

INFORMATION-ANALYTICAL ENVIRONMENT SUPPORTING INTERDISCIPLINARY RESEARCH OF NATURAL RESOURCES IN THE BAIKAL REGION

Igor V. Bychkov⁽¹⁾, Roman K. Fedorov⁽¹⁾, Yurii V. Avramenko⁽¹⁾, Alexander S. Shumilov⁽¹⁾, Alexei O. Shigarov⁽¹⁾, Alla V. Verhozina⁽²⁾, Natalia V. Emelyanova⁽³⁾, Andrei A. Sorokovoi⁽³⁾, Gennady M. Ruzhnikov⁽⁴⁾

⁽¹⁾ Matrosov Institute for System Dynamics and Control Theory SB RAS, Irkutsk, Russia

⁽²⁾ Siberian Institute of Plant Physiology and Biochemistry SB RAS, Irkutsk, Russia

⁽³⁾ V.B. Sochava Institute of Geography SB RAS, Irkutsk, Russia

⁽⁴⁾ Irkutsk Scientific Center SB RAS, Irkutsk, Russia

The technology of creating a geoportal information-analytical environment for supporting scientific research of nature management in the Baikal region is considered on the basis of a service-oriented paradigm and the integrated use of thematic and spatial data accumulated by the scientific community.

Keywords: information-analytical environment, services, geoportal, spatial data, nature management.

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СРЕДА ПОДДЕРЖКИ МЕЖДИСЦИ- ПЛИНАРНЫХ ИССЛЕДОВАНИЙ ПРИРОДОПОЛЬЗОВАНИЯ В БАЙКАЛЬСКОМ РЕГИОНЕ

И.В. Бычков⁽¹⁾, Р.К. Фёдоров⁽¹⁾, Ю.В. Авраменко⁽¹⁾, А.С. Шумилов⁽¹⁾, А.О. Шигаров⁽¹⁾,
А.В. Верхозина⁽²⁾, Н.В. Емельянова⁽³⁾, А.А. Сороковой⁽³⁾, Г.М. Ружников⁽⁴⁾

⁽¹⁾ Институт динамики систем и теории управления им. В.М. Матросова СО РАН, г. Иркутск

⁽²⁾ Сибирский институт физиологии и биохимии растений СО РАН, г. Иркутск

⁽³⁾ Институт географии им. В.Б. Сочавы СО РАН, г. Иркутск

⁽⁴⁾ Иркутский научный центр СО РАН, г. Иркутск

Рассмотрена технология создания геопортальной информационно-аналитической среды поддержки научных исследований природопользования в Байкальском регионе на основе сервис-ориентированной парадигмы и комплексного использования тематических и пространственных данных, накопленных научным сообществом

Ключевые слова: информационно-аналитическая среда, сервисы, геопортал, пространственные данные, природопользование.

Введение. Природопользование выступает объектом междисциплинарных научных исследований, так как представляет собой совокупность всех форм взаимодействия общества и природы, объединяющая не только процессы использования человеком природных условий и ресурсов, антропогенное воздействие на природу и ее изменение этим воздействием, но и воспроизводство природной среды. При рациональном природопользовании осуществляется максимально полное удовлетворение потребностей в материальных благах при сохранении экологического баланса и возможностей восстановления природно-ресурсного потенциала.

В области природопользования и охраны окружающей среды акценты ставятся на:

- решение проблем развития хозяйственного комплекса государства, при котором учитываются экологические и природно-географические условия конкретных территорий;
- достижение на каждой конкретной территории качества среды обитания, отвечающего системе оценок генетического здоровья населения;
- восстановление и сохранение биосферного равновесия, генетического фонда животного мира;
- рациональное использование всего природно-ресурсного потенциала.

Важной составляющей современного природопользования является его мониторинг, позволяющий развивать систему управления процессами природопользования. В свою очередь, интеграция, обработка пространственно-временных данных мониторинга и их комплексный анализ помогают формировать новые решения в данной области.

Научные исследования состояния и динамики природных экосистем Байкальского региона и их компонентов, активно развивающиеся в последние годы, базируются на базах данных (БД), знаниях и сервисах, локализованных в институтах СО РАН, вузах, территориальных органах власти и управления. Открытый информационный обмен поддержки природопользования в регионе не налажен, что затрудняет проведение комплексных междисциплинарных исследований и внедрение их результатов в управлении территориальным развитием.

Современное развитие информационных технологий, увеличение скорости передачи данных, развитие и стандартизация программного интерфейса браузеров и сервисов, создание центров хранения данных (ЦОД), развитие методологии «облачных вычислений» позволяют переносить информационные-аналитические системы в сеть Интернет и реализовывать открытость, масштабируемость, предоставление общих классификаторов, доступность данных и сервисов их обработки и т.д. [1, 6]

Разработка открытой информационно-аналитической среды (ИАС) поддержки междисциплинарных исследований природопользования в Байкальском регионе на основе современных информационных технологий актуальна для отработки интеграции и комплексного использования тематических и пространственных данных, накопленных научным сообществом по проблемам природопользования.

ИАС, как среда интеграции, хранения, распределённой обработки сервисами WPS и комплексного использования данных должна обеспечивать:

- создание тематических и пространственных данных природопользования по методологии краудсорсинга (crowdsourcing, crowd — «толпа» и sourcing — «использование ресурсов») по аналогии с проектами Википедия, NASA Clickworkers, eBird, Peer-to-Patent и т.д., но в отличие от них данные ИАС должны создаваться в реляционном, нормализованном виде, ориентированном на автоматическую обработку;
- создание прототипа Web-сервиса ETL интеграции неструктурированных табличных данных;
- хранение данных в центрах хранения и обработки данных (ЦОД), что обеспечит надежность, онлайн доступ с любого устройства, подключенного к Интернет;
- регламентированная работа с данными в целях сохранения интеллектуальной собственности;
- обмен данными и их объединение в общие ресурсы;

- представление данных ИАС, а также распределенная их обработка на основе WPS-сервисов и их комбинаций [3].

Разработка ИАС позволяет создать единое информационное пространство поддержки комплексных междисциплинарных исследований природных экосистем для получения новых знаний и их внедрения в управление территориальным развитием региона.

Основным объектом для апробации компонентов среды является исследование трансформации флоры территорий с изменением уровня урбанизации и разработка технологии оценки пространственного и временного изменения биоразнообразия от различных факторов. Среда распределенных («облачных») вычислений обеспечивает как ввод и хранение пользовательских данных о биоразнообразии Байкальского региона, так и выполнение анализа этих данных с помощью WPS сервисов.

Для частичного наполнения ИАС проведены исследования по следующим направлениям:

- оценка интеграционных тенденций формирования транспортной системы Байкальского региона с использованием сервисов пространственного анализа [7];
- прогнозирование с использованием тематических сервисов изменения флористического состава Байкальского региона в зависимости от социально-экономического развития региона [4];
- разработка методов и средств визуализации карт для отображения динамики изменения состояния Байкальского региона [2];
- разработка тематических карт основных водных объектов территории и пунктов мониторинга количественных и качественных характеристик водных ресурсов [2];
- создание серии социально-экономических карт развития городов и урбанизации [5, 8, 9]:
 - пространственно-временной динамики региональной системы городов;
 - дифференциации городов по уровням и типам социально-демографического развития.

Информационно-аналитическая среда мониторинга объектов природопользования

Сервис-ориентированная парадигма, стандарты OGC позволяют проводить на новом уровне междисциплинарные научные исследования природопользования Байкальского региона. Основным компонентом ИАС служит типовой геопортал, который, как точка доступа, поддерживает обмен пространственной информацией между пользователями (исследователями), а также передаёт в специальном виде данные о дизайне и способах предоставления информации (рис. 1) [1, 6].

Основные компонентами типового геопортала:

1. Каталоги данных и сервисов, подсистемы запуска сервисов и отображения данных.
2. Система ввода и редактирования реляционных данных с пространственными атрибутами, обеспечивающая доступ к данным сервисов в рамках стандарта WPS.

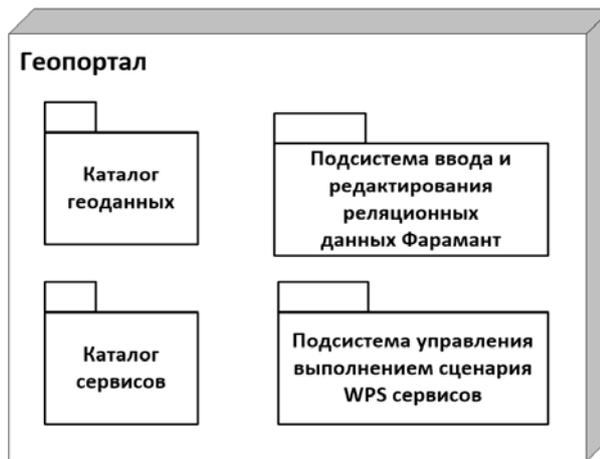


Рис. 1. Структура типового геопортала

В свою очередь, междисциплинарность задач природопользования аргументирует необходимость создания кластера тематических геопорталов, поддерживающих между собой обмен пространственными и тематическими данными и Web-сервисами их обработки на основе стандартов OGC, что повышает балансировку нагрузки сети и упрощает работу пользователей.

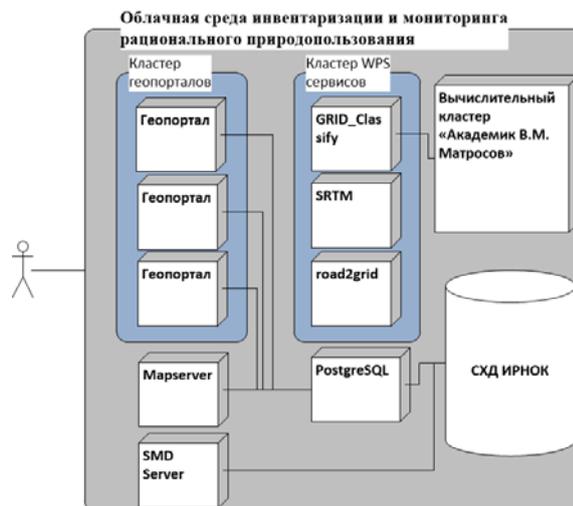


Рис. 2. Архитектура облачной среды мониторинга природопользования

В кластере геопорталов облачной инфраструктуры для быстрого разворачивания созданы шаблон нового экземпляра геопортала и кластер виртуальных машин на разных платформах для разворачивания WPS-сервисов. Для хранения пространственных данных используется СУБД PostgreSQL. Применение общей СУБД для всех геопорталов кластера позволяет упростить обмен данными и использовать в запросах различные таблицы пользователей. Mapserver и SMD Server необходимы для отображения пространственных данных в соответствии со стандартом WMS. Mapserver производит отображение постоянно изменяющихся данных, находящихся в СУБД PostgreSQL. По сравнению с другими системами, SMD Server проводит отображение статичных данных на высокой скорости без использования систем кэширования.

Для комплексного анализа данных или применения сервисов обработки проводится унификация за счет дополнительной информации (структурных спецификаций) о таблицах. Для автоматизации поиска сервисов и данных разработана спецификация в формате JSON, которая определяет входные данные сервисов, описывает структуры реляционных таблиц. Данная спецификация может конвертироваться в необходимые стандарты (XSD, WFS, WSDL-S). Кроме стандартных полей названия набора данных и описания полей (название, тип данных) структурные спецификации содержат:

- метаданные в соответствии со стандартом ISO 19115;
- единицы измерения для числовых данных;
- возможные значения для перечислимых типов данных;
- элементы управления.

Элементы управления реализуют методы необходимые для формирования пользовательского интерфейса добавления, редактирования и отображения данных атрибута.

Структурные спецификации используются пользователем для создания таблиц, поиска соответствующих сервисам данных и хранятся в специализированном каталоге и упорядочиваются в виде иерархий. Разработан механизм конвертации реляционных данных в формат DBF или SHAPE (при наличии пространственных атрибутов) для применения их в качестве входных данных WFS сервисов. В отличие от существующих, данные технологии и компоненты используют структурные спецификации представления и обработки информации, что позволяет гибко их применять для решения широкого круга проблем.

В ИАС реализованы задачи формирования, интеграции и передачи данных, запуска сервисов, организации вычислительного процесса в облачной инфраструктуре, представления результатов и т.д.

Средства визуализации карт для отображения динамики изменения состояния Байкальского региона

В рамках сервисов ввода и редактирования реляционных данных разработан инструментарий задания картографического отображения слоев геопортала на основе стандарта WMS. Редактор стилей использует метаданные каталоги пространственных данных и картографических символов (название, отображение, ключевые слова и описание). Для поиска используются атрибуты ключевые слова и описание. Отображение символа задается в формате Mapserver. Реализован модуль, позволяющий отображать символы в браузере на canvas.

Редактор стилей для каждого слоя позволяет задавать несколько классов отображения по значениям атрибутов (Рис.3). Для каждого класса можно определить свое отображение, используя символы из каталога картографических символов. Чтобы избежать задержек SLD стили транслируются в файл настройки отображения Mapserver [1, 6].

В рамках геопортала реализована возможность включения стилей отображения данных в набор метаданных сервисов, что позволяет автоматически использовать их, как для отображения параметров сервисов, так и результатов работы сервисов. Разработано несколько классификаторов, позволяющих на основе набора атрибутов автоматически формировать стили отображения объектов.

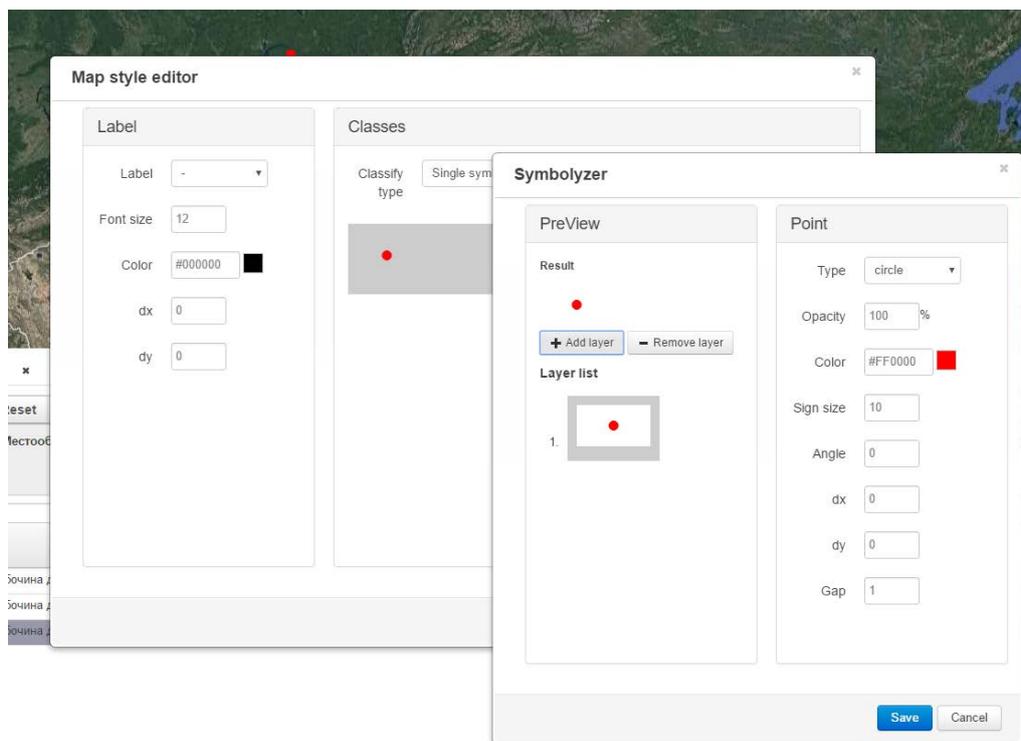


Рис.3. Редактор стилей

Web-сервис ETL интеграции неструктурированных табличных данных

Web-сервис интеграции неструктурированных табличных данных основан на правилах анализа и интерпретации произвольных размеченных таблиц, содержащих данные о природопользовании Байкальского региона, которые собираются и накапливаются в неструктурированном виде в электронных таблицах с произвольной структурой [1, 6].

Реализован прототип RESTful Web-сервиса ETL-интеграции неструктурированных табличных данных (RWS-SSDC, RESTful Web Service for SpreadSheet Data Canonicalization). Web-сервис обеспечивает: загрузку исходных табличных данных (произвольных электронных таблиц, CRL-программ, YAML-спецификаций категорий) на сервер; настройку и запуск системы трансформации данных из произвольных электронных таблиц Excel на основе исполнения CRL-программ; генерацию выходных JSON данных (исходную произвольную таблицу, восстановленные семантических отношений, выходную каноническую таблицу); сохранение на сервере выходных данных и предоставление доступа к ним (рис. 4.). Реализация Web-сервиса выполнена в виде сервлета для платформы исполнения Java EE.

В системе трансформации данных электронных таблиц разбор и загрузка исходных данных производится с помощью свободной системы Apache POI. В результате формируются объекты, представляющие таблицу, ее ячейки, строки и столбцы. Они добавляются как факты в рабочую память системы исполнения правил. Опционально одна или несколько категорий, описанных с помощью языка YAML, могут загружаться в систему, также формируя соответствующие факты в рабочей памяти Drools. При этом синтаксический разбор YAML описаний выполняется в свободной системе SnakeYAML. Структуры данных, представляющие ячейки, вхождения, метки и категории, реализованы как Java-классы в соответствии с соглашениями спецификации JavaBeans.

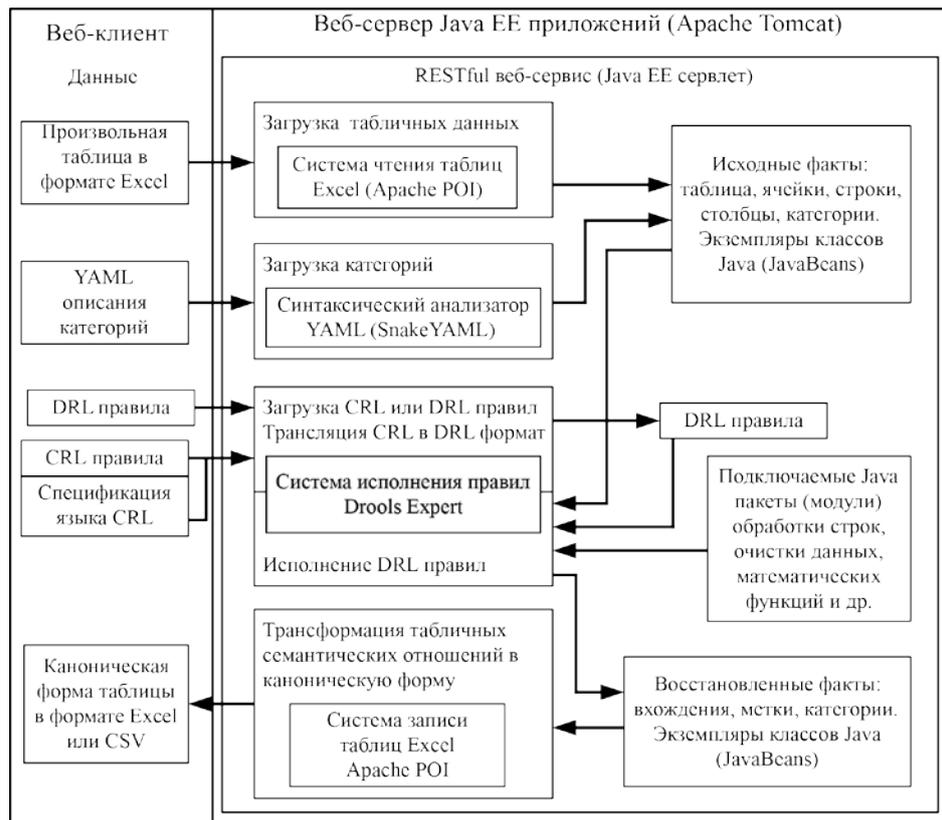


Рис. 4. Архитектура RESTful Web-сервиса ETL-интеграции неструктурированных данных из произвольных электронных таблиц

При этом каждый факт является экземпляром одного из этих классов. Правила анализа и интерпретации таблиц выражаются на специальном языке CRL. В качестве системы исполнения правил используется программное обеспечение Drools. Факты, загруженные в рабочую память, сопоставляются скомпилированным правилам. В результате создаются новые факты, описывающие семантику таблицы: вхождения, метки, категории и отношения между ними. Из них генерируется каноническая таблица.

Web-клиент доступа к RWS-SSDC обеспечивает управление сервисом, визуализацию и интерактивную работу с получаемыми от него выходными JSON данными. Web-клиент реализован в виде JavaScript приложения с помощью платформы Bootstrap (рис. 5).

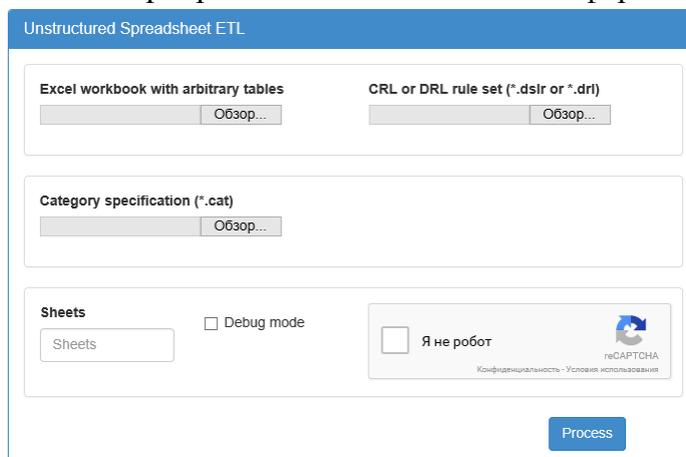


Table 0						Table 1						Table 2						Table 3						Download					
Source table												Source table																	
			A		B							Categories table																	
			C	D	E	F	Labels table																						
G	I	1	2	3	4	Canonical table																							
	J	5	5	7	8	DATA	GROUP1	GROUP2	GROUP3	GROUP4																			
H	K	9	10	11	12	1	g	c	i	a																			
Categories table												2	g	d	i	a													
Category	Labels	3	g	e	i	b	4	g	f	i	b																		
GROUP1	g,h	5	g	c	j	a	5	g	d	j	a																		
GROUP2	c,d,f,e	7	g	e	j	b	8	g	f	j	b																		
GROUP3	k,l,j	9	h	c	k	a	10	h	d	k	a																		
GROUP4	a,b	11	h	e	k	b	12	h	f	k	b																		
Labels table																													
Label	Parent	Category																											
a	null	GROUP4																											

Рис. 5. Пользовательский интерфейс Web-интерфейс

Разработанный сервис использовался для экспорта демографических данных из статистических отчетов в PDF формате геопортал.

Сервис пространственного анализа и оценки тенденций формирования транспортной системы Байкальского региона

Для оценки влияния транспортной системы на природопользование в Байкальском регионе создан WPS-сервис построения карт транспортной доступности. Исследуемая территория делится регулярной сеткой. В каждой ячейке вычисляется время движения в эту точку от ближайшего по времени муниципального образования. Предполагается, что вначале движение производится автомобильным транспортом по дорожной сети, а затем (если необходимо) пешком. На вход сервиса передаются следующие данные: границы муниципальных образований в виде полигонов, сеть автодорог в виде набора полилиний, экстенд создаваемой карты и размер ячейки. На выходе сервиса получаем регулярную сетку в формате GeoTIFF (Рис. 6).

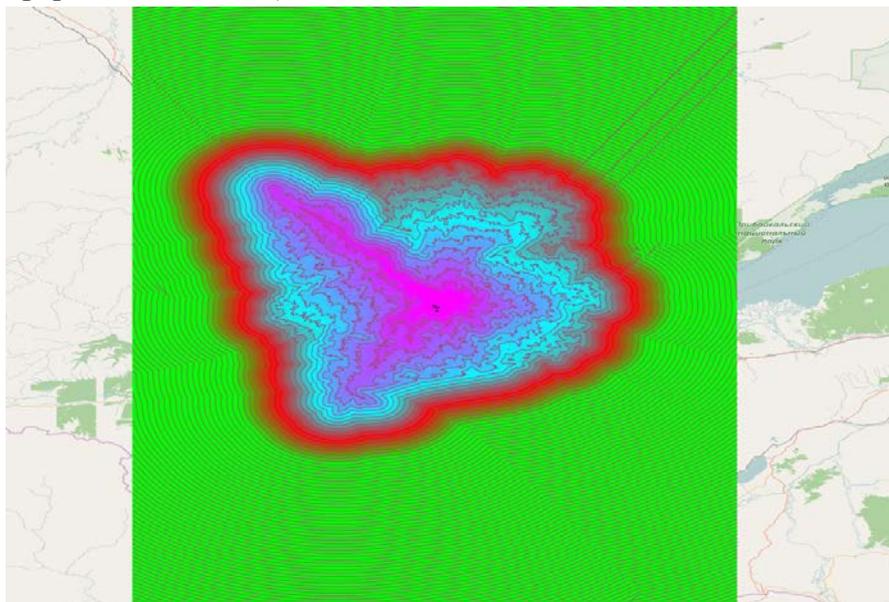


Рис. 6. Изолинии транспортной доступности с шагом 30 минут

Сервис мониторинга изменения флористического состава Байкальского региона в зависимости от социально-экономического развития региона

Изменение флористического состава региона возможно в результате двух процессов – исчезновения редких аборигенных видов и внедрения чужеродных адвентивных видов (биологическое загрязнение). Наиболее агрессивные из чужеродных видов – инвазионные, способны вытеснять или препятствовать возобновлению видов природной, аборигенной флоры, а также наносить ущерб сельскому хозяйству и здоровью населения.

Для тестирования WPS-сервиса мониторинга изменения флористического состава выбраны виды первой категории агрессивности: наземный вид ячменя гривастого (*Hordeum jubatum*), водный вид элодеи канадской (*Elodea canadensis*). В настоящее время WPS-сервис ИАС использует базу данных флоры региона Сибирского института физиологии и биохимии растений СО РАН (66 тысяч учетных записей). Созданы карты по всем видам, признанным экспертами инвазионными – еще 17 видов, и потенциально инвазионными – еще 92 вида.

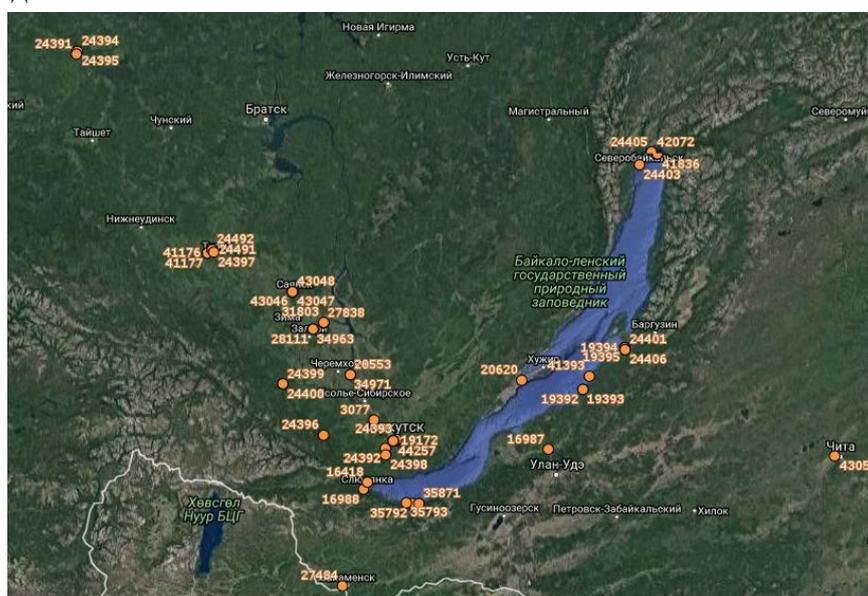


Рис. 7. Карта распространения вида «ячмень гривастый»

Работа сервиса осуществляется в два этапа: формируется регулярная сеть размером 20 x 30 км, в ячейках которой вычисляется количество видов; выделяются ячейки с количеством видов больше заданного порога.

Также при тестировании ИАС частично заполнены базы тематических пространственных данных и получены карты основных водных объектов и пунктов мониторинга характеристик водных ресурсов, а также социально-экономического развития городов и урбанизации муниципальных образований:

1. Карта «Гидрологические посты на территории российской части бассейна озера Байкал»
2. Карта «Дифференциация городов Байкальского региона по типам социально-демографического развития».
3. Картограмма «Численность постоянного населения Байкальского региона по компонентам изменения».

4. Карта «Степень урбанизированности муниципальных образований Байкальского региона».

5. Карта «Показатели экологической нагрузки и природоохранные расходы в субъектах в субъектах Байкальского региона».

Заключение. Созданы основы информационно-аналитической среды поддержки междисциплинарных исследований природопользования в Байкальском регионе. Ее особенностью является открытость, распределённость, широкое применение методов обработки тематических и пространственных данных по природопользованию Байкальском регионе в виде Web-сервисов (в соответствии со стандартом WPS), предоставляющих услуги обработки растровых и векторных данных, что позволяет расширить набор функций тематического и пространственного анализа.

Работа выполнена при финансовой поддержке: РФФИ (гранты №№: 18-07-00758-а, 17-57-44006-монг_а, 17-47-380007-р-а, 16-07-00411-а, 16-07-00554-а); Программы Президиума РАН №27; Интеграционных программ СО РАН, ИНЦ СО РАН). Результаты получены при использовании ЦКП «Интегрированная информационно-вычислительная сеть Иркутского научно-образовательного комплекса» (<http://net.icc.ru>).

ЛИТЕРАТУРА

- [1] Инфраструктура информационных ресурсов и технологии создания информационно-аналитических систем территориального управления / И.В. Бычков, Г.М. Ружников, А.Е. Хмельнов, Р.К. Фёдоров, В.В. Парамонов, А.О. Шигаров, Е.С. Фереферов, А.С. Гаченко, А.А. Михайлов, А.С. Шумилов, Ю.В. Авраменко, Т.И. Маджара / Изд-во СО РАН, 2016, 242 с.
- [2] Кузьмин С.Б., Невзорова И.В., Черкашин Е.А., Шаманова С.И. Геоинформационное картографирование на основе модели пластики рельефа и возможность ее использования при геоморфологическом анализе // Геоинформатика, 2016. № 2. С. 19-34.
- [3] Фёдоров Р. К., Шумилов А. С. Задание графа зависимостей для композиций сервисов с помощью JavaScript сценариев // География и природные ресурсы. — 2016. №6 – Спец. Выпуск. С. 160-163.
- [4] Верхозина А.В. Информационно-аналитическая система по фиторазнообразию Байкальской Сибири / А.В. Верхозина, Р.К. Федоров, С.Г. Казановский, А.С. Шумилов, Д.А. Кривенко, В.В. Мурашко // Известия Иркутского государственного университета. Серия «Биология. Экология». – 2016. – Т. 9. №3. С. 9-16.
- [5] Воробьев Н.В., Емельянова Н.В., Воробьев А.Н., Валева О.В. Расселение, динамика населения и человеческий потенциал Прибайкалья // География и природные ресурсы. – 2016. - Спец. выпуск– С. 168-178
- [6] Интеграция информационно-аналитических ресурсов и обработка пространственных данных в задачах управления территориальным развитием / И. В. Бычков, Г. М. Ружников, А. Е. Хмельнов. и др.; под ред. И. В. Бычкова - Новосибирск: Изд. СО РАН, 2012. – 369 с.
- [7] Дугарова Г., Богданов В. Анализ современного состояния придорожного сервиса в Иркутской области // ЭКО. - 2015. - №2. -С.117-126.

- [8] Воробьев Н.В. Емельянова Н.В. Вопросы разработки концепции развития Иркутской агломерации в контексте развития Прибайкалья // Материалы Всероссийской с международным участием научно-практической конференции, посвященной Всемирному дню Земли и 60-летию кафедры экономической географии КГПУ им. В.П. Астафьева. – Красноярск: Издательство КГПУ, 2015 – 4 с.
- [9] Емельянова Н.В., Сороковой А.А. Геоинформационный подход к делимитации Иркутской городской агломерации // Внеэкономические факторы пространственного развития / Под ред. В.Н. Стрелецкого и др. - М.: Эслан, 2015 – С 316-330