

**СЕКЦИЯ №4. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
И МОДЕЛИРОВАНИЕ**

УДК 631.4; 004.75

О.М. Голозубов, В.М. Колесникова

**ПРИМЕНЕНИЕ МЕХАНИЗМА БОЛЬШИХ ДАННЫХ
В АГРО-ЭКОЛОГИЧЕСКОМ МОНИТОРИНГЕ**

Факультет почвоведения МГУ имени М.В.Ломоносова
(Россия, Москва, oleggolozubov@gmail.com; v.m.kolesnikova@mail.ru)

Аннотация. Задачи агроэкологического почвенного мониторинга являются ядром Информационной системы ПГБД РФ, представляющей собой распределенную сеть АПДЦ. За более чем 10-летнюю историю развития ИС накоплен достаточный объем данных для решения ряда аналитических задач. В статье рассматривается применение теории Больших данных в качестве методологической основы интеллектуального анализа в агроэкологическом мониторинге. Принципы распределенных вычислений описаны как с точки зрения теории СУБД и BigData, так и задач мониторинга природно-почвенных параметров больших пространственных территорий и их реализации в ИС ПГБД РФ. Отмечается важность стандартизации данных в такого рода системах для сопоставимости атрибутивной информации и «бесшовности» данных в едином географическом атрибутивном пространстве.

Ключевые слова: базы данных, агроэкологический мониторинг, дата центр, атрибутивная почвенная информация.

О.М. Golozubov, V.M. Kolesnikova

BIG DATA IN AGRO-ECOLOGICAL MONITORING

Faculty of Soil Science, Lomonosov Moscow State University
(Russia, Moscow, oleggolozubov@gmail.com; v.m.kolesnikova@mail.ru)

Annotation. The tasks of agro-ecological soil monitoring are the core of the Information System of the Russian State Data Archives, which is a distributed network of ARDCs. Over the more than 10-year history of the development of IP, a sufficient amount of data has been accumulated to solve a number of analytical problems. The article discusses the application of Big Data theory as a methodological basis for intellectual analysis in agro-ecological monitoring. The principles of distributed computing are described both from the point of view of the theory of DBMS and Big Data, as well as the tasks of monitoring the natural and soil parameters of large spatial territories and their implementation in the IS of the Russian State Data Archive. The importance of data standardization in such systems is noted for the comparability of attributive information and "seamlessness" of data in a single geographic attributive space.

Key words: databases, agro-ecological monitoring, data center, attributive soil information

Задачи агроэкологического почвенного мониторинга можно отнести к классу задач Big Data. Класс задач и технологий Big Data обычно [1,2] определяют правилом 4-х «V»:

СЕКЦИЯ №4. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МОДЕЛИРОВАНИЕ

- **Velocity:** скорость оперативной обработки и формирования результата, интуитивное манипулирование (case-технологии, модели целенаправленного поиска) и on-line обработка (облачные технологии, равный доступ независимо от размещения данных).
- **Volume:** существенный объем данных в обработке - качественная характеристика, определяемая как комплексность (многомерность) и статистически достаточный объем, но при этом часто неструктурированность (разреженность, неполнота данных).
- **Variety:** разнообразие типов данных, их модальностей, разнородность (гетерогенность) источников получения и форм хранения, пространственная распределённость данных.
- **Veracity:** низкое качество «сырых» (raw) данных, потребность оценки их достоверности.

Основной объект приложения Big Data – искусственные среды (экономика, торговля, курсы валют), или «организованные» среды (здравоохранение), в которых решены проблемы получения исходных данных в цифровой форме. Естественные природные среды – метеорология, геология, экология и почвоведение, требуют и более полного сбора данных, и их «гармонизации» и включения моделей в интерпретацию и прогноз [3].

В нашей стране за более чем десятилетнюю историю развития Информационной Системы «Почвенно-географическая база данных Российской Федерации» (ИС ПГБД РФ) накоплен достаточный объем данных для решения ряда аналитических задач агроэкологического почвенного мониторинга [4]. Большой массив агроэкологических данных накоплен региональными агрохимцентрами, организациями, генерирующими мониторинговую информацию, осуществляющими наполнение баз архивных данных, выполняющими обработку данных дистанционного зондирования (ДДЗ) на основе существующих нормативных документов [5,6]. Выше, в правилах 4-х ‘V’, и далее в этой статье, мы приводим расширенные характеристики задач Big Data, основанные именно на опыте работы с мониторинговыми данными ИС ПГБД РФ.

Архитектурно-организационные принципы для построения информационных систем агроэкологического почвенного мониторинга и, более широко, пространственно-природных данных сочетают в себе как типовые принципы Big Data, так и особенности пространственно-распределенных мониторинговых сетей. Так, организационным комитетом INSPIRE («Инфраструктура пространственной информации в Евросоюзе»), образованным в 2007 г., проведена работа для создания информационной инфраструктуры, обеспечивающей свободный публичный доступ к пространственной природно-экологической информации [7,8]. В ней предусматривается

ЭКОЛОГИЯ РЕЧНЫХ БАССЕЙНОВ

а) хранение только первичных данных в одном месте сбора и обработки; б) обеспечение неразрывности пространственных данных на административных границах; в) их доступности в различных масштабах (от детального для исследований, до общего для стратегических целей) и г) обеспечивается «прозрачность» поиска нужной пространственной информации при понятных условиях доступа к ней. Отметим, что для почвенных данных (также как и для геологических, ботанических и водных) разрабатываются специальные стандарты описания и обмена данными [8], эти почвенные стандарты близки к стандартам ISO 28258, и также не являются законченным продуктом, а развиваются по мере развития проекта в целом.

Проект INSPIRE, так же как ИС ПГБД РФ и аналогичные глобальные мониторинговые проекты решают проблемы необходимости постоянного привлечения больших объемов новой информации, ее хранения и обработки. Рядом исследователей во главе с одним из пионеров технологии баз данных Майклом Стоунбрейкером (Michael Stonebraker) удалось в начале XXI века сформулировать основные принципы для решения этих проблем. В частности, система управления базами данных (СУБД) и приложения баз данных организуются таким образом, чтобы минимизировались пересылки данных по сети, связывающей узлы соответствующей вычислительной системы. Очевидно, что важность этого принципа возрастает при росте объема данных, возникновении потребности в переносе приложений баз данных (частями или полностью) на сторону сервера. Второй принцип обеспечивает возможность реального распараллеливания работы СУБД и приложений, поскольку при отсутствии общих ресурсов между узлами вычислительной системы (фактически, при использовании кластерной архитектуры) уменьшается вероятность конфликтов между частями системы и приложений, выполняемыми в разных узлах сети. Наконец, третий принцип обеспечивает эффективную параллельную обработку транзакций или эффективную поддержку оперативной аналитической обработки данных. То есть, архитектурно-организационные принципы построения глобальных мониторинговых проектов соответствуют теории Big Data и, более того, расширяют некоторые понятия и принципы. На рис.1 приведена организационная схема распределенной базы данных ИС ПГБД РФ, отражающая указанные выше принципы.

Типовые методы и механизмы систем Big Data применительно к задачам агроэкологического почвенного мониторинга имеют ряд особенностей. Исходно, технология комплексного оперативного многомерного анализа данных получила название OLAP. OLAP — это ключевой компонент организации хранилищ данных. Концепция OLAP была описана в 1993 году Эдгаром Коддом, известным исследователем баз данных и автором реляционной модели данных [9]. В литературе эти технологии часто называют *Data mining* -

СЕКЦИЯ №4. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МОДЕЛИРОВАНИЕ

методы обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности. Для принципиально неполных и разреженных почвенных данных прогнозирование — важный этап экологического мониторинга, позволяющий заполнить пробелы в данных. При этом решаются задачи регрессии и классификации, формируются сводные отчеты на многомерном «кубе» и глобальные многослойные географические карты, в основном здесь применяются «облачные» решения, поскольку исходная информация тематически и пространственно распределена между соответствующими узлами сети.



Рис. 1. Организационная схема ИС ПГБД РФ

Такие общепринятые механизмы Big Data как *Map Reduce* так же находят своё применение в мониторинговых системах. Задача распараллеливания данных и решений (Map) здесь решается естественным путем, поскольку вычисления над природно-почвенными данными, как правило, затрагивают географически локальную информацию, тематически гетерогенную и размещенную в соответствующих распределенных узлах. А задача сокращения вычислений и их типизация (Reduce) выполняется за счет а) двухтактной системы сбора данных – сначала метаданные, а затем собственно данные от определенных узлов, и б) предобработки данных в узлах сети и окончательной сборки в «облаке».

И вот здесь, в условиях пространственно и предметно распределённых гетерогенных сетей, особенностью и одной из

ЭКОЛОГИЯ РЕЧНЫХ БАССЕЙНОВ

важнейших задач является стандартизация привлекаемых в ИС разнородных данных. Аналогично системам стандартизации проекта INSPIRE, опирающимся на серии стандартов почвенных обследований и описаний ISO серии TS190, и другим глобальным проектам, для ИС ПГБД РФ разработаны прототипы стандартов почвенных описаний, форматов получения, хранения, представления и обмена почвенной информацией [10]. Также как и современные международные стандарты, этот прототип представляется собой комплекс структурированных гипертекстовых описаний (XML\XSD) и программного обеспечения, реализованного как в виде автономных приложений для полевых обследований, так и в виде веб-приложений для загрузки и модификации почвенных данных. Использование такого программного обеспечения позволяет сочетать региональную и тематическую специализацию с типизацией алгоритмов и процедур, что обеспечивает совместимость и сопоставимость данных, формирует единое «бесшовное» географическое пространство на уровне страны.

Список использованных источников

1. Лесковец Ю., Раджараман А., Ульман Д. Д. Анализ больших наборов данных //М.: ДМК Пресс. – 2016. – Т. 488с.
2. Кузнецов С. К свободе от проблемы Больших Данных //Открытые системы. СУБД, – 2012. – №. 2. – С. 22-22.
3. Ужинский А. Интеллектуальная платформа экологического мониторинга //Открытые системы. СУБД. - 2021. - №. 02. //https://www.osp.ru/os/2021/02/13055942.
4. Technologies and Standards in the Information Systems of the Soil-Geographic Database of Russia // Golozubov O.M., Rozhkov V.A., Alyabina I.O., Ivanov A.V., Kolesnikova V.M., Shoba S.A.// Eurasian Soil Science, 2015 ,vol. 48, № 1, pp. 1-10.
5. Приказ Минсельхоза России от 04.05.2010 №150 «Об утверждении Порядка государственного учета показателей состояния плодородия земель сельскохозяйственного назначения».
6. Приказ Минсельхоза России от 02.12.2020 №729 «О внесении изменений в приказ Минсельхоза России от 04.05.2010 №150 «Об утверждении Порядка государственного учета показателей состояния плодородия земель сельскохозяйственного назначения».
7. INSPIRE Principles <https://inspire.ec.europa.eu/inspire-principles/9>
8. INSPIRE D2.8.III.3 Data Specification on Soil – Technical Guidelines// <https://inspire.ec.europa.eu/data-specifications/2892>.
9. E.F. Codd, S.B. Codd, and C.T. Salley, Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate. Technical report, 1993.
10. Голозубов О.М., Исмаилов А.И., Колесникова В.М., Морозов И.В., Розлога Ю.Г., Филипчук В.Ф., Чернова О.В., Многоязычная

СЕКЦИЯ №4. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МОДЕЛИРОВАНИЕ

база метаданных, объектная модель и программный комплекс для стандартизации и обмена почвенной информацией Soil_ML_MultyL (SOFT), авторское свидетельство #24121 от 7 июня 2019 г.

УДК 502.51+504.45+ 504.4.054

С.М. Клубов¹, Ю.О. Рожкова, В.Ю. Третьяков, В.В. Дмитриев

ФОРМИРОВАНИЕ СТОКА АЗОТА И ФОСФОРА НА УРБАНИЗИРОВАННЫХ ВОДОСБОРАХ МАЛЫХ РЕК

Санкт-Петербургский государственный университет
(Россия, Санкт-Петербург, ¹klubov_stepan@mail.ru)

Аннотация. Рассматриваются модули стока биогенных элементов (общего азота и фосфора) с водосборов ряда рек бассейна Балтийского моря (реки Охта, Волковка, Нева, Ижора, Луга и другие). Для расчета этих модулей стока использованы результаты собственных исследований водосборов рек Охты и Волковки в Санкт-Петербурге. Применены результаты моделирования стока биогенных элементов, выполненного в Институте озераведения РАН с использованием модели ILLM. Рассчитаны модули стока азота и фосфора с водосборов водных объектов, на которых отсутствует мониторинг гидрохимического состава воды. Выполнено сравнение рассчитанных модулей стока общего азота и фосфора с модулями, определёнными на основании максимально допустимых количеств поступления азота и фосфора в Балтийское море в соответствии с решениями Хельсинкской комиссии по защите морской среды Балтийского моря (HELCOM).

Ключевые слова: эвтрофирование Балтийского моря, модули стока азота и фосфора

S.M. Klubov¹, I.O. Rozkova, V.Yu. Tretyakov, V.V. Dmitriev

NITROGEN AND PHOSPHORUS DISCHARGES FORMATION WITHIN URBANIZED WATERSHEDS OF SMALL RIVERS

Saint Petersburg State University
(Russia, Saint Petersburg, ¹klubov_stepan@mail.ru)

Abstract. Specific discharges of total nitrogen and phosphorus from watersheds of several rivers of the Baltic Sea catchment area are considered. The discharges values were calculated on the base of our researches of the Okhta River and Volkovka River watersheds. We used some results of the biogenic elements outflow simulation, carried out in Institute of Limnology of Russian Academy of Sciences with usage of ILLM model. The discharges from water objects watersheds without the water chemical state monitoring stations have been calculated. The calculated values of the discharges were compared with the discharges corresponded the maximum permissible nitrogen and phosphorus intakes into the Baltic Sea according to Baltic Marine Environment Protection Commission (the Helsinki Commission, HELCOM) conclusions.

Keywords: the Baltic Sea eutrophication; nitrogen and phosphorus specific discharges.

Введение

Специфика Балтийского моря заключается в наличии обширного водосборного бассейна, расположенного в пределах ряда стран с развитой промышленностью и сельским хозяйством. Это определяет высокую степень антропогенной нагрузки на экосистему моря. На побережье Балтийского моря расположено много крупных