

А.В. ВОРОБЬЕВ, Г.Р. ВОРОБЬЕВА, Н.И. ЮСУПОВА
**КОНЦЕПЦИЯ ЕДИНОГО ПРОСТРАНСТВА ГЕОМАГНИТНЫХ
ДАННЫХ**

Воробьев А.В., Воробьева Г.Р., Юсупова Н.И. **Концепция единого пространства геомагнитных данных.**

Аннотация. Задача мониторинга параметров геомагнитного поля и его вариаций преимущественно решается сетью магнитных обсерваторий и вариационных станций, однако значимым препятствием при обработке и анализе получаемых таким образом данных наряду с их пространственной анизотропией являются пропуски (или полное отсутствие) достоверных значений и частичное несоответствие установленному формату. Неоднородность и аномальность данных исключает (существенно усложняет) возможность их автоматической интеграции и применения к ним инструментария для частотного анализа. Известные решения по интеграции разнородных геомагнитных данных базируются преимущественно на модели консолидации и лишь частично решают данную проблему. Получаемые в результате наборы данных, как правило, не соответствуют требованиям IAGA (International Association of Geomagnetism and Aeronomy — Международной ассоциации геомагнетизма и аэрономии), рекомендуемым к представлению результатов геомагнитных наблюдений. При этом пропуски во временных рядах устраняются известными средствами обработки геомагнитных данных путем исключения отсутствующих или аномальных значений из конечной выборки, что, очевидно, может привести как к потере актуальной информации о ходе изменения параметров геомагнитного поля и его вариаций, нарушению шага дискретизации, так и к неоднородности временного ряда. Предлагается подход к созданию единого пространства геомагнитных данных, основанный на комбинировании моделей консолидации и федерализации, включающий предварительную обработку исходных временных рядов с опционально доступной процедурой их восстановления и верификации, ориентированный на применение технологий облачных вычислений и иерархического формата с целью повышения вычислительной скорости обработки больших объемов данных и, как следствие, обеспечивающий получение пользователями более качественных и однородных данных.

Ключевые слова: геомагнитные данные, магнитные обсерватории, временные ряды, большие данные, единое информационное пространство, параллельные вычисления.

1. Введение. В настоящее время установлены и активно изучаются многочисленные негативные эффекты воздействия космической среды на объекты народного хозяйства, наиболее ярко проявляющиеся в периоды так называемых магнитных бурь. Особого внимания заслуживают такие эффекты, как:

- магнитное торможение искусственных спутников Земли;
- нарушение коротковолновой радиосвязи;
- дополнительные погрешности прецизионной магнитометрической аппаратуры;
- радиационное воздействие на биологические объекты, находящиеся в верхних слоях атмосферы;

– токовые наводки в трубопроводах, трансокеанских кабелях, системах автоматики высокоширотных железных дорог.

Актуальная в этой связи проблема многопараметрового мониторинга геомагнитного поля и его вариаций преимущественно решается посредством магнитных обсерваторий, аэромагнитных, гидромагнитных съемок, спутниковых и подземных скважинных наблюдений, а также с помощью портативных магнитометров различного принципа действия и динамического диапазона [1]. Каждый из способов имеет свои характерные преимущества и недостатки относительно других и используется для решения определенного круга задач как прикладного, так и фундаментального характера. Однако, благодаря реализуемой концепции объединения магнитных обсерваторий в сети (INTERMAGNET [2, 3], IMAGE [4], AUTUMNX [5, 6] и пр. [7]) и открытому удаленному доступу к регистрируемым ими данным, именно они являются наиболее достоверным, распространенным и доступным для большинства ученых и специалистов методами наблюдения вариаций геомагнитного поля.

По этой и другим причинам (по состоянию на 2018 г.) свыше 30 известных широкой общественности сетей магнитных обсерваторий объединяют более 300 магнитных станций и обсерваторий, чье неравномерное распределение по поверхности Земли обуславливает относительный избыток геомагнитных данных в одних регионах планеты и их дефицит в других [1]. Однако, помимо проблем, сопряженных с геопространственным распределением магнитных обсерваторий, значимым препятствием на пути интеграции, обработки и анализа получаемой с их помощью информации являются несоблюдение единого формата и несогласованность представления результатов наблюдений параметров геомагнитного поля и его вариаций, а также различный период их регистрации. При этом разнородность геомагнитных данных обусловлена гетерогенностью их источников, одни из которых представлены магнитными обсерваториями, регистрирующими абсолютные значения параметров магнитного поля Земли, а другие — вариационными станциями, осуществляющими наблюдение за параметрами геомагнитных вариаций.

Предлагается одно из возможных решений задачи интеграции гетерогенных источников геомагнитных данных в единое информационное пространство, позволяющее потребителям получать полную и достоверную информацию о состоянии магнитного поля Земли в любой известный момент времени и в любой точке пространства. Приводится описание предложенной концепции, архитектуры единого пространства геомагнитных данных, а также инфокоммуникационных технологий, реализующих указанный подход.

2. Состояние вопроса. До начала 90-х годов XX века основным способом представления результатов наблюдений параметров магнитного поля Земли и его вариаций являлись магнитограммы на бумажном носителе [8, 9]. С развитием технологий вариационные станции и магнитные обсерватории перешли на цифровую систему регистрации и обработки геомагнитных данных, результатом которой стали ежесуточные файлы минутных значений вариаций геомагнитного поля с привязкой к абсолютным наблюдениям. И, наконец, эволюция инфокоммуникационных технологий позволила обеспечить открытость геомагнитных данных посредством стандартных сетевых протоколов и веб-ориентированных интерфейсов.

В итоге на сегодняшний день большинство магнитных обсерваторий и станций располагают размещенными в сети Интернет ресурсами, на которых доступны результаты многолетних наблюдений параметров геомагнитного поля и его вариаций. Так, к примеру, на веб-ресурсе сети INTERMAGNET (<ftp://ftp.seismo.nrcan.gc.ca/intermagnet>) размещены вариативные, предварительные, квазиокончателные и окончательные результаты поминутных и посекундных наблюдений параметров геомагнитного поля с 1991 года по настоящее время [2, 3]. Также в репозитории сети IMAGE (<http://space.fmi.fi/image/www/index.php?page=home>) доступны статистические таблицы и графики, отражающие результаты наблюдений компонент вектора геомагнитного поля (с шагом дискретизации 10-20 с или 1 мин) [4].

Однако и на современном уровне развития технологий существует ряд технических проблем, возникающих при обработке и анализе больших объемов гетерогенных геомагнитных данных, которые характерны и для других типов данных о состоянии окружающей среды [10, 11]. В первую очередь, магнитные сети, как правило, не соблюдают общепринятого формата представления результатов геомагнитных наблюдений, внося изменения в стандартный формат IAGA2002 [7] или используя собственные специфичные форматы, которые базируются на текстовом представлении табличных данных TSV (tab-separated values) [4]. Кроме того, наблюдаемые станциями и обсерваториями параметры геомагнитного поля и его вариаций отличаются: одни из них регистрируют направление магнитного поля, выраженное измеряемыми в градусах склонением (D) и наклонением (I), а другие — описывают полный вектор напряженности геомагнитного поля (F) на основе трех его компонент, измеряемых в нТл в различных системах координат [12-14]. И, наконец, может варьироваться шаг дискретизации регистрации геомагнитных данных, который принимает значения от полсекунды [6] до минуты [2, 4].

В определенном смысле усугубляет обозначенную проблему и тот факт, что регистрируемые станциями и обсерваториями геомагнитные данные содержат систематические пропуски (или полное от-

сутствие) достоверных значений за определенный период наработки ИМО-станции (так, например, в период с 00:00 по 23:59 ч (UTC) 9 марта 2017 г. магнитной обсерваторией «Arti» пропущено 26 % значений временного ряда, магнитной обсерваторией «Guam» – 0.17 %, данные по обсерватории «Vostok» отсутствуют и т.д.). Следует также отметить и то, что геомагнитные данные, публикуемые INTERMAGNET и другими магнитными сетями, могут быть как очевидно недостоверными, так и не соответствовать формату IAGA-2002. Немаловажен и тот факт, что неполнота временных рядов (уже на теоретическом уровне) исключает возможность применения к ним математического и программно-алгоритмического инструментария частотного анализа и влияет на достоверность результатов сопряженных исследований, опирающихся на геомагнитные данные, получаемые посредством этих обсерваторий [1].

При этом информационные нужды потребителей геомагнитных данных не всегда ограничиваются масштабами одной магнитной сети и могут быть связаны с анализом и обработкой результатов наблюдений нескольких станций и обсерваторий, относящихся в том числе и к разным сетям. В этом случае потребитель сталкивается с очевидной технической проблемой, связанной с высокой трудоемкостью и затратами времени на поиск и получение необходимых данных. При этом все выявленные ранее проблемы геомагнитных данных (формат, структура, шаг дискретизации и т.д.) выходят на первый план и делают невозможным их автоматизированный анализ и обработку.

Попытка объединить результаты разнородных наблюдений параметров геомагнитного поля и его вариаций была предпринята в рамках проекта SuperMag (<http://supermag.uib.no/index.html>), инициированного в составе программы Electronic Geophysical Year (eGY, 2007–2008). Проект обеспечивает единый веб-ориентированный доступ к геомагнитным данным вариационных станций в единой координатной системе и унифицированных единицах измерения с заданным временным разрешением [14, 15]. В настоящее время ресурс предоставляет потребителям геомагнитные данные за период 1980–2010 годов, а количество доступных станций при этом варьируется от 90 до 165 в зависимости от анализируемого временного интервала.

В литературе [15] в качестве основного недостатка проекта SuperMag выделяют его ограниченность: в репозитории доступны только вариационные данные, в то время как отсутствуют абсолютные измерения, необходимые при исследовании динамики главного геомагнитного поля, а также практического применения для навигации, ориентации и геологии. В базе данных SuperMag также представлены данные, как правило, не менее семидневной давности, поэтому невозможно исполь-

зовать проект в качестве источника данных в режиме реального (или близком к нему) времени, в то время как такая необходимость часто возникает при решении различного рода прикладных задач.

Немаловажен и тот факт, что выбросы и пропущенные во временных рядах геомагнитных данных значения удаляются из набора предоставляемых потребителю данных, что может привести к потере критически важной информации, а также негативно сказывается на качестве и информативности систем визуализации данных. Еще один недостаток связан с тем, что поисковая система не отлажена и работает нестабильно, пропуская подавляющую часть запросов к станциям. И, наконец, проект SuperMag, как яркий пример консолидации данных, наследует его главный недостаток – избыточность, поскольку данные дублируются как в отдельных источниках, так и в едином репозитории.

Обозначенные проблемы интеграции результатов наблюдения параметров магнитного поля Земли и его вариаций, а также выявленные недостатки существующих подходов обуславливают актуальность научно-технической задачи, связанной с совершенствованием методов и средств распространения, обработки, анализа и визуализации больших объемов гетерогенных геомагнитных данных. Ее очевидным решением является объединение множества гетерогенных источников в единое пространство геомагнитных данных под управлением централизованного метода доступа, а также инструментария, обеспечивающего возможность их анализа и визуализации.

3. Постановка задачи. С целью расширения доступности результатов наблюдений параметров магнитного поля Земли и его вариаций необходимо разработать веб-ориентированную систему, реализующую инструментарий ведения и использования единого пространства геомагнитных данных как результата интеграции распределенных разнородных источников. Разработка информационной системы предполагает решение следующих задач:

1. Формулировка принципов объединения источников гетерогенных геомагнитных данных на основе моделей консолидации и федерализации для интеграции распределенных наборов разнородных данных.

2. Формализация структуры единого пространства геомагнитных данных, обеспечивающая объединение разнородных источников и повышение качества наборов данных за счет предварительной обработки для восстановления временных рядов и устранения аномальных значений в них.

3. Повышение вычислительной скорости получения и обработки больших объемов геомагнитных данных посредством создания виртуального вычислительного кластера, основанного на модели распределенных вычислений MapReduce и отличающегося тем,

что исходный выполняемый процесс выступает в качестве главного узла, а порождаемые им параллельные процессы являются рабочими узлами, реализующими изолированные запросы к гетерогенным источникам геомагнитных данных.

4. Принципы организации единого пространства геомагнитных данных. Под единым пространством геомагнитных данных, по аналогии с [16], будем понимать совокупность гетерогенных источников данных наблюдений за параметрами магнитного поля Земли и его вариаций, а также инфокоммуникационных технологий их интеграции, обработки, анализа и визуализации, функционирующих на основе единых принципов и обеспечивающих информационное взаимодействие поставщиков и потребителей геомагнитных данных, равно как и удовлетворение их информационных потребностей при решении прикладных и научно-исследовательских задач.

Согласно требованиям к структурной интеграции разнородных информационных ресурсов [17], единое пространство геомагнитных данных должно удовлетворять следующим основным принципам:

– прозрачности, согласно которому запросы потребителей геомагнитных данных не связаны с физическим расположением источников данных и формой представления информации в них;

– системности, что определяет необходимость формирования единого пространства как целостного образования, способного удовлетворить информационные потребности пользователей при решении прикладных и научно-исследовательских задач без выделения отдельных источников данных;

– технологичности, предполагающего комплексное использование различных технологий накопления и обработки геомагнитных данных, включая программно-инструментальные средства их комплексного анализа и визуализации.

В основе предлагаемой концепции лежит идея интеграции моделей консолидации и федерализации данных при их объединении в единое пространство с унифицированным для пользователей представлением информации [18]. Обе архитектуры являются однонаправленными и используются для предоставления наборов распределенных данных потребителю. Отличаются они друг от друга тем, что в первом случае происходит копирование данных в отдельное информационное хранилище, а во втором наборы данных формируются динамически без перемещения их из исходных источников.

С учетом перечисленных особенностей архитектур интеграции предлагается использовать консолидацию для сбора и долговременно-го физического хранения аналитических геомагнитных данных, а фе-

дерализацию — для формирования и виртуального хранения наборов оперативных геомагнитных данных. При этом под аналитическими данными будем понимать результаты всех предшествующих текущим суткам наблюдений параметров магнитного поля Земли, а под оперативными — геомагнитные данные, зарегистрированные магнитными обсерваториями и вариационными станциями в течение последних суток и автоматически преобразуемые в аналитические по их истечении. Результирующий набор может быть получен как интеграцией оперативных и аналитических данных, так и сформирован из выборки одного из них. При этом потребители геомагнитных данных имеют возможность в реальном или близком к реальному времени получить нужную информацию, охватываемую этим пространством [17, 18].

Еще одной отличительной особенностью предлагаемой концепции является развитие архитектуры федерализации данных: результаты наблюдений за параметрами магнитного поля Земли копируются в аналитическое хранилище после прохождения процедуры предварительной обработки. Необходимость такой доработки вызвана тем, что временные ряды геомагнитных данных являются, как правило, неполными и содержат различного рода аномалии, что может, в свою очередь, сказаться на достоверности результатов, их обработки и анализа. В рамках концепции восстановление временных рядов геомагнитных данных реализуется предложенным авторами индуктивным методом [19], основанным на статистических методах обработки временных рядов и принципах машинного обучения с использованием размеченных данных и отличающегося тем, что признаковым описанием фрагмента временного ряда выступает пара предшествующего и следующего за ним фрагментов того же ряда, в совокупности образующих обучающую выборку для поиска недостающего фрагмента по набору его признаков с последующим линейным масштабированием для восстановления исходного тренда информационного сигнала.

Таким образом, физическая независимость и скрытая от потребителя базовая неоднородность геомагнитных данных обеспечивают оперативность и высокую эффективность выполнения множественных обращений к данным под управлением единого метода доступа. При этом нивелируется проблема асинхронного обмена данными между несколькими источниками, свойственная консолидации данных и приводящая к тому, что запрашиваемая информация теряет свою актуальность на момент обращения к ней потребителя.

С учетом вышесказанного и согласно известной классификации [17], единое пространство геомагнитных данных является смешанным, поскольку сочетает в себе элементы централизован-

ной (с общим хранилищем данных) и децентрализованной (с распределенным хранилищем данных) архитектур. Присутствие в его структуре эвристически сложных вычислительных процедур позволяет полностью автоматизировать процедуру формирования хранилища геомагнитных данных, оставляя человеку-оператору только роль в подготовке данных на уровне их источников (станций и обсерваторий) за пределами единого пространства геомагнитных данных. При этом указанные вычислительные процедуры предназначены не только для сбора данных из распределенных источников, но и для выполнения аналитических операций и визуализации различных срезов данных.

5. Формализация структуры единого пространства геомагнитных данных. Структура единого пространства геомагнитных данных представлена тремя разделами, обеспечивающими сбор, хранение и обработку гетерогенных данных (рисунок 1). Основным его компонентом является распределенный набор источников геомагнитных данных, представленный как магнитными сетями, так и отдельными станциями и обсерваториями, которые осуществляют мониторинг параметров магнитного поля Земли и его вариаций.

Обозначим совокупность источников геомагнитных данных как D . Тогда распределенные средства сбора и хранения разнородных геомагнитных данных можно представить, как:

$$D = \langle M_n, M_s, M_o \rangle; \forall s \in M_s : s \notin \{M_n\}; \forall o \in M_o : o \notin \{M_n\}; s \neq o,$$

где M_n — множество магнитных сетей, M_s — множество автономных вариационных станций, M_o — множество автономных магнитных обсерваторий.

Отметим, что справедливо соотношение:

$$\exists d \in \{M_n\} : d \in M_{n_i}, d \in M_{n_j}, i \neq j,$$

где d — произвольный экземпляр из множества магнитных сетей M_n .

Важно, что поставщики геомагнитных данных предоставляют специализированные веб-сервисы, которые по стандартным инфокоммуникационным протоколам осуществляют передачу данных потребителям, но не обеспечивают их предварительную обработку. В результате конечный информационный продукт представляет собой набор неполных временных рядов с пропусками и аномалиями, содержащих разноформатные результаты неоднородных геомагнитных наблюдений, проведенных с различной частотой (в среднем — от нескольких секунд до минуты).

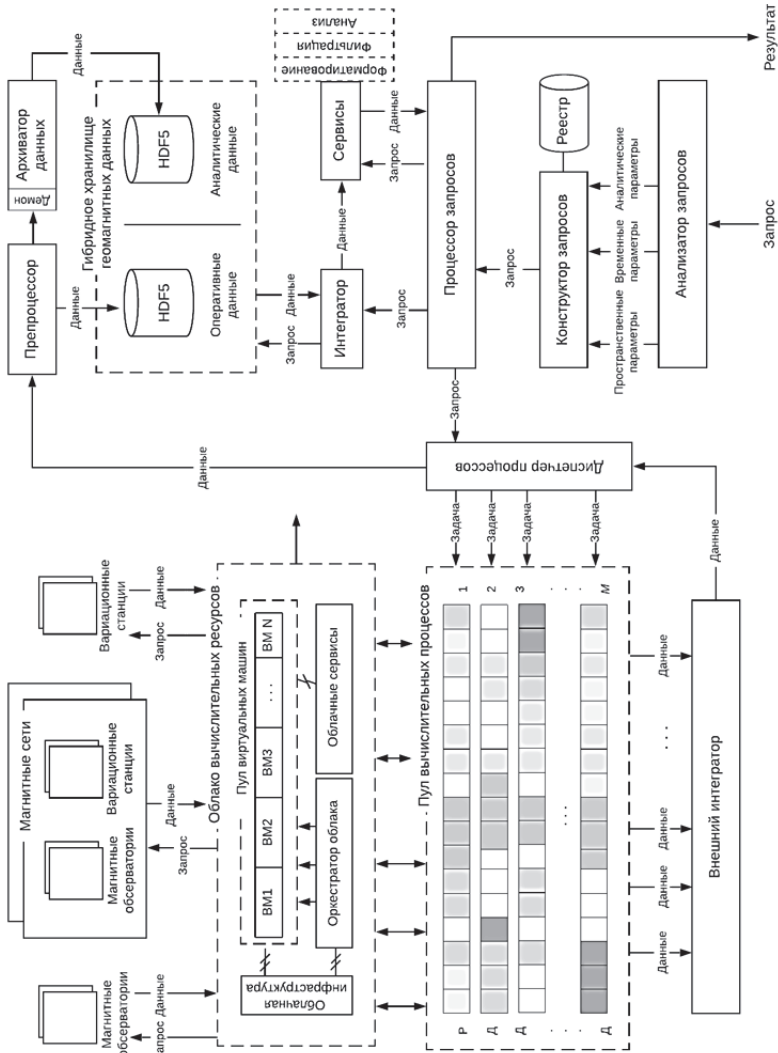


Рис. 1. Концепция единого пространства геомагнитных данных

Все доступные в едином пространстве поставщики геомагнитных данных обозначены в специальном информационном разделе — реестре R , содержащем URL доступа к соответствующим сервисам данных:

$$R = \{R_i\}, i = \overline{1, n}; n = (l_1 - l_2) + l_3 + l_4,$$

$$l_1 = \|M_n\|; l_2 = \|M_n^*\|, M_n^* = \cap M_n; l_3 = \|M_s\|; l_4 = \|M_o\|.$$

Следующий компонент единого пространства – гибридное хранилище геомагнитных данных H , сочетающее в себе физически хранимые аналитические данные A и размещенные в виртуальном кэше оперативные данные B , прошедшие предварительную обработку:

$$H = A \cup B.$$

В зависимости от параметров запроса результат формируется как комбинированием данных обоих хранилищ, так и применением к ним фильтрации по пространственно-временным признакам.

Формирование как содержимого хранилища, так и ответа на запрос к нему осуществляется связанной модульной структурой M единого пространства геомагнитных данных:

$$M = \{M_i\}, i = \overline{1, 8}.$$

Всего в архитектуре единого пространства геомагнитных данных предложены восемь модулей, каждый из которых обеспечивает определенные этапы сбора, обработки и анализа гетерогенных данных. Так, анализатор запросов декомпозирует полученное от потребителя сообщение на набор параметров P , выделяя среди них пространственные, временные и аналитические:

$$P = \{S, T, A\},$$

где S — множество пространственных параметров (географические координаты, перечень станций, обсерваторий и пр.), T — множество временных параметров (временной промежуток, за который требуется получить данные), A — множество аналитических параметров (при необходимости — вид применяемого анализа (Фурье-преобразование, частотный, амплитудный и пр.)).

На следующем шаге выделенные параметры передаются конструктору запросов, формирующему тексты команд T для обраще-

ния к разделам гибридного хранилища геомагнитных данных, в том числе предварительно из реестра извлекаются локаторы для требуемых источников:

$$C = U \cup T,$$

где U — множество локаторов источников на основании множества S .

Основным компонентом модульной структуры единого пространства геомагнитных данных является процессор запросов, который на программном уровне оперирует текстами полученных от конструктора команд и передает их выполнение соответствующим программным интерпретаторам. При этом направление передачи команд от процессора к интерпретаторам зависит от типа запроса.

Так, если запрос предполагает обращение только к аналитическим данным, то управление получает интегратор, объединяющий наборы результирующих и разреженных по параметрам данных. Интегратор, в свою очередь, может передать полученный массив аналитическим сервисам, обеспечивающим возможность форматирования данных (приведение к одному из поддерживаемых форматов, в том числе IAGA2002), их фильтрации, а также аналитической обработки. Полученный в результате набор данных возвращает управление процессору запросов для дальнейшей работы с потребителем.

В случае, если выполнение запроса предполагает обращение к исходным распределенным источникам данных (на основании выборки локаторов из реестра), то процессор передает управление диспетчеру процессов, декомпозирующему исполняемую процедуру выборки на множество параллельно исполняемых подпроцессов. Результирующие данные поступают обратно в диспетчер процессов из внешнего интегратора, получающего управление по завершению соответствующих вычислительных подпроцессов и объединяющего полученные в ходе их выполнения результаты.

На следующем шаге полученные геомагнитные данные передаются в препроцессор для прохождения процедуры предварительной обработки — обнаружения выбросов и аномалий, восстановления временного ряда, форматирования и прочее. Если массив данных содержит наблюдения параметров геомагнитного поля только за текущие сутки, то он передается в виртуальный кэш гибридного хранилища для последующего его использования интегратором и аналитическими сервисами. В противном случае данные передаются в специальный модуль — архиватор. Особенностью последнего является то, что он выполняется процессом-демоном, автоматически инициирующим процедуру заполнения хранилища данными по истечении суток. При этом оперативные данные

формируются по запросу, заданному не только потребителем геомагнитных данных, но и самим процессом-демоном (работа процесса ведется по принципу классического планировщика CronTab), который, в свою очередь, эмулирует запрос на получение геомагнитных данных от всех доступных источников за текущие сутки.

Еще один компонент в структуре единого пространства геомагнитных данных играет решающую роль в обеспечении высокой вычислительной скорости обращения к распределенным источникам геомагнитных данных. Таким компонентом является облако вычислительных ресурсов O , предоставляющее среду выполнения программного кода на базе модели PaaS (Platform as a Service, платформа как сервис) [20, 21]. Облачная инфраструктура единого пространства геомагнитных данных обслуживает ряд виртуальных машин, каждой из которых выделяется один или несколько заданных вычислительных процессов по получению и обработке массивов данных. При этом задачи управления виртуальными машинами, их диспетчеризации, распределения аппаратных и программных ресурсов решает специализированный модуль — оркестратор облака.

Одновременное завершение всех параллельных вычислительных процессов не является обязательным, но каждый из них формирует одну часть результирующего набора разнородных геомагнитных данных, который после дополнительной обработки унифицируется и передается во внешний интегратор, формирующий данные для гибридного хранилища (как аналитические, так и оперативные). Таким образом, время ожидания результирующего набора данных соответствует максимальному времени, отводимому на выполнение процессам из вычислительного пула облачной инфраструктуры. При этом отметим, что число виртуальных машин, как правило, меньше выделенных вычислительных процессов, поэтому требуется диспетчеризация и повторное использование ресурсов, что осуществляется оркестратором облачной инфраструктуры.

6. Технологии единого пространства геомагнитных данных.

С технической точки зрения единое пространство геомагнитных данных представляет собой распределенную информационную систему, обеспечивающую обработку, анализ и визуализацию больших объемов гетерогенных геомагнитных данных. Однако реализация такой системы имеет ряд нюансов, связанных со спецификой как самих данных, так и их источников.

Требование доступности единого пространства геомагнитных данных широкому кругу потребителей обуславливает необходимость разработки веб-ориентированной информационной среды как промежуточного звена между распределенными источниками данных

и пользователями. Традиционная клиент-серверная архитектура в сочетании с современными технологиями веб-программирования и дизайна (PWA, progressive web applications) позволяет создать независимое от платформы и простое в использовании приложение, доступное пользователям, независимо от типа устройства. Дополнительно к этому большой объем данных, сложность аналитических операций, необходимость доступного представления потребителям результатов обработки геомагнитных данных и обеспечение высокой вычислительной скорости в веб-среде обуславливают выбор в пользу некомпилируемого расширяемого языка программирования, текст на котором можно выполнить на виртуальной машине со специально выделенными областями памяти и стека.

Далее в рамках предложенной концепции для решения проблемы высоких аппаратных требований к организации гибридного хранилища предлагается хранить геомагнитные данные, которые прошли предварительную обработку, в иерархическом бинарном формате HDF5 (hierarchical data format), специально разработанном для больших объемов цифровой информации. Структура документов HDF5 схожа с иерархической файловой системой, при этом для доступа к данным применяются пути, сформированные на основании POSIX-синтаксиса, а метаданные задаются в виде набора именованных атрибутов соответствующих объектов [22]. Применяемые алгоритмы сжатия сокращают физический объем требуемого для хранения данных дискового пространства, а широкий спектр инструментально-программных средств обработки формата позволяет существенно повысить вычислительную скорость выполнения основных операций над данными. Указанные преимущества являются доводом в пользу выбора формата HDF5 для организации хранилища геомагнитных данных.

Кроме того, с целью повышения вычислительной скорости получения геомагнитных данных из гетерогенных источников предлагается введение виртуального вычислительного кластера на базе модели распределенных вычислений MapReduce и принципов многопроцессности. Однако реализация предлагаемого подхода на веб-платформе технически невозможна из-за использования здесь концепции мьютекса GIL (Global Interpreter Lock). В этой связи предлагается использовать принцип многопоточности, разделив главный узел на множество параллельно выполняемых независимых потоков, число которыхратно вычислительной мощности серверов, обеспечивающих работу единого пространства геомагнитных данных. Тогда вычислительная нагрузка будет распределена между всеми доступными (или их большинством) ядрами процессора, а параллельно выполняемые процессы не замедлят работу друг друга и в совокупности приведут к увеличе-

нию вычислительной скорости работы с большими объемами геомагнитных данных. Предлагаемая схема многопоточной обработки больших объемов геомагнитных данных приведена на рисунке 2.

Показано, что вычислительный процесс декомпозируется на множество групп, состоящих из n потоков. Диспетчер процессов присваивает каждому из них идентификатор, направляя его значение в модуль управления ресурсами потоков — мьютекс. Здесь потоки формируют очередь, при этом каждый из них гарантированно получает управление и свой набор вычислительных ресурсов — состояние потока, которое, в свою очередь, выделяется в результате взаимодействия мьютекса, диспетчера процессов и пула вычислительных ресурсов. Аналогичным образом происходит освобождение ресурсов: завершивший выполнение поток передает идентификатор мьютексу, который направляет его диспетчеру, освобождающему соответствующее состояние потока и формирующему новое для следующего в очереди.

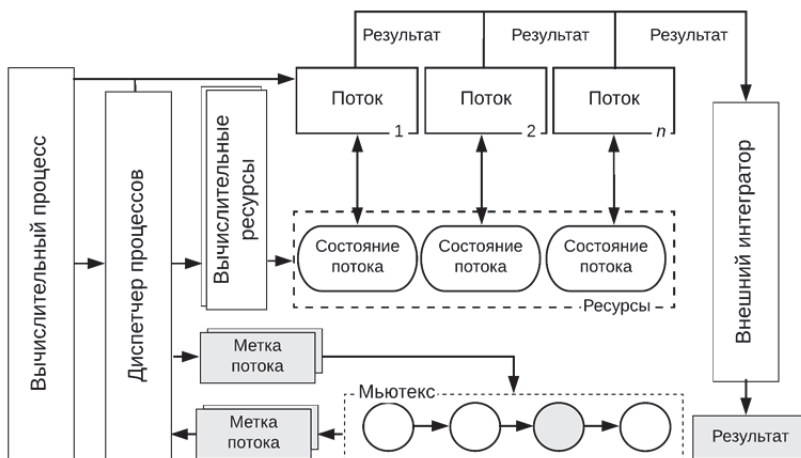


Рис. 2. Схема многопоточной загрузки больших объемов геомагнитных данных

Важно отметить, что параллельно выполняемые вычислительные потоки гораздо менее требовательны к ресурсам, чем параллельно выполняемые вычислительные процессы, и с ними гораздо практичнее выполнять в однопроцессорных системах программы, логика которых требует применения нескольких потоков исполнения. Большинство современных хостингов позволяют разделить таким образом вычислительный процесс в среднем на полтора десятка

потоков и, соответственно, увеличить вычислительную скорость в то же количество раз.

Однако, если выйти за пределы возможностей хостинга и использовать облачную инфраструктуру, то предоставляемые при этом вычислительные скорости и мощности можно дополнительно увеличить. В этом случае вычислительные потоки должны быть равномерно распределены между доступными виртуальными машинами облака, образующими виртуальный вычислительный кластер и также реализующими модель MapReduce [23].

В основе облачной платформы лежит концепция виртуализации, которая обеспечивает предоставление пользователям и системам набора вычислительных ресурсов, абстрагированное от аппаратной реализации. Физические ресурсы и сервисы облака объединяются, и из них оркестратором выделяется пул виртуальных машин, где выполняются гостевые вычислительные процессы. Почасовая тарификация, возможность конфигурирования используемых виртуальных машин (число и степень загрузки ядер процессора) и количество выделяемой памяти позволяют гибко настраивать вычислительные возможности облачной инфраструктуры. В результате число параллельно исполняемых потоков будет кратно вычислительной мощности арендуемой облачной инфраструктуры (количеству виртуальных машин и ядер в них).

7. Пример реализации. На текущем этапе исследований разработан прототип веб-ориентированной информационной системы, реализующей концепцию единого пространства геомагнитных данных на основе инфокоммуникационных технологий и геоинформационного моделирования и включающей ряд сервисов обработки и визуализации геомагнитных данных.

Так, в составе информационной системы реализован сервис «Геомагнитный калькулятор» (рисунок 3), осуществляющий расчет параметров невозмущенного геомагнитного поля в точке с заданными пространственно-временными координатами [24, 25].

Сервис доступен по адресу https://www.geomagnet.ru/geomagnetic_calculator.html и предоставляет пользователям возможность рассчитать следующие параметры для заданной географической точки и временной метки:

- геомагнитные координаты;
- параметры геомагнитного диполя: координаты северного магнитного полюса (в градусах) и магнитный момент (Тл · м³);
- параметры геомагнитного поля: магнитная индукция (нТл · км), компоненты и полный вектор поля (нТл), магнитное склонение и магнитное наклонение (градусы).

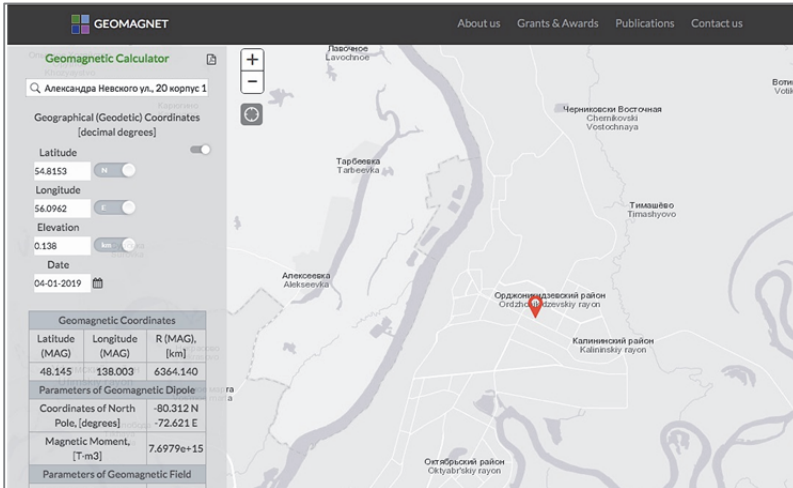


Рис. 3. Экранная форма сервиса «Геомагнитный калькулятор»

Сервис трехмерной визуализации обеспечивает графическую интерпретацию геомагнитных данных и характеризует распределение параметров главного геомагнитного поля по земной поверхности посредством комбинируемых линий уровня (изолиний) и сплайнов, расположенных на интерактивном виртуальном глобусе (рисунок 4).

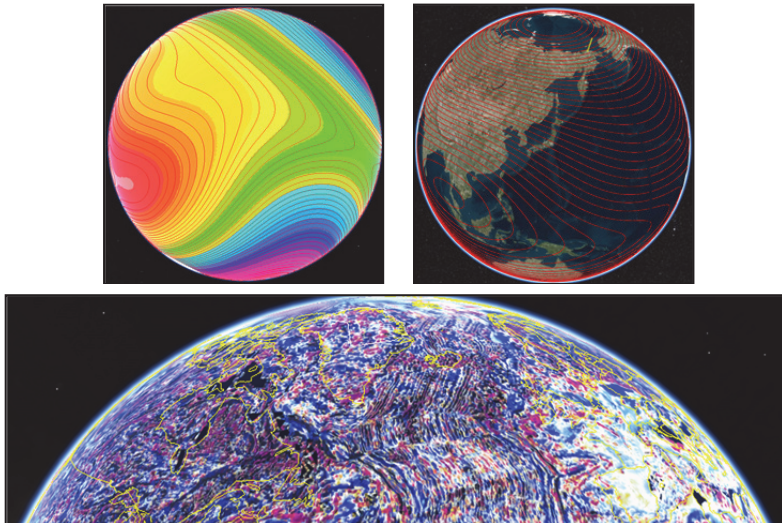


Рис. 4. Экранная форма сервиса трехмерной визуализации геомагнитных данных

Пользователь взаимодействует как непосредственно с трехмерным представлением земной поверхности, используя средства масштабирования и инструментарий геоинформационных технологий (геолокация, прямое и обратное геокодирование и пр.), так и с результатом визуализации параметров главного поля посредством меню различных уровней, ассоциированных с соответствующими графическими элементами (например, каждой линии уровня ставится в соответствие контекстное меню, отражающее значения параметра магнитного поля) [24].

Отличительной особенностью представленной визуализации геомагнитных данных является возможность простого переключения между режимами двух- (2D), трех- (3D) и псевдотрехмерного изображения (2.5D) посредством соответствующего элемента управления, доступного на веб-странице. При этом результаты графической интерпретации геомагнитных данных автоматически адаптируются под выбранный режим представления, что расширяет возможности пользователя при проведении их визуального анализа [24, 25].

Сервис «Магнитные обсерватории» предназначен для предоставления пользователям доступа к результатам наблюдений за параметрами геомагнитного поля и его вариаций, выполняемых одной или несколькими магнитными станциями и обсерваториями.

Особенностью предлагаемого технического решения является информационное обеспечение системы, представленное полученными из распределенных источников геомагнитными данными и регулярно актуализируемое с помощью настраиваемых процессов-демонов. С точки зрения функциональности, сервис обеспечивает возможность поиска магнитных станций и обсерваторий, их фильтрацию, анализ, отображение и вывод доступных данных за указанный период.

На рисунке 5 приведен пример работы информационной системы, демонстрирующий способ выбора магнитной обсерватории с учетом ее пространственной привязки и результаты амплитудно-временного анализа соответствующих геомагнитных данных, полученных по результатам семидневного наблюдения параметров геомагнитного поля и его вариаций.

Применение геоинформационных технологий в рамках данного сервиса позволяет предоставить пользователю несколько альтернативных способов выбора магнитной станции или обсерватории для анализа геомагнитных данных:

– прямое геокодирование: указание значений пространственных координат искомой станции. При этом координаты могут быть указаны в соответствующих полях ввода на экранной форме информационной системы или выбраны путем позиционирования пользователем курсора на виртуальной карте;

– обратное геокодирование: указание наименования, аббревиатуры станции или адреса ее местоположения, которые посредством специализированных веб-ориентированных сценариев преобразуются в наборы пространственных координат. Сопоставление полнотекстовых поисковых запросов осуществляется как с серверной базой пространственных данных, так и с реестром источников геомагнитных данных, предусмотренном предложенной структурой единого пространства геомагнитных данных;

– геолокация: определение реального географического положения электронного устройства, с которого осуществляется выход в Сеть. Для реализации данной функциональности в структуре информационной системы предусмотрен механизм выбора станции или обсерватории, ближайшей к точке, зарегистрированной системой позиционирования.

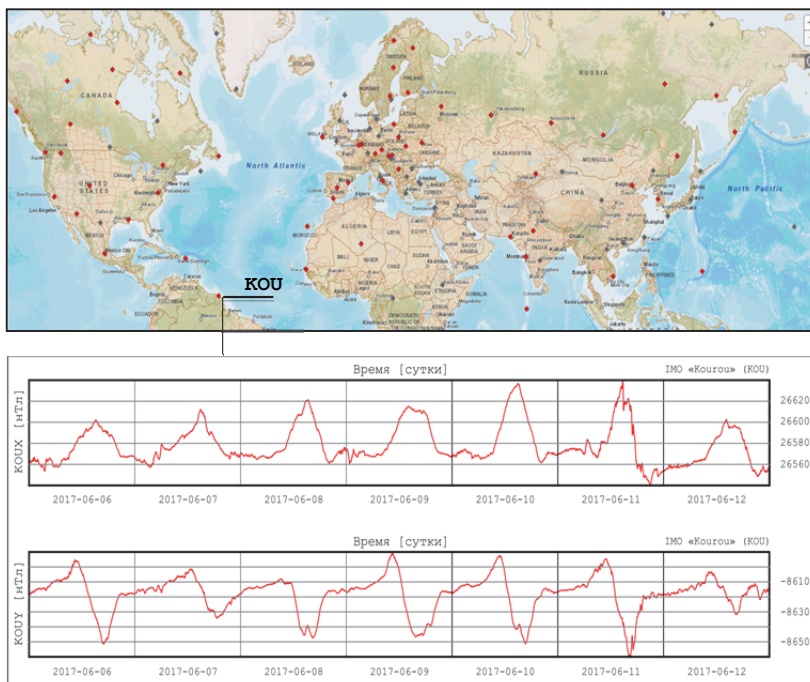


Рис. 5. Экранная форма сервиса «Магнитные обсерватории» по выбору магнитной станции и ее данных

Для выбранного источника геомагнитных данных информационной системой формируется график амплитудно-временного анализа

результатов наблюдений, по умолчанию ориентированный на неделю от заданной временной метки. Временной период можно указать, воспользовавшись соответствующими полями ввода на экранной форме, нагруженными функцией элемента-календаря.

Кроме этого, сервис включает в себя набор инструментов цифровой обработки геомагнитных данных, основанных на классических механизмах цифровой обработки сигналов (ЦОС). Среди них:

- линейная, нелинейная и адаптивная фильтрация информационного сигнала, обеспечивающая как подавление шума и селекцию сигнала в частотной области, так и анализ корреляций параметров смежных информационных сигналов;

- анализ сигнала во временной области, обеспечивающий расчет таких параметров информационного сигнала, как максимальное, минимальное и среднее значения, а также дисперсия и среднеквадратическое отклонение;

- спектральный и частотно-временной анализ, реализуемый как посредством исследования периодограммы, то есть за счет оценки спектральной плотности мощности, основанной на вычислении квадрата модуля преобразования Фурье последовательности данных с использованием статистического усреднения информационного сигнала, так и посредством оценки его вейвлет-скалограммы.

Указанные информационные сервисы реализованы посредством описанного выше набора технологий. Сервис «Магнитный калькулятор» размещен на внешнем сервере и находится на этапе тестирования. Остальные сервисы реализованы в виде исследовательского прототипа и проходят стадию отладки и тестирования на локальном виртуальном сервере.

Проведенные экспериментальные исследования результата реализации предложенной концепции единого пространства геомагнитных данных с применением рассмотренного набора технологий (п. 5) показали, что:

- скорость обработки больших объемов геомагнитных данных за счет применения многопоточного режима увеличивается в число раз, кратное вычислительной мощности сервера;

- аппаратные требования информационной системы в контексте размеров дискового пространства, физически занимаемого гибридным хранилищем, существенно сокращены (более чем в 2,3 раза по сравнению с CSV-представлением) за счет использования иерархического формата больших данных HDF5.

7. Заключение. Всевозрастающая потребность в своевременных достоверных геомагнитных данных предъявляет новые требования

к техническим средствам их сбора, обработки, передачи и анализа. Существующие сервисы лишь частично решают такие задачи, фокусируясь на отдельных магнитных сетях, обсерваториях или станциях. Вместе с тем реалии таковы, что для решения научно-практических задач необходима интеграция геомагнитных данных, получаемых из множества гетерогенных распределенных источников. Кроме того, проблема усугубляется непрерывно возрастающими объемами геомагнитной информации и обнаруживает низкую эффективность традиционных методов и средств их хранения, обработки и анализа.

В статье предложена концепция единого пространства геомагнитных данных, основанная на сочетании моделей консолидации и федерализации при интеграции данных и отличающаяся наличием гибридного информационного хранилища, заполняемого данными по выполнению процесса-демона, предварительной обработкой для повышения качества хранимых временных рядов, иерархическим форматом для снижения вычислительных затрат на размещение данных, модульной структурой и облачными вычислениями. Применение указанной концепции позволит:

1. Обеспечить конечным пользователям единый доступ к геомагнитным данным, предоставляемым гетерогенными источниками, без необходимости обращения к несвязанным информационным ресурсам, а также без избыточной предварительной загрузки разнородных данных с их последующей интеграцией для выполнения аналитических операций. Известные решения в области обработки геомагнитных данных лишь частично решают такую задачу, например, в рамках проекта SuperMag доступны только данные вариационных станций, а информационные ресурсы отдельных станций, обсерваторий и магнитных сетей предоставляют результаты суточных геомагнитных наблюдений, физически представленных в отдельных текстовых файлах.

2. Формировать и хранить полные временные ряды геомагнитных данных, не содержащие аномальных значений. При этом использование ранее предложенного авторами метода обеспечивает ошибку восстановления временного ряда в пределах погрешности геомагнитных измерений, определенной международной ассоциацией IAGA. Существующие решения по восстановлению временных рядов геомагнитных данных не обеспечивают приемлемого значения ошибки и применимы только для устранения единичных пропусков. Кроме того, один из немногих информационных проектов по геомагнитным наблюдениям SuperMag предусматривает удаление обнаруженных во временных рядах аномальных значений, что негативно сказывается на результатах аналитической обработки геомагнитных данных.

3. Сократить вычислительные затраты на хранение и обработку больших объемов геомагнитных данных за счет использования иерархического формата HDF5 и технологий облачных вычислений. В настоящее время массивы геомагнитных данных хранятся в виде набора текстовых файлов и доступны, как правило, по протоколу FTP. Специфика протокола и физическое разделение результатов наблюдений обуславливают низкую вычислительную скорость их загрузки и обработки. Так, к примеру, получение результатов годовых геомагнитных наблюдений с ресурса сети INTERMAGNET и их последующая аналитическая обработка требуют в среднем 4.5 минуты процессорного времени, что с учетом требований к эргономике программного обеспечения (время ожидания отклика не должно превышать 3-5 с) неприемлемо. Предложенная архитектура и стек технологий успешно решают поставленную задачу, сокращая затраты процессорного времени до допустимых пределов.

Перечисленное в совокупности позволит повысить как доступность разнородных результатов наблюдений за параметрами магнитного поля Земли и его вариаций, так и вычислительную скорость сбора и обработки больших объемов геомагнитных данных.

Предложен набор инфокоммуникационных технологий, обеспечивающих эффективную реализацию предложенной концепции. Ожидается, что веб-ориентированная реализация (клиент-серверные приложения по принципам PWA) концепции единого пространства существенно расширит круг потребителей геомагнитных данных, а вычислительные технологии, предназначенные для хранения и обработки больших объемов данных (формат HDF5, облачная инфраструктура и вычисления), позволят обеспечить высокую скорость и сравнительно небольшие вычислительные затраты на выполнение процедур интеграции разнородных геомагнитных данных по запросу потребителя. Кроме того, предлагаемый набор сервисов обеспечит потребителей новыми результатами, позволяющими получить дополнительную информацию из геомагнитных данных без применения дополнительных сторонних средств и систем.

Разработан исследовательский прототип информационной системы, включающий в себя ряд веб-ориентированных сервисов, которые обеспечивают:

- расчет параметров невозмущенного геомагнитного поля в точке с пространственно-временными координатами, заданными посредством геоинформационных технологий геолокации, прямого и обратного геокодирования;

- визуализацию геомагнитных данных с использованием платформонезависимой программируемой веб-ориентированной графики, обеспечивающей возможность переключения между двумерным (2D),

псевдотрехмерным (2.5D) и трехмерным (3D) форматами графического представления распределения параметров геомагнитного поля по земной поверхности;

– получение и амплитудно-временной анализ временных рядов геомагнитных данных, регистрируемых одной или несколькими магнитными обсерваториями и вариационными станциями на протяжении заданного временного интервала.

Ряд сервисов информационной системы размещен на внешнем веб-сервере с открытым физическим доступом, а остальные проходят стадии интеграционного тестирования и отладки на локальном виртуальном сервере. Дальнейшие исследования предполагают полное размещение сервисов на внешнем ресурсе и последующее развитие функциональности, заявленной при описании предложенной структуры единого пространства геомагнитных данных.

Литература

1. *Воробьев А.В., Воробьева Г.П.* Подход к оценке относительной информационной эффективности магнитных обсерваторий сети INTERMAGNET // Геомагнетизм и аэрономия. 2018. Т. 58. № 5. С. 648–652.
2. *St-Louis B.J. et al.* Intermagnet Technical Reference Manual // The INTERMAGNET office. Retrieved from the World Wide Web. 2011. 100 p.
3. *Love J.J., Chulliat A.* An International Network of Magnetic Observatories // EOS Transactions: American Geophysical Union. 2013. vol. 94. no. 42. pp. 373–374.
4. *Sandhu J.K. et al.* Variations of high latitude geomagnetic pulsation frequencies: A comparison of time of flight estimates and IMAGE magnetometer observations // Journal of Geophysical Research Space Physics. 2018. vol. 123. no. 1. pp. 567–586.
5. *Connors M. et al.* The AUTUMNX magnetometer meridian chain in Québec, Canada // Earth, Planets and Space. 2016. vol. 68. no. 1. pp. 2.
6. *Hughes W.J., Engebretson M.J.* MACCS: Magnetometer array for cusp and cleft studies // Satellite – Ground Based Coordination Sourcebook. 1997. vol. 1198. pp. 119.
7. *Reay S.J. et al.* Magnetic Observatory Data and Metadata: Types and Availability // Geomagnetic Observations and Models. 2011. pp. 149–181.
8. *Холутов С.Ю.* Обработка магнитных данных на обсерваториях // ИКИР ДВО РАН. 2017. 114 с.
9. *Анисимов С.В. и др.* Информационные технологии в геомагнитных измерениях на геофизической обсерватории Борок // Геофизические исследования. 2008. Т. 9. № 3. С. 62–76.
10. *Золотов С.Ю., Додолин Е.Л.* Методологические основы распределенной информационной системы мониторинга состояния окружающей среды // Доклады ТУСУРа. 2010. № 1(21). С. 203–206.
11. *Кокорев В.А., Шерстюков А.Б.* О метеорологических данных для изучения современных и будущих изменений климата на территории России // Арктика XXI век. Естественные науки. 2015. № 2. С. 5–23.
12. *Соловьев А.А. и др.* Новая геомагнитная обсерватория «Климовская» // Геомагнетизм и аэрономия. 2016. Т. 56. № 3. С. 342–354.
13. *Gvishiani A., Soloviev A.* Geoinformatic advances in geomagnetic data studies and Russian INTERMAGNET segment // Исследования по геоинформатике: труды Геофизического центра РАН. 2016. Т. 4. № 2. С. 8.
14. *Mandea M, Korte M.* Geomagnetic Observations and Models // IAGA Special Sopron Book Series. 2011. vol. 5. pp. 149–181.

15. *Dods J., Chapman S.C., Gjerloev J.W.* Network analysis of geomagnetic substorms using the SuperMAG database of ground-based magnetometer stations // Journal of Geophysical Research-Space Physics. 2015. vol. 120. pp. 7774–7784.
16. Концепция формирования и развития единого информационного пространства России и соответствующих государственных информационных ресурсов. URL: http://www.nsc.ru/win/laws/russ_kon.htm (дата обращения: 04.01.2019).
17. *Куяев В.О. и др.* Варианты построения единого информационного пространства для интеграции разнородных автоматизированных систем // Информация и космос. 2015. № 4. С. 83–87.
18. Методы интеграции данных в информационных системах. URL: <http://www.ipras.ru/articles/kogalov10-05.pdf> (дата обращения: 04.01.2019).
19. *Воробьев А.В., Воробьева Г.П.* Индуктивный метод восстановления временных рядов геомагнитных данных // Труды СПИИРАН. 2018. № 57. С. 104–133.
20. *Teixeira C. et al.* The Building Blocks of a PaaS // Journal of Network and Systems Management. 2014. vol. 22. pp. 75.
21. *Van Eyk E. et al.* Serverless is More: From PaaS to Present Cloud Computing // IEEE Internet Computing. 2018. vol. 22. no. 5. pp. 8–17.
22. The HDF Group. Hierarchical Data Format, version 5. URL: <http://www.hdfgroup.org/HDF5/> (дата обращения: 04.01.2019).
23. *Miyazaki R., Matsuzaki K., Sato D.* A Generator of Hadoop MapReduce Programs that Manipulate One-dimensional Arrays // Journal of Information Process. 2017. vol. 25. pp. 841–851.
24. *Воробьев А.В., Воробьева Г.П.* Веб-ориентированная 2D/3D-визуализация параметров геомагнитного поля и его вариаций // Научная визуализация. 2017. Т. 9. № 2. С. 94–101.
25. *Yusupova N. et al.* Web-based solutions in modeling and analysis of geomagnetic field and its variations // Proceedings of REMS 2018 – Russian Federation & Europe Multidisciplinary Symposium on Computer Science and ICT. URL: <http://ceur-ws.org/Vol-2254/10000282.pdf> (дата обращения: 04.01.2019).

Воробьев Андрей Владимирович — канд. техн. наук, доцент, кафедра геоинформационных систем факультета информатики и робототехники, Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: геоинформационные технологии, цифровая обработка сигналов. Число научных публикаций — 147. geomagnet@list.ru; К.Маркса, 12, 450007, Уфа, Российская Федерация; р.т.: +7(917)345-2299.

Воробьева Гульнара Равилевна — канд. техн. наук, доцент, кафедра вычислительной математики и кибернетики факультета информатики и робототехники, Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: геоинформационные и веб-технологии, системы хранения и обработки информации. Число научных публикаций — 121. gulnara.vorobeva@gmail.com; К. Маркса, 12, 450008, Уфа, Российская Федерация; р.т.: +7(917)417-411.

Юсупова Нафиса Исламовна — д-р техн. наук, профессор, заведующий кафедрой, кафедра вычислительной математики и кибернетики факультета информатики и робототехники, Уфимский государственный авиационный технический университет (УГАТУ). Область научных интересов: интеллектуальные методы обработки информации и управления с приложениями в социальных, экономических и технических системах. Число научных публикаций — 566. yussupova@ugatu.ac.ru; К.Маркса, 12, 450007, Уфа, Российская Федерация; р.т.: +7 (347)273-7717.

Поддержка исследований. Работа частично выполнена при финансовой поддержке РФФИ (проект № 18-07-00193-а).

A.V. VOROBEV, G.R. VOROBEVA, N.I. YUSUPOVA
CONCEPTION OF GEOMAGNETIC DATA INTEGRATED SPACE

Vorobev A.V., Vorobeva G.R., Yusupova N.I. Conception of Geomagnetic Data Integrated Space.

Abstract. As is known, today the problem of geomagnetic field and its variations parameters monitoring is solved mainly by a network of magnetic observatories and variational stations, but a significant obstacle in the processing and analysis of the data thus obtained, along with their spatial anisotropy, are omissions or reliable inconsistency with the established format. Heterogeneity and anomalousness of the data excludes (significantly complicates) the possibility of their automatic integration and the application of frequency analysis tools to them. Known solutions for the integration of heterogeneous geomagnetic data are mainly based on the consolidation model and only partially solve the problem. The resulting data sets, as a rule, do not meet the requirements for real-time information systems, may include outliers, and omissions in the time series of geomagnetic data are eliminated by excluding missing or anomalous values from the final sample, which can obviously lead to both to the loss of relevant information, violation of the discretization step, and to heterogeneity of the time series. The paper proposes an approach to creating an integrated space of geomagnetic data based on a combination of consolidation and federalization models, including preliminary processing of the original time series with an optionally available procedure for their recovery and verification, focused on the use of cloud computing technologies and hierarchical format and processing speed of large amounts of data and, as a result, providing users with better and more homogeneous data.

Keywords: Geomagnetic Data, Magnetic Observatories, Time Series, Big Data, Integrated Information Space, Parallel Computing.

Vorobev Andrei Vladimirovich — Ph.D., Associate Professor, Geoinformation Systems Department of Computer Science and Robotics Faculty, Ufa State Aviation Technical University (USATU). Research interests: geoinformation technologies, digital signal processing. The number of publications — 147. geomagnet@list.ru; 12, K. Marx, 450007, Ufa, Russian Federation; office phone: +7(917)345-2299.

Vorobeva Gulnara Ravilevna — Associate Professor, Computational Mathematics and Cybernetics Department of Computer Science and Robotics Faculty, Ufa State Aviation Technical University (USATU). Research interests: geoinformation and web technologies, systems of information storing and processing. The number of publications — 121. gulnara.vorobeva@gmail.com; 12, K. Marx, 450008, Ufa, Russian Federation; office phone: +7(917)417-411.

Yusupova Nafisa Islamovna — Ph.D., Dr.Sci., Professor, Head of Department, Computational Mathematics and Cybernetics Department of Computer Science and Robotics Faculty, Ufa State Aviation Technical University (USATU). Research interests: intelligent methods of information processing and management with applications in social, economic and technical systems. The number of publications — 566. yussupova@ugatu.ac.ru; 12, K.Marx, 450007, Ufa, Russian Federation; office phone: +7 (347)273-7717.

Acknowledgements. This research is partially supported by RFBR (grant 18-07-00193-a).

References

1. Vorobev A.V., Vorobeva G.R. [Approach to Assessment of the Relative Informational Efficiency of Intermagnet Magnetic Observatories]. *Geomagnetizm i aeronomia – Geomagnetism and Aeronomy*. 2018. Issue 58. vol. 5. pp. 648–652. (In Russ.).

2. St-Louis B.J. et al. Intermagnet Technical Reference Manual. The INTERMAGNET office. Retrieved from the World Wide Web. 2011. 100 p.
3. Love J.J., Chulliat A. An International Network of Magnetic Observatories // *EOS Transactions: American Geophysical Union*. 2013. vol. 94. no. 42. pp. 373–374.
4. Sandhu J.K. et al. Variations of high latitude geomagnetic pulsation frequencies: A comparison of time of flight estimates and IMAGE magnetometer observations. *Journal of Geophysical Research Space Physics*. 2018. vol. 123. no. 1. pp. 567–586.
5. Connors M. et al. The AUTUMNX magnetometer meridian chain in Québec, Canada. *Earth, Planets and Space*. 2016. vol. 68. no. 1. pp. 2.
6. Hughes W.J., Engebretson M.J. MACCS: Magnetometer array for cusp and cleft studies. Satellite – Ground Based Coordination Sourcebook. 1997. vol. 1198. pp. 119.
7. Reay S.J. et al. Magnetic Observatory Data and Metadata: Types and Availability. *Geomagnetic Observations and Models*. 2011. pp. 149–181.
8. Khomutov S.Yu. *Obrabotka magnitnykh dannykh na observatorijah* [Magnetic data processing at observatories]. IKIR FEB RAS. 2017. 114 p. (In Russ.).
9. Anisimov S.V. et al. [Information technology in geomagnetic measurements at the Borok geophysical observatory]. *Geofizicheskie issledovaniya – Geophysical research*. 2008. vol. 9. no. 3. pp. 62–76. (In Russ.).
10. Zolotov S.Yu., Dodolin E.L. [Methodological basis of a distributed information system for monitoring the state of the environment]. *Doklady TUSURa – TUSUR reports*. 2010. vol. 1(21). pp. 203–206. (In Russ.).
11. Kokorev V.A., Sherstukov A.B. [About meteorological data to study current and future climate change in Russia]. *Arktika XXI vek. Yestestvennyye nauki – Arctic XXI century. Natural Sciences*. 2015. vol. 2. pp. 5–23. (In Russ.).
12. Soloviev A.A. et al. [New geomagnetic observatory "Klimovskaya"]. *Geomagnetizm i aeronomiya – Geomagnetism and Aeronomy*. 2016. Issue 56. vol 3. pp. 342–354. (In Russ.).
13. Gvishiani A., Soloviev A. [Geoinformatic advances in geomagnetic data studies and Russian INTERMAGNET segment]. *Issledovaniya po geoinformatike: trudy Geofizicheskogo centra RAN – Research in geoinformatics: the works of the Geophysical Center RAS 4*. vol. 2. pp. 8.
14. Manda M, Korte M. Geomagnetic Observations and Models. IAGA Special Sopron Book Series. 2011. vol. 5. pp. 149–181.
15. Dods J., Chapman S.C., Gjerloev J.W. Network analysis of geomagnetic substorms using the SuperMAG database of ground-based magnetometer stations. *Journal of Geophysical Research-Space Physics*. 2015. vol. 120. pp. 7774–7784.
16. Konceptsiya formirovaniya i razvitiya edinogo informacionnogo prostranstva Rossii i sootvetstvujushhih gosudarstvennykh informacionnykh resursov [The concept of the formation and development of a single information space of Russia and the relevant state information resources]. Available at: http://www.nsc.ru/win/laws/russ_kon.htm (accessed: 04.01.2019). (In Russ.).
17. Kuvaev V.O. et al. [Variants of building a single information space for the integration of heterogeneous automated systems]. *Informacija i kosmos – Information and space*. 2015. vol. 4. pp. 83–87. (In Russ.).
18. Metody integracii dannykh v informacionnykh sistemah [Methods of data integration in information systems]. Available at: <http://www.ipr-ras.ru/articles/kogalov10-05.pdf> (accessed: 04.01.2019). (In Russ.).
19. Vorobei A.V., Vorobeva G.R. [Inductive method of geomagnetic data time series recovery]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2018. vol. 57. pp. 104–133. (In Russ.).
20. Teixeira C. et al. The Building Blocks of a PaaS. *Journal of Network and Systems Management*. 2014. vol. 22. pp. 75.

21. Van Eyk E. et al. Serverless is More: From PaaS to Present Cloud Computing. *IEEE Internet Computing*. 2018. vol. 22. no. 5. pp. 8–17.
22. The HDF Group. Hierarchical Data Format, version 5. Available at: <http://www.hdfgroup.org/HDF5/> (accessed: 04.01.2019).
23. Miyazaki R., Matsuzaki K., Sato D. A Generator of Hadoop MapReduce Programs that Manipulate One-dimensional Arrays. *Journal of Information Process*. 2017. vol. 25. pp. 841–851.
24. Vorobev A.V., Vorobeva G.R. [Web-based 2D / 3D visualization of geomagnetic field and its variations parameters]. *Nauchnaja vizualizacija – Scientific Visualization*. 2017. Issue 9. vol. 2. pp. 94–101. (In Russ.).
25. Yusupova N. et al. Web-based solutions in modeling and analysis of geomagnetic field and its variations. Proceedings of REMS 2018 – Russian Federation & Europe Multidisciplinary Symposium on Computer Science and ICT. Available at: <http://ceur-ws.org/Vol-2254/10000282.pdf> (accessed: 04.01.2019).