество различных программных платформ, предоставляющих широкий набор инструментов по обработке и анализу разнотипных данных. Интеграция такого количества разнородных источников инструментов обработки в рамках единой программной платформы становится крайне сложной задачей.

Последнее десятилетие в IT отрасли происходит активный переход от платформенной архитектуры к использованию цифровых экосистем.

Цифровая экосистема представляет собой распределенную, адаптивную, открытую социально-техническую систему со свойствами самоорганизации, масштабируемостью и устойчивостью. Модели цифровых экосистем основаны на знаниях о природных экосистемах, особенно в аспектах, связанных с конкуренцией и сотрудничеством между различными организациями. Экосистемный подход позволяет организовать взаимодействие различных сервисов, платформ и

продуктов компании, а также сервисов внешних партнёров, в рамках единой среды.

Основными свойствами цифровой экосистемы являются: стабильность, омниканальность и отказоустойчивость.

Для взаимодействия двух и более цифровых платформ в рамках единой системы, экосистема обеспечивает следующие функции:

1) Единые нормативы и требования к данным, которыми обмениваются цифровые платформы;

2) Единообразные схемы и шаблоны интеграции, стандартизированные интерфейсы, унифицированная архитектура и логика поведения;

3) Комплексное решение для обеспечения безопасности работы цифровых платформ;

4) Единые требования к аналитической составляющей каждой цифровой платформы, готовые унифицированные алгоритмы, шаблоны и сравнительные показатели;

5) Сервисы проверки контроля качества входящих и исходящих цифровых данных, качества применяемых инструментов и интерфейсов взаимодействия;

6) Готовые правила проектирования и развития цифровых платформ для устранения ненужной закрытости и лишней фрагментации цифровых платформ на узкие сегменты.

В основе многих современных экосистем лежит так называемое ядро. Ядро экосистемы обеспечивает минимальный набор технологических сервисов и продуктов, позволяющий экосистеме функционировать и выполнять основную задачу в случае нестабильной работы других элементов экосистемы.

Для обеспечения персонализированного подхода при взаимодействии пользователей с сервисами и продуктами экосистемы используются следующие механизмы:

1) Сквозная аутентификация и авторизация – позволяет взаимодействовать с элементами экосистемы, используя единую учётную запись пользователя и пользовательский профиль;

2) Единое хранилище данных – позволяет организовать единое пространство пользовательских данных для использования различными элементами экосистемы;

3) Событийная интеграция систем – позволяет уведомлять о действиях пользователя другие элементы экосистемы.

Основные функции экосистемы доступны через специализированные программные интерфейсы.

Среди наиболее крупных IT-компаний, использующих экосистемный подход для организации доступа к своим сервисам и платформам, можно выделить Apple, Google, Microsoft, Meta (Facebook), Amazon. В Российском сегменте можно отметить компании Яндекс, Сбер и МТС.

Одним из главных преимуществ использования цифровых экосистем для владельцев различных сервисов является возможность бесшовного перехода между сервисами экосистемы на основе анализа поведения конкретного пользователя. Таким образом, используя один из сервисов экосистемы, аудитория постепенно начинает использовать и другие сервисы, входящие в состав экосистемы.

Несмотря на то, что процесс развития экосистемы в каждом случае индивидуален в зависимости от целей и задач, которые она решает, существуют некоторые общие архитектурные принципы построения экосистем.

В основе архитектуры экосистемы лежат три блока:

- 1) Бэк-офис, состоящий из собственных сервисов компании;
- 2) Бэк-офис из сторонних систем;
- 3) Фронт-офис, включающий основные интерфейсы проекта.

Объединяет их корпоративная шина данных, которая передаёт информацию по всей экосистеме.

Использование единого протокола доступа для организации единого рабочего пространства геолога, как в случае WPS платформы для интеграции веб-сервисов, затруднительно из-за разнородности вышеперечисленных программных платформ. Однако, если использовать некоторые общие свойства платформ, такие как наличие централизованных хранилищ данных и систем единой аутентификации, появляется возможность организовать взаимодействие посредством

передачи данных от имени конкретного пользователя для анализа в хранилища данных, соответствующие платформам. Реализация подобного механизма стала возможна при использовании технологии веб-приложений, позволяющей запросить у пользователя разрешение на доступ к определённым возможностям пользовательского аккаунта различных поставщиков облачных сервисов. Такая технология поддерживается, например, Microsoft, Google, Yandex, ESRI.

Перед использованием соответствующего сервиса пользователю предлагается загрузить данные для анализа в его персональное хранилище. Для этого пользователь проходит авторизацию на сайте поставщика сервиса, после чего приложение запрашивает у пользователя разрешение на скачивание и публикацию данных в его хранилище.

Для организации подобного подхода необходимо разработать программную экосистему, которая будет осуществлять контроль данных для анализа, предоставляя пользователю инструменты для «бесшовного» перемещения данных между платформами, с отслеживанием изменений данных после работы с ними на определённой платформе.

Такая мультиплатформенная экосистема будет предоставлять пользователю возможность выбора платформы обработки для его данных, передавая на выбранную платформу данные и дальнейшее управление ими в рамках выбранной платформы [16].

На рис. 2 представлена архитектурная схема для реализации мультиплатформенной геологической экосистемы.

Экосистема состоит из следующих основных модулей:

1. GAM – модуль общей авторизации. Основная функция модуля: обеспечение единой точки аутентификации и авторизации. Модуль обращается к сервису авторизации на базе подхода Single Sign-On (SSO) с использованием протокола OAuth.

2. MICS – модуль взаимодействия с общим хранилищем данных пользователя. Основная функция модуля: обеспечение интерфейса доступа к хранилищу данных пользователя. Результаты обработки также могут быть сохранены пользователем в хранилище для их последующего использования в других сервисах. Взаимодействие с платформой осуществляется через REST интерфейс.

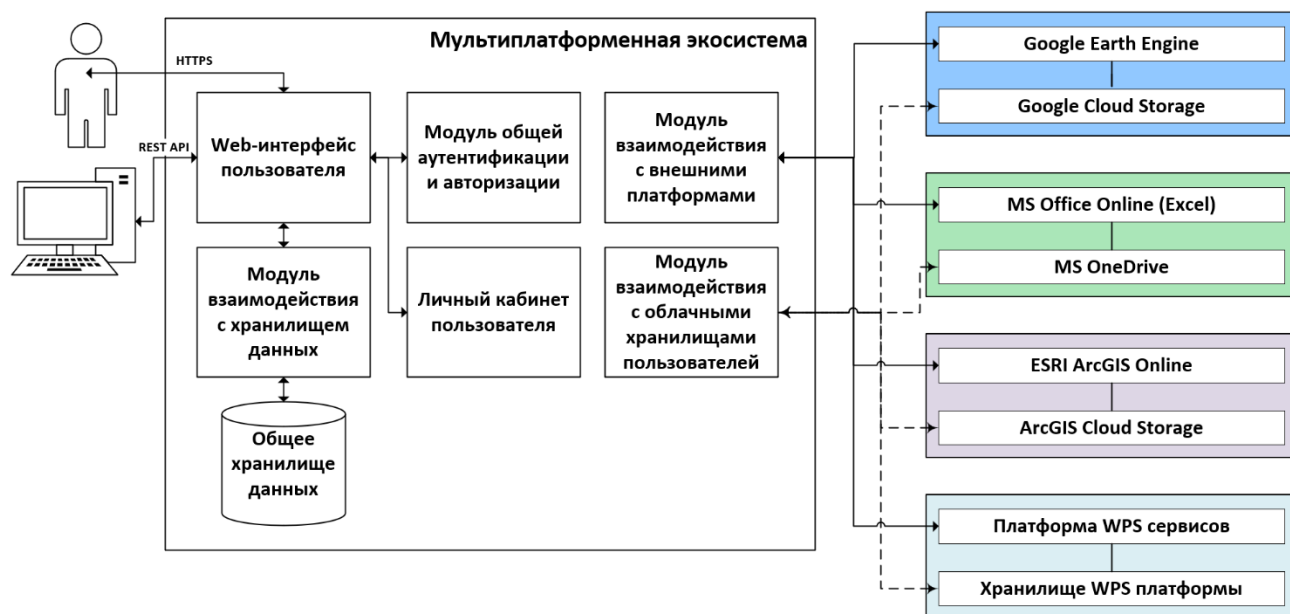


Рис. 2. Общая архитектурная схема мультиплатформенной геологической экосистемы

3. ССМ – модуль взаимодействия с внешними облачными хранилищами пользователей. Основная функция модуля: организация единого пространства данных на основе имеющихся у пользователей аккаунтов поставщиков услуг облачного хранения для дальнейшего использования данных при работе с разнородными интерактивными облачными сервисами. Поставщики облачных сервисов часто требуют предварительной загрузки данных для обработки в их собственные облачные хранилища. Модуль позволит исследователю при переходе от одного сервиса к другому в автоматическом режиме перемещать данные в требуемое хранилище. Кроме того, модуль является основным инструментом, позволяющим интегрировать разнородные интерактивные сервисы в рамках разрабатываемой платформы. Взаимодействие с платформой осуществляется через REST интерфейс.

4. SSMM – Модуль мониторинга состояния системы и сервисов. Основная функция модуля: отслеживание статуса каждого внешнего поставщика сервиса. Мониторинг состояния сервисов производится по нескольким параметрам, включая доступность узла, на котором размещён сервис, доступность сервиса по указанному протоколу и проверку сервиса на изменения путём отправки тестовых

запросов к сервису. Взаимодействие с экосистемой осуществляется через REST интерфейс.

5. РАМ – Модуль взаимодействия с внешними платформами. Основная цель модуля: передача управления данными соответствующей платформе обработки.

Подход по организации единого рабочего пространства геолога-исследователя с использованием мультиплатформенной экосистемы позволяет организовать «бесшовное» перемещение данных между платформами обработки геологических данных. Экосистема позволяет пользователям эффективно управлять геологическими данными и их визуализацией, используя при этом общедоступные программные платформы по анализу и обработке геологической информации.

5. ЕДИНОЕ ЦИФРОВОЕ ПРОСТРАНСТВО НАУЧНЫХ ГЕОЛОГИЧЕСКИХ ЗНАНИЙ

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 [17, 18].

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 результаты данного научного направления, и динамическую часть, включающую информацию о новейших достижениях в данной научной области.

В последние десятилетия происходит быстрое развитие научных информационных ресурсов и сервисов, которые образуют научное информационное пространство. Возникают их новые типы, изменяют свою форму и содержание традиционные ресурсы и способы коммуникации.

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

Примером современной научной электронной библиотеки может служить Национальная цифровая библиотека по науке США (The National Science Digital Library – NSDL¹¹), которая основана Национальным научным фондом (NSF) в 2000 г. с целью обеспечить доступ к сервисам высокого качества по многим областям науки, технологий и техники. NSDL агрегирует ресурсы из цифровых библиотек и проектов, поддерживаемых NSF, а также формирует собственные обзоры вебсайтов по своим направлениям.

В качестве примера отечественной электронной библиотеки можно привести Научную электронную библиотеку¹² содержащую метаданные о нескольких миллионах отечественных публикаций и предоставляющую коммерческий и свободный доступ к значительному количеству их полных текстов. В последние годы развивается электронная библиотека «Киберленинка»¹³, построенная на основе принципов открытой науки и содержащая более 1,3 млн полных текстов.

Предполагается, что единое цифровое пространство научных геологических знаний поможет сформировать новые процессы генерации данных и знаний. В последнее время нами осуществлялась 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 распределенной геологической информации с использованием специализированных служб её анализа и обработки [19].

Использованные подходы и модульная архитектура при проектировании Информационно-аналитической геологической среды позволяют трансформировать ее в Единое цифровое пространство научных геологических знаний для решения широкого круга задач для наук о Земле (рис. 3).

¹¹ <https://nsdl.oercommons.org>

¹² <https://elibrary.ru>

¹³ <https://cyberleninka.ru>

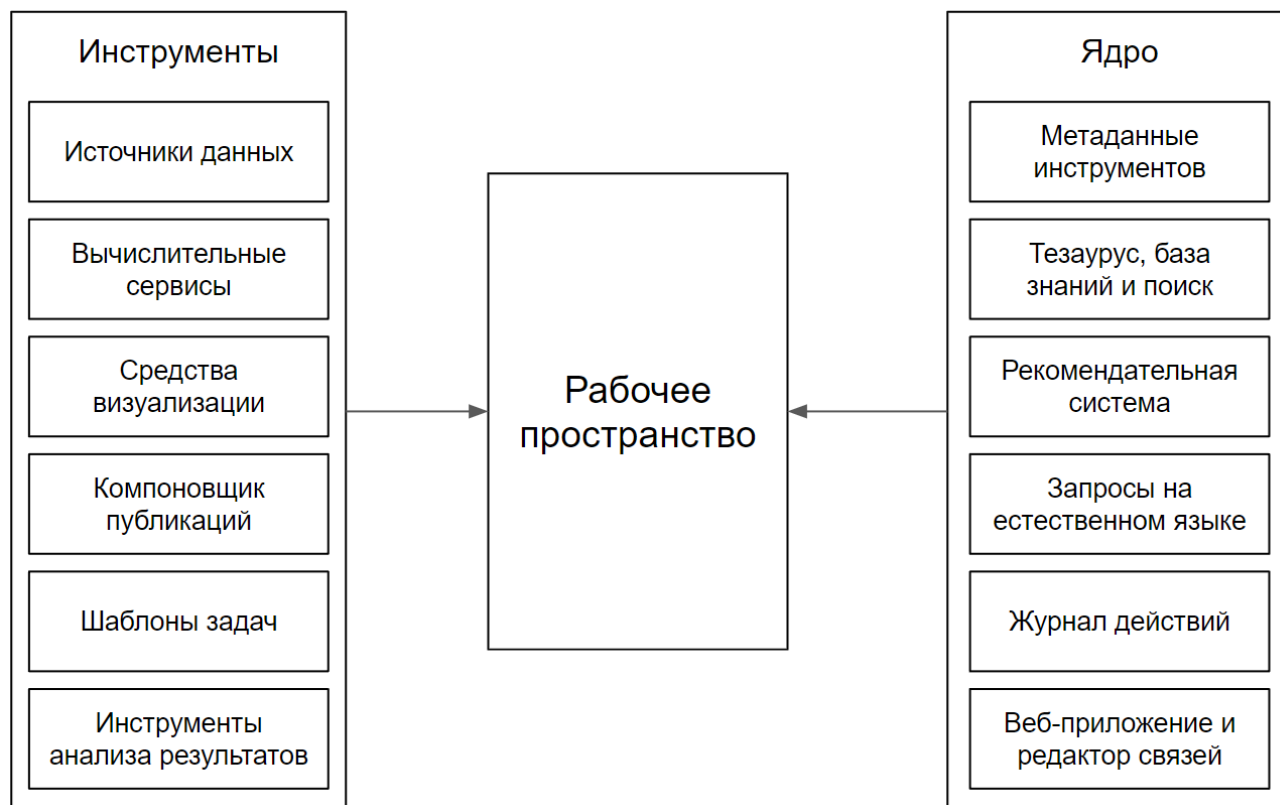


Рис. 3. Концептуальная схема Единого цифрового пространства научных геологических знаний, созданная на базовой основе Информационно-вычислительной геологической среды по поддержке научных исследований

Простота использования и гибкость персональной настройки под определённые задачи позволят привлечь к использованию этой среды 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.

6. «УМНОЕ» ЦИФРОВОЕ РАБОЧЕЕ ПРОСТРАНСТВО ПОЛЬЗОВАТЕЛЯ В ЕДИНОМ ЦИФРОВОМ ПРОСТРАНСТВЕ НАУЧНЫХ ГЕОЛОГИЧЕСКИХ ЗНАНИЙ

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 распределенных источников научных данных. Такой набор необходимо структурировать и установить взаимосвязи между данными, если это возможно, исходя из контекста задачи.

По результатам литературных обзоров, тестовых исследований и обсуждений нами предложены основные компоненты для ядра будущей системы, включающей в себя рабочие пространства для научных сотрудников в геологии.

А именно:

а) Компонент, отвечающий за «погружение» различных видов данных (файлы, таблицы, ссылки на внешние ресурсы, текстовые данные, географически привязанные данные, мультимедиа информация и т. д.) в рабочее пространство;

б) Компонент, отвечающий за загрузку данных из 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 нами осуществляется проектирование Единого цифрового пространства геологических знаний, а рабочая версия реализуется на Портале GeologyScience.ru.

Благодарности

Работы выполняются в рамках Государственного задания ГГМ РАН по теме № 0140-2019-0005 «Разработка информационной среды интеграции данных естественнонаучных музеев и сервисов их обработки для наук о Земле», а также темы государственного задания № 1021061009468-8-1.5.1 «Цифровая платформа интеграции и анализа геологических и музейных данных».

СПИСОК ЛИТЕРАТУРЫ

1. *Лаврова О.Ю., Лупян Е.А., Митягина М.И., Уваров И.А.* Информационная система See the Sea: текущие возможности и перспективы развития // "Информационные технологии в дистанционном зондировании Земли – RORSE 2018". ИКИ РАН, 2019. С. 367–373.

2. *Толпин В.А., Балашов И.В., Ефремов В.Ю., Лупян Е.А., Прошин А.А., Уваров И.А., Флитман Е.В.* Создание интерфейсов для работы с данными современных систем дистанционного мониторинга (система GEOSMIS) // Современные проблемы дистанционного зондирования Земли из космоса. 2011. Т. 8. № 3. С. 93–108.

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*. Preprint available on EarthArXiv 27, Jan. 2018.
<http://doi.org/doi:10.17605/OSF.IO/YNAXW>

4. *Klein L.J., Marianno F.J., Albrecht C.M., Freitag M., Hamann H.F.* PAIRS: A scalable geo-spatial data analytics platform, *IEEE Conference on Big Data (Big Data)* 290-1298, (2015). <http://doi.org/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, Russia, October 14–18, 2019. Published on CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073). Vol.-2527. P. 36–41.

6. *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.
<http://doi.org/10.2205/2019ES000696>

7. *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 // Труды XVII Международной конференции «Распределенные информационно-вычислительные ресурсы. Цифровые двойники и большие данные (DICR-2019)». 2019. С. 139–147.

8. *Eremenko V.S., Naumova V.V., Platonov K.A., Dyakov S.E., Eremenko A.S.* The main components of a distributed computational and analytical environment for the scientific study of geological systems // *Russian Journal of Earth Sciences*. 2018. Vol. 18, Issue 6.

9. *Zagumennov A.A., Naumova V.V.* Development of the computing node for processing satellite imagery and spatial data for earth sciences // *Proceedings of the All-Russian Conference with International Participation "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes" (SDM-2021)*. Novosibirsk, Russia, August 24–27, 2021. P. 272–279.

10. *Platonov K.A., Naumova V.V.* The Center of quantitative data on geology: current state and prospects for development // *Russian Journal of Earth Sciences*. 2020. Vol. 20. <https://doi.org/10.2205/2020ES000755>

11. *Иванов С.Д.* Интерактивный реестр геосенсоров на основе веб-приложения // *Компьютерные исследования и моделирование*. 2016. Т. 8, № 4. С. 621–632.

12. *Tkaczyk D., Szostek P., Fedoryszak M., Dendek P., Bolikowski L.* CERMINE: automatic extraction of structured metadata from scientific literature // *International Journal on Document Analysis and Recognition*. 2015. Vol. 18, No. 4. P. 317–335.

13. *Maynard D., Bontcheva K., Augenstein I.* *Synthesis Lectures on the Semantic Web: Theory and Technology* // December 2016. Vol. 6, No. 2. P. 1–194.

14. *Eremenko V.S., Naumova V.V., Zagumennov A.A., Bulov S.V.* Cloud technologies for development of geographically distributed computational and analytical Geological environment // *Computational technologies*. 2021. Vol. 26. Issue 1. P. 86–98.

15. *Ерёменко В.С., Наумова В.В.* Система каталогизации и мониторинга территориально-распределенных вычислительных узлов в среде WPS сервисов для решения геологических задач // *Вестник НГУ. Серия: Информационные технологии*. 2019. Т. 17, № 2. С. 39–48.

16. Eremenko V.S., Naumova V.V. A multi-platform ecosystem for computing in Earth sciences // Proceedings of the All-Russian Conference with International Participation "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes" (SDM-2021). Novosibirsk, Russia, August 24–27, 2021. P. 67–73.

17. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестник РАН, 2019. Т. 89, № 7. С. 728–735. <https://doi.org/10.31857/S0869-5873897728-735>

18. Единое цифровое пространство научных знаний: проблемы и решения: сборник научных трудов / под ред. Н.Е. Каленова, А.Н. Сотникова, 2021. <https://doi.org/10.23681/610687>.

19. Патук М.И., Наумова В.В., Еременко В.С. Цифровой репозиторий "GeologyScience.ru": открытый доступ к научным публикациям по геологии России // Электронные библиотеки. 2020. Т. 23, №6. <https://doi.org/10.26907/1562-5419-2020-23-6-1324-1338>

FROM AN INFORMATION AND ANALYTICAL ENVIRONMENT TO SUPPORT SCIENTIFIC RESEARCH IN GEOLOGY TO A SINGLE DIGITAL SPACE OF GEOLOGICAL SCIENTIFIC KNOWLEDGE

V. V. Naumova¹ [0000-0002-3001-1638], V. S. Eremenko² [0000-0002-5250-5743],
A. S. Eremenko³ [0000-0003-1923-8417], A. A. Zagumennov⁴ [0000-0003-1923-8417],
M. I. Patuk⁵ [0000-0003-3036-2275]

^{1, 2, 5}*State Geological Museum named after Vladimir Vernadsky, Moscow*

^{3, 4}*Institute of Automation and Control Processes FEB RAS, Vladivostok*

¹naumova_new@mail.ru, ²vitaer@gmail.com, ³academy21@gmail.com,

⁴truepikvic@gmail.com, ⁵patuk@mail.ru

Abstract

The article describes all stages of the development of an Information system for the creation, development and support of scientific research in geology: from the In-

formation and analytical environment to the Unified 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.

Keywords: *Web services, cloud services, geological data processing, geological software platform, software ecosystem for geological research, unified digital space of geological scientific knowledge.*

REFERENCES

1. *Lavrova O.Ju., Lupjan E.A., Mitjagina M.I., Uvarov I.A.* Informacionnaja sistema See the Sea: tekushhie vozmozhnosti i perspektivy razvitija // "Informacionnye tehnologii v distancionnom zondirovanii Zemli - RORSE 2018". IKI RAN, 2019. P. 367–373.
2. *Tolpin V.A., Balashov I.V., Efremov V.Ju., Lupjan E.A., Proshin A.A., Uvarov I.A., Flitman E.V.* Sozdanie interfejsov dlja raboty s dannymi sovremennyh sistem distancionnogo monitoringa (sistema GEOSMIS) // *Sovremennye problemy distancionnogo zondirovanija Zemli iz kosmosa*. 2011. Vol. 8. No 3. P. 93–108.
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*. Preprint available on EarthArXiv 27, Jan. 2018.
<http://doi.org/doi:10.17605/OSF.IO/YNAXW>
4. *Klein L.J., Marianno F.J., Albrecht C.M., Freitag M., Hamann H.F.* PAIRS: A scalable geospatial data analytics platform, *IEEE Conference on Big Data (Big Data)* 290–1298, (2015). <http://doi.org/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, Russia, October 14–18, 2019. Published on CEUR Workshop Proceedings (CEUR-WS.org, ISSN 1613-0073). Vol. 2527, P. 36–41.

6. *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.

<http://doi.org/10.2205/2019ES000696>

7. *Naumova V.V., Platonov K.A., Eremenko V.S., Patuk M.I., Dyakov S.E.* Informacionno-analiticheskaja sreda dlja podderzhki nauchnyh issledovanij v geologii: tekushhee sostojanie i perspektivy razvitija // Trudy XVII Mezhdunarodnoj konferencii «Raspredelennye informacionno-vychislitelnye resursy. Cifrovye dvojniki i bolshie dannye (DICR-2019). 2019. P. 139–147.

8. *Eremenko V.S., Naumova V.V., Platonov K.A., Dyakov S.E., Eremenko A.S.* The main components of a distributed computational and analytical environment for the scientific study of geological systems // Russian Journal of Earth Sciences. 2018. Vol. 18, Issue 6.

9. *Zagumennov A.A., Naumova V.V.* Development of the computing node for processing satellite imagery and spatial data for earth sciences // Proceedings of the All-Russian Conference with International Participation "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes" (SDM-2021). Novosibirsk, Russia, August 24–27, 2021. P. 272–279.

10. *Platonov K.A., Naumova V.V.* The Center of quantitative data on geology: current state and prospects for development // Russian Journal of Earth Sciences. 2020. Vol. 20. <https://doi.org/10.2205/2020ES000755>

11. *Ivanov S.D.* Interaktivnyj reestr geosensorov na osnove veb-prilozhenija // Kompjuternye issledovanija i modelirovanie. 2016. Vol. 8, No. 4. P. 621–632.

12. *Tkaczyk D., Szostek P., Fedoryszak M., Dendek P., Bolikowski L.* CERMINE: automatic extraction of structured metadata from scientific literature // In International Journal on Document Analysis and Recognition. 2015. Vol. 18, No. 4. P. 317–335.

13. *Maynard D., Bontcheva K., Augenstein I.* Synthesis Lectures on the Semantic Web: Theory and Technology // December 2016. Vol. 6, No. 2. P. 1–194.

14. *Eremenko V.S., Naumova V.V., Zagumennov A.A., Bulov S.V.* Cloud technologies for development of geographically distributed computational and analytical Geological environment // Computational technologies. 2021. Vol. 26. Issue 1. P. 86–98.

15. *Eremenko V.S., Naumova V.V.* Sistema katalogizacii i monitoringa territorialno raspredelennyh vychislitelnyh uzlov v srede WPS servisov dlja reshenija geologicheskikh zadach // Vestnik NGU. Serija: Informacionnye tehnologii. 2019. Vol. 17, No. 2. P. 39–48.

16. *Eremenko V.S., Naumova V.V.* A multi-platform ecosystem for computing in Earth sciences // Proceedings of the All-Russian Conference with International Participation "Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes" (SDM-2021). Novosibirsk, Russia, August 24–27, 2021. P. 67–73.

17. *Antopolskij A.B., Kalenov N.E., Serebrjakov V.A., Sotnikov A.N.* O edinom cifrovom prostranstve nauchnyh znaniy // Vestnik RAN, 2019. Vol 89, No. 7. P. 728–735. <https://doi.org/10.31857/S0869-5873897728-735>

18. Edinoe cifrovoe prostranstvo nauchnyh znaniy. Problemy i reshenija. Sbornik nauchnyh trudov. Edited by *Kalenov N.E., Sotnikov A.N.*, 2021. <https://doi.org/10.23681/610687>.

19. *Patuk M.I., Naumova V.V., Eremenko V.S.* Cifrovoy repozitorij "GeologyScience.ru": otkrytyj dostup k nauchnym publikacijam po geologii Rossii // Elektronnye biblioteki, 2020, Vol. 23, No. 6. <https://doi.org/10.26907/1562-5419-2020-23-6-1324-1338>

СВЕДЕНИЯ ОБ АВТОРАХ

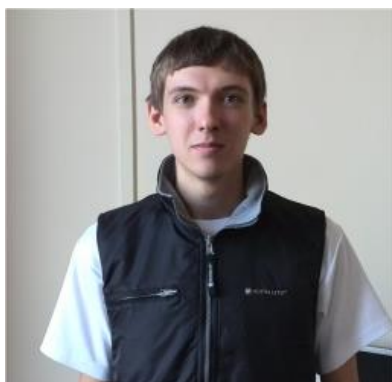


НАУМОВА Вера Викторовна – д. г.-м. н., г. н. с., зав. Научным отделом Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Vera Viktorovna NAUMOVA – Prof., head of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: naumova_new@mail.ru,

ORCID: 0000-0002-3001-1638



ЕРЁМЕНКО Виталий Сергеевич – младший научный сотрудник, Государственный геологический музей им. В.И. Вернадского РАН, Москва, Россия.

Vitaliy Sergeevich EREMENKO – Junior researcher of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: vitaer@gmail.com,

ORCID: 0000-0002-5250-5743



ЕРЕМЕНКО Александр Сергеевич – программист (внештатный сотрудник), Государственный Геологический музей им. В.И. Вернадского РАН; старший научный сотрудник, кандидат технических наук, Институт автоматизации и процессов управления ДВО РАН, Владивосток, Россия.

Aleksandr Sergeevich EREMENKO – contract programmer of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia); Ph.D, Senior researcher of Institute of automatic and control processes of FEB RAS, Vladivostok (Russia).

email: academy21@gmail.com,

ORCID: 0000-0003-1923-8417



ЗАГУМЕННОВ Алексей Андреевич – программист (внештатный сотрудник), Государственный Геологический музей им. В.И. Вернадского РАН, Москва, Россия; младший научный сотрудник, Институт автоматизации и процессов управления ДВО РАН, Владивосток, Россия.

Aleksey Andreevich ZAGUMENNOV – contract programmer of SGM scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia); Junior researcher of Institute of automatic and control processes of FEB RAS, Vladivostok (Russia).

email: truepikvic@gmail.com,

ORCID: 0000-0002-0501-5362



ПАТУК Михаил Иванович – к. г.-м. н., и. о. н. с., научный отдел Государственного геологического музея им. В.И. Вернадского РАН, Москва, Россия.

Michail Ivanovich PATUK – PhD, scientific department, Vernadsky State Geological Museum RAS, Moscow (Russia).

email: patuk@mail.ru

ORCID: 0000-0003-3036-2275

Материал поступил в редакцию 18 февраля 2022 года