

УДК 004.75

АВТОМАТИЗАЦИЯ ИНТЕГРАЦИИ СЕРВИСОВ ВЕБ-ОБРАБОТКИ ДАННЫХ ЭКОЛОГИЧЕСКОГО МОНИТОРИНГА С РАСПРЕДЕЛЁННЫМИ НАУЧНЫМИ ПРИЛОЖЕНИЯМИ

© И. В. Бычков, А. Г. Феоктистов, С. А. Горский, Р. О. Костромин,
Р. К. Фёдоров

*Институт динамики систем и теории управления им. В. М. Матросова СО РАН,
664033, Иркутск, ул. Лермонтова, 134, а/я 29
E-mail: agf@icc.ru*

Решение задач экологического мониторинга обоснованно требует осуществления сбора, оцифровки, хранения и анализа большого объёма пространственно-временных данных. Зачастую их обработка обуславливает необходимость применения распределённых вычислений. В известных геоинформационных системах средства для поддержки такого рода вычислений в полной мере, как правило, отсутствуют. Предложен подход к решению данной проблемы на основе интеграции геоинформационной системы со средствами разработки и применения распределённых научных приложений в гетерогенной вычислительной среде посредством представления этих приложений сервисами веб-обработки данных. Преимущество такого подхода заключается в снижении трудоёмкости подготовки и проведения экспериментов с геоданными.

Ключевые слова: экологический мониторинг, сервисы веб-обработки данных, геоинформационные системы, распределённые вычисления, научные рабочие процессы.

DOI: 10.15372/AUT20220407

Введение. Разработка и применение цифровых технологий исследования состояния атмосферного воздуха, поверхностных вод суши, морских вод, почвенно-земельного покрова и других важных компонентов окружающей среды являются чрезвычайно актуальными [1]. Особую роль такие технологии играют в рамках экологического мониторинга уникальных водных, земельных, лесных, биологических и других ресурсов охраняемых природных территорий [2], позволяющего в динамике осуществлять контроль текущего состояния исследуемых территорий с целью дальнейшего анализа изменений и принятия решений по их сохранению и развитию. В общем случае процесс мониторинга предполагает выбор мест и параметров наблюдения, размещение контрольно-измерительной аппаратуры, приём и передачу сигналов, обработку и анализ данных, визуализацию полученных данных и результатов их анализа, поддержку принятия решений и т. д., осуществляемых специалистами разных научных и прикладных дисциплин [3].

В настоящее время такой мониторинг базируется на использовании разнообразных специализированных информационно-вычислительных комплексов [4], интегрирующей основой которых зачастую являются геоинформационные системы (ГИС) [4]. Учитывая междисциплинарный характер исследований, выполняемых в рамках экологического мониторинга, важной задачей ГИС является обеспечение гибких и удобных возможностей для специалистов из разных предметных областей с целью обеспечения им доступа к распределённым алгоритмам, моделям, данным и датчикам для обработки геоданных [5]. Среди таких возможностей следует отметить разработку, публикацию и комплексирование веб-сервисов, включая сервис-ориентированные средства, предназначенные для трудоёмкого анализа текущей и ретроспективной информации, который основан на динамическом вы-

делении, интеграции и применении суперкомпьютерных, облачных и туманных ресурсов [6].

Организация и эффективное использование такой гетерогенной распределённой вычислительной среды (ГРВС) в рамках ГИС является актуальной и в то же время нетривиальной проблемой. В частности, требуется разработка и применение дополнительного программного обеспечения (ПО), инструментов для создания сервисов веб-обработки (Web Processing Services, WPS) больших данных и приложений для выполнения параллельных и распределённых вычислений в ГРВС [7]. Решению этих вопросов в той или иной степени посвящён ряд научных работ [8, 9]. Примеры успешной реализации WPS-сервисов для трудоёмкой обработки данных дистанционного зондирования на высокопроизводительных вычислительных системах приведены в [10, 11]. Однако в целом вышеупомянутая проблема не решена. В этой связи рассматривается новый подход к интеграции WPS-сервисов данных экологического мониторинга с распределёнными научными приложениями, расширяющими возможности ГИС по организации высокопроизводительных вычислений при подготовке и проведении крупномасштабных экспериментов на основе геоданных. Исследование направлено на развитие геопортала [12], используемого специалистами из разных предметных областей для сбора, обработки, распространения, анализа и визуализации пространственно-временных данных Байкальской природной территории (БПТ) с целью автоматизации алгоритмического обмена услугами (сервисами) между ними в рамках единой информационно-вычислительной среды.

Геопортал. Геопортал реализует типовые функции: картографическую визуализацию на основе стандарта Web Map Service (WMS) [13], доступ к каталогу пространственных данных (посредством метаданных), поиск пространственных данных, набор сервисов, обеспечивающих типовые запросы, и т. д. В дополнение к подсистемам, реализующим типовые функции, разработаны следующие средства: сервисы ввода и редактирования реляционных данных, позволяющие создавать новые пространственные данные, каталог WPS-сервисов и подсистема запуска сервисов, которые позволяют расширять функциональные возможности геопортала.

Схема работы с геоданными представлена на рис. 1. Сбор, предварительная обработка и передача информации о состоянии наблюдаемых объектов БПТ осуществляются с помощью контрольно-измерительной аппаратуры метеостанций, мониторов уровня воды, датчиков сейсмических колебаний, спутников и других технических средств. На основе наблюдений формируются файлы и базы данных разных форматов, включающие также результаты работы различных веб-сервисов. Метаданные, используемые в рамках геопортала, обеспечивают унификацию процессов поиска, получения, обработки, анализа и визуализации геоданных. Доступ пользователей к средствам геопортала осуществляется с помощью веб-интерфейса. Пользователи могут как работать с информацией в системе хранения данных (СХД), так и использовать результаты работы WPS-сервисов. Им обеспечивается возможность создания своих WPS-сервисов и их регистрации на геопортале.

В целом геопортал обеспечивает доступ к данным специалистов из разных предметных областей при решении научных и прикладных задач в рамках экологического мониторинга БПТ, в том числе проблем, связанных с актуализацией топографических и навигационных карт, сельскохозяйственным мониторингом, наблюдением за динамикой и состоянием вырубки лесов, контролем обстановки на водоёмах и другими важными вопросами. Представление и визуализация данных в процессе исследований осуществляется с помощью электронных карт, таблиц, диаграмм и графиков.

Интеграция WPS-сервисов с распределёнными приложениями. Автоматизация обработки геоданных с применением высокопроизводительных вычислений в ГРВС через WPS-сервисы осуществляется с помощью инструментального комплекса (ИК) Orlando Tools (OT) [14]. Пользователь ИК OT разрабатывает распределённое научное приложение,



Рис. 1. Общая схема обработки геоданных

схема решения задачи в котором является аналогом абстрактного научного рабочего процесса (АНРП) [15], который описывает процесс решения задачи в терминах параметров (значимых величин предметной области) и операций, выполняемых над этими параметрами и реализуемых программными модулями приложения. Концептуальные связи между модулями и операциями, а также соответствие формальных параметров модулей параметрам операций задаются в вычислительной модели приложения, формируемой с помощью её конструктора в графическом режиме или описываемой на языке XML. Инструментальный комплекс ОТ предоставляет своим пользователям специализированную подсистему (менеджера ПО) для непрерывной интеграции разрабатываемого ПО. Менеджер поддерживает работу с Git-репозиториями, а также контроль версий, отладку, тестирование, доставку и размещение ПО на ресурсах ГРВС [16]. Модули могут выполняться на любых ресурсах, характеристики которых удовлетворяют требованиям, указанным разработчиками модулей в их спецификациях при формировании вычислительной модели. Таким образом, АНРП определяет порядок обработки данных, абстрагируясь от конкретных ресурсов и версий прикладного ПО, применяемого в процессе решения задачи. Построение АНРП производится в конструкторе вычислительной модели. Планирование и выделение ресурсов для выполнения модулей АНРП, запуск заданий, балансировку вычислительной нагрузки на ресурсы и мониторинг заданий осуществляет планировщик вычислений.

Новая подсистема (конструктор WPS-сервисов) ИК ОТ предназначена для автоматизации разработки WPS-сервисов, создаваемых на основе АНРП. Конструктор поддерживает процедуры выбора АНРП приложения, его спецификации и регистрации в виде асинхронного WPS-сервиса на геопортале. Этот процесс может включать вызовы других АНРП, представленных WPS-сервисами. Схема выполнения WPS-сервиса представлена на рис. 2.

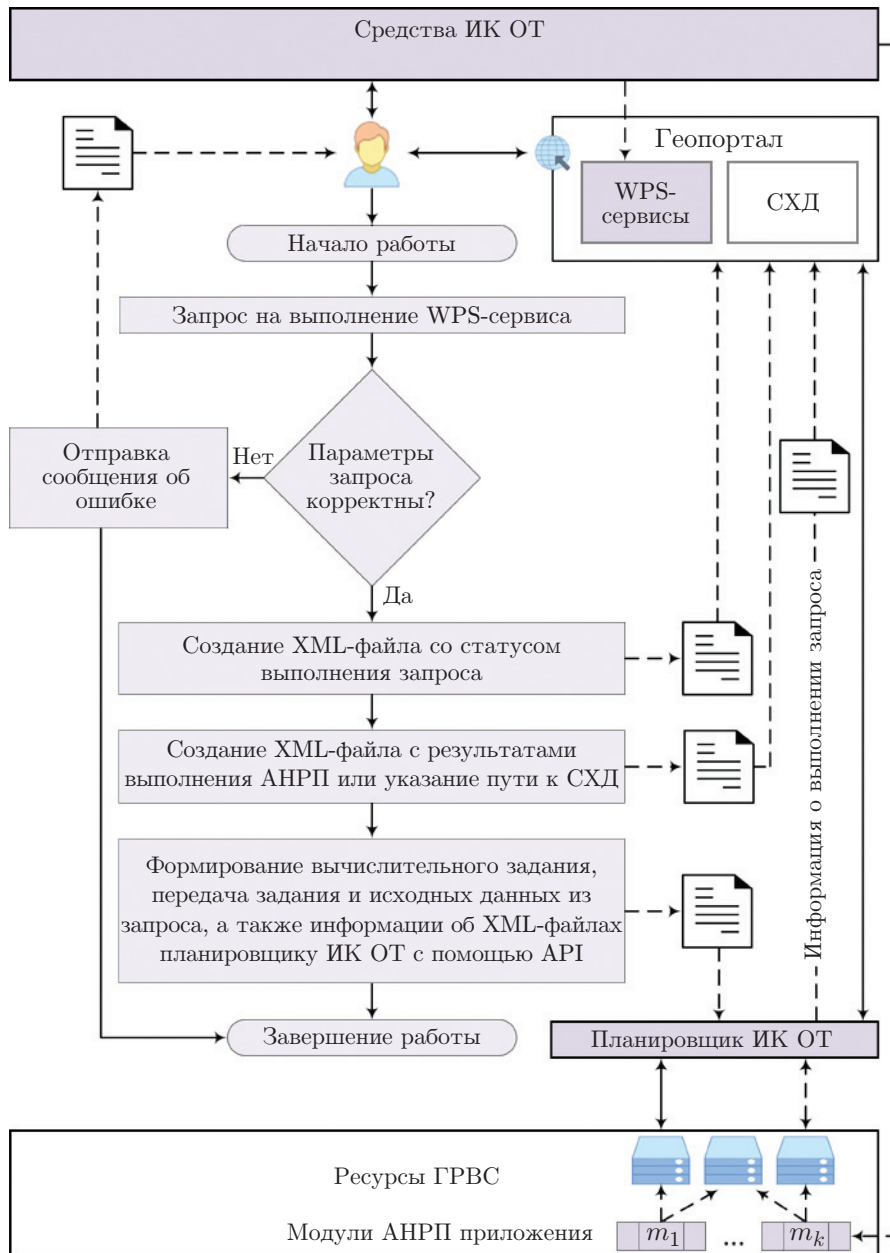


Рис. 2. Схема работы с WPS-сервисами

Пользователь обращается к WPS-сервису, в процессе выполнения которого создаются необходимые информационные структуры для отображения статуса выполнения запроса и сохранения результатов обработки данных, а также вычислительное задание для планировщика ИК ОТ. Планировщик получает вычислительное задание и исходные данные, производит декомпозицию вычислительной нагрузки по выполнению задания на подзадачи и распределяет их по ресурсам ГРВС. В дальнейшем он периодически проверяет статус выполнения подзадач, обновляет статус выполнения запроса в XML-файле на геопортале и при необходимости производит передачу данных на геопортал. При успешном завершении всех подзадач или обнаружении сбоя вычислительного процесса планировщик заканчивает обработку задания и формирует финальный статус выполнения запроса.

Сравнительный анализ. Проведено исследование известных подходов к разработке

АНРП для ГИС [17]. С точки зрения поддержки требуемых функциональных возможностей особое внимание заслуживают два программных средства: BPEL Designer Project [17], позволяющий работать с WPS-сервисами и создавать рабочие процессы на основе композиции таких сервисов, и Virtual Geographic Environments (VGE) [18] этих ГИС с возможностями создания рабочих процессов и управления ими в ГРВС. Недостатками средств является низкая степень автоматизации построения рабочих процессов и создания WPS-сервисов, а также сложность их интеграции с внешними ГИС и вычислительными системами.

Проведён сравнительный анализ функциональных возможностей ИК ОТ с возможностями BPEL Designer Project и VGE. Рассматривалась автоматизация поддержки распределённых вычислений (c_1), АНРП (c_2), регистрации АНРП на геопортале в виде WPS-сервисов (c_3) и непрерывной интеграции ПО в процессе разработки и применения приложений (c_4). Оценка E каждого программного средства выполнена следующим образом: $E = \sum_{i=1}^n e_i w_i$, где e_i — это степень автоматизации поддержки функциональной возможности c_i в рамках ГИС (1,00 — полная автоматизация, 0,75 — большая степень автоматизации, 0,50 — частичная автоматизация, 0,25 — низкая степень автоматизации, 0 — отсутствие автоматизации), а w_i — относительный вес i -й возможности с точки зрения её важности для пользователей ГИС, $\sum_{i=1}^n w_i = 1$. Значения e_i и w_i , а также оценка E , рассчитанная для каждого программного средства, приведены в таблице. Здесь значения e_i получены на основе анализа функциональных возможностей рассматриваемых программных средств. Веса w_i отражают в таблице агрегированные субъективные оценки пользователей. Проведённое сравнение программных комплексов показало преимущество ИК ОТ по ряду важных показателей.

Средство	e_1	e_2	e_3	e_4	w_1	w_2	w_3	w_4	E
BPEL Designer Project	0,75	0,75	0,75	0	0,33	0,27	0,22	0,18	0,62
VGE	1,00	1,00	0,75	0,25	0,33	0,27	0,22	0,18	0,81
ИК ОТ	1,00	1,00	1,00	1,00	0,33	0,27	0,22	0,18	1,00

Применение. Разработанный подход применён при решении задачи предсказания температуры воздуха на определённый период на основе анализа многомерного временного ряда метеорологических данных одного из районов БПТ. Прогноз температуры осуществляется путём поиска в многомерном ряду фрагментов данных (рис. 3), подобных заданному (предшествующему прогнозируемому периоду), с помощью корреляционной функции с последующей статистической обработкой найденных значений.

Наиболее трудоёмким этапом в решении задачи, требующим применения распределённых вычислений, является выбор параметров функции предсказания температуры воздуха. Настраиваемыми параметрами функции являются длины фрагментов многомерного временного ряда, предельные значения разности элементов компонентов ряда и ограничения на коэффициенты корреляции фрагментов этих компонентов. Выбор параметров производится путём генерации комбинаций их допустимых значений (множества начальных точек) с последующим применением метода мультистарта для поиска глобального минимума ошибки прогнозирования. Настройка параметров на конкретный временной ряд позволяет учесть особенности всех его компонентов, оказывающих влияние на изменение температуры воздуха, и существенно повысить точность прогноза. Для реализации такой настройки создано приложение с АНРП, зарегистрированным на геопортале в качестве WPS-сервиса.

Этот процесс описывает порядок выполнения модулей $m_1 - m_5$ приложения (рис. 4), реализующих метод мультистарта. Модуль m_1 реализует вызов WPS-сервиса для формирования многомерного временного ряда на основе данных из открытого источника [19]. Модуль m_2 захватывает и выделяет ресурсы ГРВС для экземпляров модуля m_4 на основе

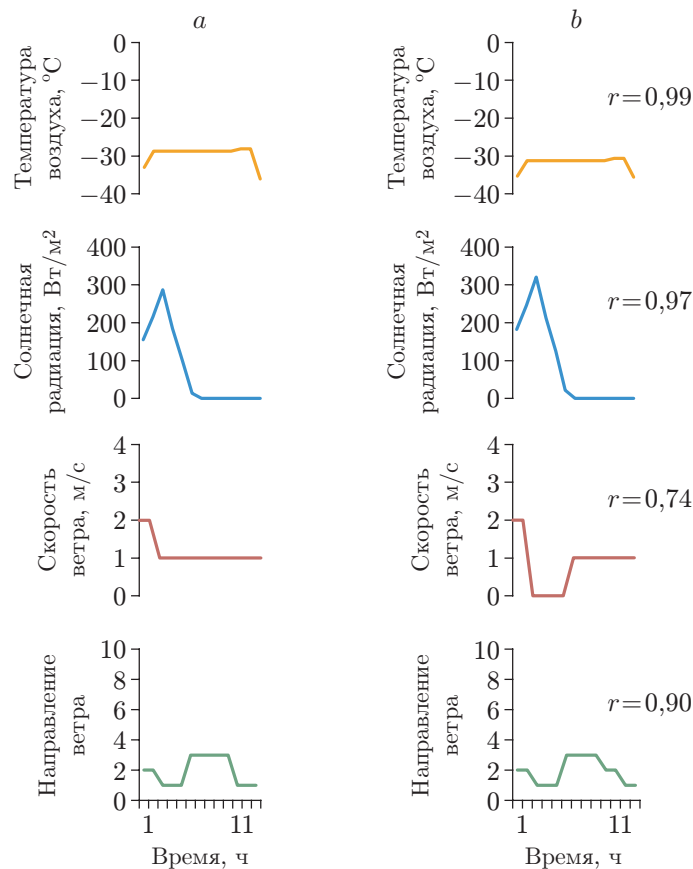


Рис. 3. Фрагменты многомерного временного ряда: заданные (а) и сравниваемые (b)

информации о выполнении данного модуля на ресурсах полученной в процессе непрерывной интеграции ПО приложения. Модуль m_3 генерирует множество начальных точек и декомпозирует его по экземплярам модуля m_4 с учётом производительности выделенных ресурсов. Каждый экземпляр модуля m_4 реализует алгоритм спуска из начальной точки в локальный минимум по методу Нелдера — Мида [20]. Модуль m_5 объединяет локальные минимумы, найденные экземплярами модуля m_4 , и возвращает минимальное значение среди найденных локальных минимумов как глобальный минимум. Координаты выбранной точки, соответствующей глобальному минимуму, применяются для настройки параметров функции.

В качестве примера выполнена настройка параметров функции предсказания температуры воздуха на ресурсах Центра коллективного пользования Иркутского суперкомпьютерного центра СО РАН (ИСКЦ) [21]. На рис. 5, а, b отражены время выполнения АНРП и ускорение вычислений соответственно, полученные при увеличении числа узлов в процессе подбора параметров функции предсказания для 500, 1000, 2000, 4000 и 8000 начальных точек. Данные, приведённые на этих рисунках, очевидно показывают масштабируемость вычислений при увеличении числа узлов. Расчёты выполнены на узлах со следующими характеристиками: 2 процессора Intel Xeon CPU X5670 (18 ядер, 2,1 ГГц), 128 Гб оперативной памяти. Накладные расходы на выполнение WPS-сервиса не превышают 5 с. WPS-сервис успешно использован для модификации сервиса имитационного моделирования природосберегающего оборудования [22], применяемого для оценки снижения CO_2 за счёт частичного использования тепловых насосов вместо угольных котлов в системах

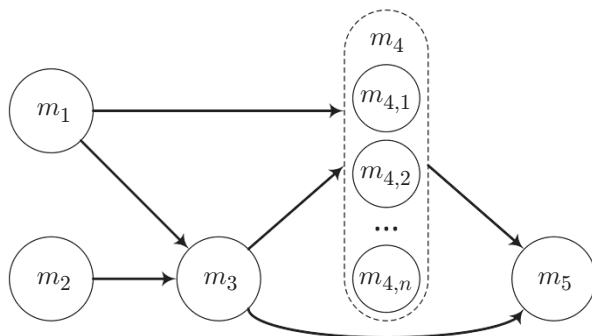


Рис. 4. Схема выполнения модулей

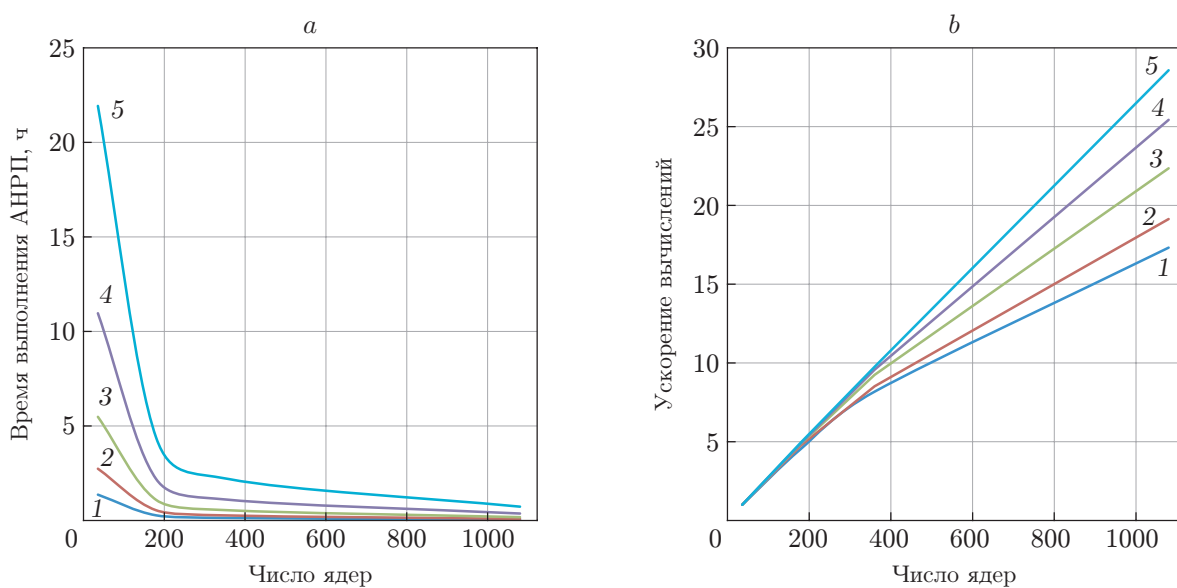


Рис. 5. Показатели выполнения АНРП: время (а) и ускорение вычислений (b) (1 — 500 начальных точек, 2 — 1000, 3 — 2000, 4 — 4000, 5 — 8000)

теплоснабжения объектов БПТ. Применение предложенной функции почасового прогноза температуры, параметры которой были скорректированы с учётом погодных условий в районе расположения моделируемых объектов, вместо прежней функции подобного назначения позволило уменьшить ошибку прогнозирования, уточнить характеристики отопительного сезона и потенциальное снижение выбросов CO_2 .

Заключение. Рассмотрен новый подход к расширению возможностей геопортала применительно к ресурсоёмкой обработке геоданных при проведении исследований в рамках экологического мониторинга БПТ путём интеграции WPS-сервисов с распределёнными научными приложениями, поддерживающими АНРП. Предложенный подход обеспечивает пользователям геопортала возможность объединения веб-сервисов, доступ к требуемым пространственно-временным данным и средства выполнения вычислений в ГРВС.

Финансирование. Работа выполнена при поддержке Министерства науки и высшего образования РФ (грант № 075-15-2020-787 на выполнение крупных научных проектов по приоритетным направлениям научно-технического развития (проект «Фундаментальные основы, методы и технологии цифрового мониторинга и прогнозирования экологической обстановки Байкальской природной территории»)).

СПИСОК ЛИТЕРАТУРЫ

1. **Makhonko N. I., Belousov S. A., Tarasova E. A. et al.** Information and communication technologies in environmental monitoring of climate change // *IOP C. Ser. Earth Environ.* 2021. **808**, N 1. P. 012045.
2. **Bychkov I. V., Ruzhnikov G. M., Hmelnov A. E. et al.** Digital monitoring of lake Baikal and its coastal area // *Proc. of the 2nd Scientific-practical Workshop on Information Technologies: Algorithms, Models, Systems (ITAMS 2019)*. Irkutsk, Russia, Sept. 20, 2019. Vol. 2463. P. 13–23.
3. **Якубайлик О. Э., Кадочников А. А., Токарев А. В.** Геоинформационная веб-система и приборно-измерительное обеспечение оперативной оценки загрязнения атмосферы // *Автометрия*. 2018. **54**, № 3. С. 39–46. DOI: 10.15372/AUT20180305.
4. **Paul P., Aithal P. S., Bhuimali A. et al.** Geo information systems and remote sensing: Applications in environmental systems and management // *Int. Journ. Manag. Technol. Soc. Sci.* 2020. **5**, N 2. P. 11–18.
5. **Zhang M., Jiang L., Zhao J. et al.** Coupling OGC WPS and W3C PROV for provenance-aware geoprocessing workflows // *Comput. Geosci.* 2020. **138**. P. 104419.
6. **Breunig M., Bradley P. E., Jahn M. et al.** Geospatial data management research: Progress and future directions // *ISPRS Int. Journ. Geo-Inf.* 2020. **9**, N 2. P. 95.
7. **Baranski B.** Grid computing enabled web processing service // *Proc. of the 6th Geographic Information Days. IfGI prints.* 2008. **32**. P. 243–256.
8. **Ulutaş Karakol D., Cömert Ç.** Architecture for semantic web service composition in spatial data infrastructures // *Surv. Rev.* 2020. **54**, N 382. P. 1–16.
9. **GeoServer.** URL: <http://geoserver.org/> (дата обращения: 25.04.2022).
10. **Шокин Ю. И., Пестунов И. А., Смирнов В. В. и др.** Корпоративная информационная система СО РАН для сбора, хранения и обработки спутниковых и наземных данных // *Горный информационно-аналитический бюллетень*. 2009. **17**, № 12. С. 9–15.
11. **Bychkov I. V., Ruzhnikov G. M., Fedorov R. K. et al.** Classification of Sentinel-2 satellite images of the Baikal natural territory // *Comp. Opt.* 2022. **46**, N 1. P. 90–96.
12. **Bychkov I. V., Ruzhnikov G. M., Fedorov R. K. et al.** Digital environmental monitoring technology Baikal natural territory // *Proc. of the 3rd Workshop on Information Technologies: Algorithms, Models, Systems (ITAMS 2020)*. Irkutsk, Russia, Sept. 3, 2020. Vol. 2677. P. 1–7.
13. **The OpenGIS Web Processing Service.** URL: <https://www.ogc.org/standards/wps/> (дата обращения: 25.04.2022).
14. **Tchernykh A., Bychkov I., Feoktistov A. et al.** Mitigating uncertainty in developing and applying scientific applications in an integrated computing environment // *Program. Comput. Soft.+.* 2020. **46**, N 8. P. 483–502.
15. **Belhajjame K., Vargas-Solar G., Collet C.** A flexible workflow model for process-oriented applications // *Proc. of the Second Intern. Conf. on Web Information Systems Engineering*. Kyoto, Japan, Dec. 3-6, 2001. **1**. P. 72–80.
16. **Gorsky S., Kostromin R., Feoktistov A. et al.** Orlando tools: Supporting high-performance computing in distributed environments // *Proc. of the 6th Intern. Conf. on Information Technology and Nanotechnology (ITNT 2020)*. Samara, Russia, May 26-29, 2020. P. 1–6.
17. **Tan X., Jiao J., Chen N. et al.** Geoscience model service integrated workflow for rainstorm waterlogging analysis // *Int. Journ. Digit. Earth.* 2021. **14**, N 7. P. 851–873.
18. **Li H., Zhang C., Xiao Z. et al.** A Web-based geo-simulation approach integrating knowledge graph and model-services // *Environ. Modelling and Software.* 2021. **144**. P. 105160.
19. **rp5.ru.** Расписание погоды. Погода в 243 странах мира. URL: <https://rp5.ru/> (дата обращения: 25.04.2022).

-
20. **Nelder J. A., Mead R.** A simplex method for function minimization // Comput. Journ. 1965. **7**. P. 308–313.
 21. **ИСКЦ СО РАН.** URL: <https://hpc.icc.ru/> (дата обращения: 25.04.2022).
 22. **Kostromin R., Basharina O., Feoktistov A. et al.** Microservice-based approach to simulating environmentally-friendly equipment of infrastructure objects taking into account meteorological data // Atmosphere. 2021. **12**, N 9. P. 1217.

Поступила в редакцию 16.05.2022

После доработки 18.05.2022

Принята к публикации 20.05.2022
