



УДК 550.3; 523.9; 004.031; 004.032; 004.05

ПРИНЦИПЫ РАЗРАБОТКИ ИНФОРМАЦИОННЫХ СИСТЕМ СБОРА, ОБРАБОТКИ И ХРАНЕНИЯ ГЕЛИОГЕОФИЗИЧЕСКИХ ПРОСТРАНСТВЕННО-РАСПРЕДЕЛЕННЫХ ДАННЫХ В ИНТЕРЕСАХ ПОИСКА РЕШЕНИЙ НАУЧНЫХ И ПРИКЛАДНЫХ ЗАДАЧ

О. В. Никифоров¹, А. К. Кузьмин¹, А. М. Мёрзлый¹, В. Ю. Попов^{1,2}, И. А. Шагурин^{1,2}, А. Т. Янаков¹¹Институт космических исследований РАН (ИКИ РАН), г. Москва, Россия²Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), г. Москва, Россия

В статье рассматриваются подходы к разработке информационной системы в области сбора, обработки, хранения и анализа данных, а также представлены существующие решения этих задач. Описан ряд факторов, влияющих на программную совместимость таких систем. Рассмотрены технологические особенности задач обработки пространственных данных, а также основы проектирования структуры ИАС. Проанализированы программно-архитектурные направления при создании ИАС, в том числе с анализом типов и схем проектирования архитектур систем, принципов сбора и обработки пространственно-распределенных данных, схем проектирования топологии поступления данных, языков программирования и других технологических факторов, влияющих на архитектуру ИАС. В выводах сформулированы принципы проектирования и разработки ИАС, использующих гелиогеофизические пространственно-распределенные данные, т.е. сведения, показатели, факты, полученные с научных и других сенсоров/приборов, выраженные как в числовой, так и в любой другой форме.

КЛЮЧЕВЫЕ СЛОВА: ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЕ СИСТЕМЫ, ГЕОИНФОРМАЦИОННЫЕ СИСТЕМЫ, СБОР, ОБРАБОТКА, ХРАНЕНИЕ, АНАЛИЗ, ПРОСТРАНСТВЕННО-РАСПРЕДЕЛЕННЫЕ ДАННЫЕ

EDN: UGOIZI

1. ВВЕДЕНИЕ

В рамках серии работ, посвященных характеристикам информационно-аналитической системы (ИАС), нацеленной на систематический мониторинг и контролирование условий распространения радиосигналов коротковолнового диапазона, оценки и прогнозирования условий работы радиотехнических средств в полярной ионосфере [Кузьмин и др., 2021] создаются ИАС «Гелиогеофизика» и геоинформационная система (ГИС) «Аврора-Арктика» [Петрукович и др., 2017; Никифоров и др., 2018; Умрихин и др., 2019].

Стимулом для авторов к конкретизации целей и обозначению задач разработки информационных систем сбора, обработки и хранения гелиогеофизических пространственно-распределенных данных послужила необходимость построения эффективной коммуникации научного сообщества с пользователями результатов научных исследований.

Целью создания ИАС ИКИ ГГФ является информационное обеспечение потребителей системы магнитосфера-ионосфера-термосфера. Данные о структурах эмиссий, наблюдавшихся в ионосфере и

Электронная почта авторов для переписки:

Никифоров Олег Вячеславович, e-mail: o.nikiforov@cosmos.ru
Кузьмин Александр Константинович, e-mail: alkuzmin@mail.ru
Мёрзлый Алексей Михайлович, e-mail: pinegal42@yandex.ru
Попов Виктор Юрьевич, e-mail: masterlu@mail.ru
Шагурин Иван Александрович, e-mail: iashagurin@mail.ru
Янаков Альберт Триандофилович, e-mail: alexander.7@mail.ru

<https://elibrary.ru/ugoizi>Адрес редакции журнала
«Гелиогеофизические исследования»:

ФГБУ «ИПГ»
129128; Россия, Москва
ул. Ростокинская, 9.
e-mail: vestnik@ipg.geospace.ru

верхней атмосфере, в первую очередь в полярной зоне, специалисты используют для решения научных и прикладных задач, для разработки и внедрения методов распознавания авроральных структур различных масштабов в полученных изображениях, типов энергетических спектров высыпающихся частиц, классификации, кластеризации, и их интеграции с другими ГИС и базами данных, разрабатываемыми в ИКИ РАН [Никифоров и др., 2024].

Задачами являются:

- обеспечение гелиогеофизическими данными, полученными с наземных станций, орбит КА (космических аппаратов), платформ, отделяемых от зондирующих ракет, БПЛА (беспилотных летательных аппаратов), тематических серверных систем и пр.;
- разработка универсального стенда для тестирования моделей ионосферы;
- решение научных и прикладных задач на основе эмпирических и теоретических моделей ионосферы, распространения радиоволн, характеристик аврорального овала и др.;
- модельное сопровождение анализа экспериментальных данных и подбор эффективных моделей в конкретных областях и конкретных условиях;
- использование в образовательной деятельности (подготовка студентов и аспирантов);
- решение задач национальной экономики (обеспечение устойчивой радиосвязи, позиционирования и пеленгации в арктической зоне РФ для госслужб, туристической отрасли и промышленности);
- популяризация науки [Никифоров, 2021].

Одним из важнейших критериев актуальности и значимости научных исследований, мониторинга и прогнозирования является востребованность результатов. В настоящее время широко распространены и продолжают совершенствоваться и развиваться средства и методы наблюдения, обработки и анализа данных. С разных орбит КА, платформ зондирующих ракет и поверхности земли собраны, накоплены, структурированы зеттабайты пространственно-распределенных данных и сопутствующей информации, что представляет интерес для дальнейшего анализа, осмысления и интерпретации.

В соответствии с определением Клода Элвуда Шеннона информация — это субстанция, которая увеличивает степень знания человека об интересующем объекте и, соответственно, снижает степень неопределенности (энтропии) знаний об этом объекте [Shannon, 1993]. Информация — это многозначный феномен. Основными характеристиками информации являются достоверность, полнота, актуальность, ясность, ценность и ряд других. Первой из научных теорий информации стала математическая, или вероятностно-статистическая, концепция. В ней уделяется особое внимание количественным аспектам и отождествляются термины «информация» и «количество информации». Понятие информации относится к передаче, приему и обработке каких-либо сведений или сигналов. Информация о солнечно-земных процессах необходима для анализа конкретных данных и получения результатов их интерпретации. Именно поэтому важно дальнейшее развитие подходов, методик и решений доведения данных и информации до конкретных пользователей, в конечном счете и до общества в целом.

2. ОСНОВЫ ПРИ РАЗРАБОТКЕ СИСТЕМЫ СБОРА, СТРУКТУРЫ И ОБРАБОТКИ, ХРАНЕНИЯ И АНАЛИЗА ДАННЫХ

Системы обработки пространственно-распределенных данных (СОПД) представляют собой информационные системы с обязательным пространственным распределением наблюдаемого предмета, в которых использование пространственно-распределенных данных обеспечивает решение задач, возлагаемых на систему.

Системы сбора, обработки, хранения и анализа (СОХА) можно условно разделить на три большие группы (рис. 1): универсальные, специальные и комбинированные.

Универсальные СОХА (ГИС, системы автоматизированного проектирования и др.) характеризуются принципами мультиприменения, т.е. не ограничены областью применения, выбором методик выполнения работ, и, как правило, нацелены на решение унифицированных задач.

В специальных СОХА реализуется противоположный подход – персонализированный, т.е. подразумевается создание программно-аппаратных комплексов под решение конкретных узкоспециализированных задач.



Рис. 1. Классификация систем сбора, обработки, хранения и анализа

Создание систем СОХА при универсальном подходе подразумевает быструю реализацию новых функциональных возможностей в практически любых областях деятельности.

Отличие универсального подхода в контексте адаптивности подразумевает использование существующих готовых решений (как собственных, так и внешних) для решения наиболее приближенных к предметной области задач, а также минимальность трудовых и вычислительных затрат.

Специальный подход подразумевает максимально конкретизированное техническое описание решаемых задач и требуемого функционала. Это описание становится предметом разработки с помощью создания новой СОХА или полной модернизации разработанных ранее комплексов программ.

Существующие готовые программные решения СОХА, как правило, имеют принципиальные ограничения применимости и совместимости. Это связано с учётом решаемых задач, технологических факторов, значительных трудозатрат на преодоление программно-аппаратных преград, не связанных с решением конкретных задач, а также с учетом мест разработки таких готовых решений. Эти и другие ограничения вынуждают специалистов в конкретной тематике идти по пути формирования комбинированного подхода, который заключается в использовании готовых решений, дополненных создаваемыми программными комплексами, позволяющими решать узкоспециальные задачи.

3. СУЩЕСТВУЮЩИЕ РЕШЕНИЯ В ОБЛАСТИ СОХА ПРОСТРАНСТВЕННО-РАСПРЕДЕЛЕННЫХ ДАННЫХ, И РЯД ФАКТОРОВ, ВЛИЯЮЩИХ НА ИХ ПРОГРАММНУЮ СОВМЕСТИМОСТЬ

При конструировании информационно-аналитической системы (ИАС) разумно применять принцип использования программных компонент и комплексов (ПКиК) российской и/или зарубежной свободно распространяемой разработки, наиболее близкой к предметной области разрабатываемой системы. Это позволяет рационально использовать трудовые, временные и вычислительные ресурсы.

Подход многократного использования программных компонентов и комплексов, часто заменяющий собой более классические каскадный и итеративный [Власов, 2015; Первухин и др., 2020] подходы, является основным и наиболее используемым при проектировании и конструировании программного обеспечения (ПО) в целом, ИАС и СОХА в частности. Например, концепция RAD (rapid application development — быстрая разработка приложений) при разработке программного обеспечения позволяет получить быстрый результат при минимально допустимых временных и финансовых ресурсах, а также при недостаточно проработанном техническом задании на разработку ПО. Эффективной эта концепция становится при постоянном взаимодействии с заказчиком, использовании уже разработанных ПКиК, а также автоматизации операций, использовании визуального программирования. «Быстрая разработка» позволяет получить предсказуемый и воспроизводимый результат, обеспечить компактность и выразительность кода ПКиК, возможность его доработки с целью обеспечения конкретных требований технического задания (ТЗ). При использовании этой концепции многократного использования ПКиК в целом, часто в организации-заказчике разработчик ТЗ и тематический пользователь (интересант) — это разные лица, а уже созданным ПКиК могут потребоваться слишком масштабная доработка, которая становится нецелесообразной по сравнению с разработкой нового кода. Наиболее заметны эти особенности в научной, инновационной деятельности и других подобных, в которых требуется нестандартность реализации новых замыслов, идей и проектов [Кнут, 2006; Larman, 2006; Cohn, 2011].

К оценке возможности вторичного использования ПКиК необходимо подходить предельно взвешено, иначе плюсы конкретного решения могут становиться минусами. Оценка возможностей

вторичного использования ПКК, в частности для создания ИАС, должна учесть возможности объединения двух и более ПКК путем анализа целей и задач предполагаемых работ, приоритетных характеристик, основополагающих требований к компонентам и программным комплексам, факторов, влияющих на весь жизненный цикл разработки, например, по принципам SDM (Software Development Methodology). Кратко представить последовательность, предусмотренную SDM, следует с помощью жизненного цикла разработки программного обеспечения (software development life cycle, SDLC): планирование (planning) – анализ (analysis) – дизайн (design) – выполнение (implementation) – тестирование и интеграция (testing & integration) – обслуживание (maintenance). SDLC общепринятый процесс разработки, который является международным и национальным стандартом (ISO/IEC 12207:2008; ГОСТ Р ИСО/МЭК 12207-2010) системной и программной инженерии.

3.1. Технологические особенности задач обработки пространственно-распределенных данных

Тематические пространственно-распределенные данные (ТПД) представляют собой качественное или количественное отражение изучаемого явления, процесса, события или феномена с обязательными пространственно-временными характеристиками.

Виды представления ТПД характеризуются большим разнообразием. В самом общем виде корректно разделить их на 2 группы (рис. 2): растровую и векторную [Капралов, 2010].

Работа с растровым изображением обуславливается следующими характеристиками: размером отображаемого пространства, разрешения, количества каналов, цветовой палитрой и разрядностью, способами кодирования.

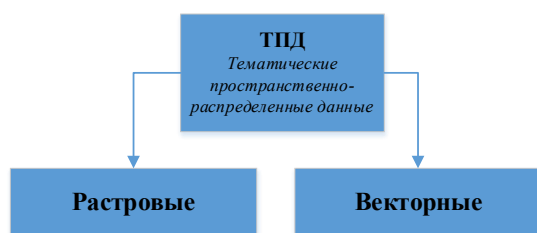


Рис. 2. Группы тематических пространственно-распределенных данных

При работе с векторным изображением характеристики иные: морфология объектов, их количество и внутренние связи между ними, а также количество и виды используемых систем координат и требуемая точность географической, геомагнитной или иной привязки. Кодирование растровых данных складывается из ТПД, которые представляют собой двоичный поток, и метаданных, объединенных с ТПД GeoTIFF [QGIS...; GRASS GIS...], либо представленных в отдельном потоке (например, ENVI Header).

Морфология векторных данных значительно сложнее растровых, т.к. активно применяемых вариантов кодирования свыше 50. Применение каждого обуславливается структурой данных, свойствами применяемых алгоритмов и особенностью задач [QGIS...; GRASS GIS...]. Часто в пространственных характеристиках векторных данных также содержится взаимное пространственное расположение точечных, линейных и полигональных объектов [Warmerdam et al., 1998; gvSIG ...]. В пакетных геоинформационных системах (QGIS, ArcGIS и прочие) применяются собственно разработанные форматы, например, WKT (WKB) по стандарту OGC [OGC...] и Shapefile (.shp, .shx, .dbf, .sbn) [ESRI...], развиваемому компанией ESRI.

В процессе исследования ТПД, их взаимосвязей, соотношений и характеристик, атрибутивные данные (АД) неотделимы от пространственно-временных и должны регулярно актуализироваться. Задачи актуализации решаются с использованием систем управления базами данных (СУБД), в которых реализован функционал работы с пространственно-распределенными данными (PostgreSQL и PostGIS, Oracle Spatial, Microsoft SQL Server, MySQL/MariaDB и др.) [ORACLE...; PostGIS...]. Одновременная обработка пространственно-распределенных и атрибутивных данных является преимуществом благодаря сохранению целостности массива данных, при этом некоторые СУБД содержат специальные решения для обработки специфичных ТПД в сочетании с АД, а также предлагают пользователям специфический функционал. Например, работу со структурами данных с разбиением пространства для упорядочивания точек в k-мерном пространстве (KD-tree), R-дерево, R+-дерево, R*-дерево и т.п. [Rigaux et al., 2002] [Шикин и Боресков, 2020]. Недостаток совместной обработки ТПД и АД без учёта их особенностей приводит к снижению эффективности решения таких задач, как визуализация. Альтернативным решением является раздельная обработка ТПД и АД, что позволяет обеспечить большую эффективность и функциональность доступа к данным, но приводит к усложнению структуры хранения. Кроме того, раздельная обработка ТПД и АД ведет к проблемам целостности архива и взаимной между ТПД и АД ссылочной системы, особенно при обновлении

массивов данных, например, при работе с постоянным потоком данных или при пересмотре структуры данных.

Результатом работы любой ГИС, СОХА, СОПД является представление данных и информации конечному пользователю с обязательным эффективным набором функциональных возможностей интерфейса, интуитивно понятным, гибким и персонализированным, что позволяет сократить время получения, понимания и, при необходимости, принятия решения. Решение задач визуализации требует глубокой технической проработки. Кратко перечислим основные общепринятые характеристики решения задач по визуализации:

- большие объемы данных;
- широкий диапазон масштабов отображения;
- множество систем координат исходных данных и предоставляемых данных;
- множество проекций;
- интерактивность (управление визуализацией);
- адаптивность к устройствам;
- производительность;
- совместимость с браузерами (в случае, если разрабатывается веб-приложение);
- сквозная аналитика (интеграция визуализаций с другими инструментами и системами аналитики);
- управление цветом и стилем, т.е. эффективное управление цветами и стилями визуализаций требует тщательного подхода, чтобы обеспечить четкое и понятное пользователю представление данных, особенно в случае цветовой карты и отображения значений.

Отдельно необходимо выделить выполнение стандартных и специфических пространственных операторов [Паутов, 2023] над полученными данными и проанализированной информацией. Стандартные операторы содержатся во множестве ГИС систем, согласно стандарту OGC [OGC...; Железняков, 2011], в котором они разделены на категории измерения, функций, предикатов, построения и пр.

Оптимальный выбор методов пространственной индексации [Шикин, 2020] и математической обработки данных и информации позволяет наиболее эффективно выполнять задачи визуализации, а также работы системы в целом.

Решения задач поддержки пересчета различных систем координат при решении задач визуализации являются востребованными и необходимыми. Кроме стандартных глобальных прямоугольных, геоцентрических, астрономических, географических, геомагнитных систем координат, часто нужна поддержка конкретных территориальных (региональных, местных) систем координат [Флейс, 2002; GISLAB...; Капралов и др., 2010]. Для корректной визуализации необходима не только возможность пересчёта систем координат, но и информация о методе пересчета, его точности, времени и множестве других параметров, набор которых формируется от конкретных решаемых задач, например, настройки цвета, параметров анимации, возможности группировать и фильтровать данные по различным параметрам, производительности отрисовки (рендеринга).

Собираемые данные часто характеризуется нелинейностью, уникальностью, специфичностью для каждой конкретной задачи. В зависимости от направления исследования, в идеале, цикл задается исходя из следующих свойств:

- пространственных;
- временных;
- вещественных (материя);
- энергетических;
- гравитационных;
- магнитных;
- взаимодействий: с физическими полями; по пространству.

Технологической основой ИАС обработки гелиогеофизических пространственно-распределенных данных являются специфические и уникальные наборы алгоритмов и структур общих и специальных данных. Эти наборы структурно сложны и требуют специальной квалификации для сбора, обработки, хранения и анализа, тестирования и сопровождения. Соответственно, при разработке таких ИАС или ГИС использование существующих готовых ПКК имеет низкую эффективность, и она еще более снижается при грамотной и профессиональной постановке конкретных задач, решаемых системой.

3.2. Подходы к проектированию структуры ИАС

Основные проектируемые структуры – логическая, тематическая и техническая (рис. 3), с последующей увеличивающейся подробностью каждой из них. Важнейшим решением является уровневая приоритезация структур, из которой рождается подход к проектированию. Например, тематическая структура в рамках проектирования ИАС с данными по гелиогеофизике, согласно принципам (см. Заключение), имеет высшую приоритетность. Она формирует логическую структуру, подчиняет техническую структуру и схему, создавая перфектный подход.



Рис. 3. Основные проектируемые структуры информационно-аналитических систем

Исходя из приоритетности тематической структуры, на её основе формируется каркас информационной системы. Этот каркас представляет собой физически обоснованный параметрический набор контролируемых внешних воздействующих гелиогеофизических и других характеристик, каждая из которых имеет пространственное, временное описание, описание состава, свойств и др.

Например, пространственное

описание характеристики включает такие позиции:

- общие: полный охват, площадь, высота «от» и «до»;
- внутривидовые (строение целого): типы составных частей, количество составных частей в общем объёме по типам, сегменты (рабочая зона), общая форма (геометрическая), формы составных частей, формы сегментов, плотность составных частей в общем объёме, плотность составных частей в сегменте, размер максимальной формы составной части, размер минимальной формы составной части, взаимное наложение / проникновение составных частей друг в друга и т.д.

Пространственное описание характеристик позволяет определить рабочие масштабы (площади и объёмы), параметры генерализации и др., в том числе те, которые подчиняют логическую и техническую структуры (формируя их), а также предопределяя параметры и характеристики доведения информации, например, визуализацию.

Временное описание характеристик включает, например:

- статический / динамический характер;
- циклы временных изменений физические;
- циклы временных изменений требуемые;
- максимальное значение;
- минимальное значение;
- максимальная амплитуда изменений по каждому циклу;
- минимальная амплитуда изменений по каждому циклу;
- необходимая частота измерения;
- фактическая частота измерения.

Такой подход можно назвать перфектным.

При создании ИАС по перфектным принципам предполагается исходить из главенства требуемых свойств и решений основной тематической задачи или задач, не учитывая дополнительных факторов (экономических, организационных, временных и пр.). Фактическая реализация подхода представляет собой максимально полное описание предмета в концепции с формированием требований к ИАС. При разработке технического задания возможна корректировка требований в сторону огрубления с обязательным условием сохранения максимально полного объёма требуемых свойств и решений задач с последующим выходом ИАС на максимальные характеристики. Такой подход подразумевает полное описание решаемых задач с помощью ИАС, который обеспечит достижение цели или целей без потерь в качестве на каждом из этапов работы ИАС. Например, проведение детального анализа характеристик всего спектра разных готовых решений с выводами о возможности их использования при обязательном сохранении всех ключевых свойств ИАС, совместимости и надёжности.

Формирование и разработка архитектуры ИАС на принципах оптимизации подразумевает корректировку требуемых свойств и решений с учётом рационального использования ресурсов,

сложности и скорости реализации, трудоёмкости, удобства разработки и эксплуатации, универсализации для дальнейшего использования при решении других задач и т.д.

Оптимальный или перфектный подход к проектированию архитектуры ИАС в корне меняет процесс разработки. В частности, пересматриваются первые 3 позиции проектирования (в соответствии с тематическим ГОСТ 34.601-90): формирование требований, разработка концепции, разработка технического задания.

Оптимизация разработки архитектуры ИАС в первую очередь подразумевает использование готовых программных продуктов для решения однотипных задач. Подбор готовых решений при оптимизации ведет к известным плюсам и минусам, показанным в таблице 1.

Таблица 1. Плюсы и минусы при выборе готовых решений

Плюсы	Минусы
Простота реализации	Шаблонность
Время реализации	Ограниченный набор реализаций
Низкая трудоёмкость	Выбор реализации на основании частного предпочтения
Удобство разработки	Слабая выраженность требуемых свойств и решений
Универсализм	Точечное применение

Например, создание пользовательского интерфейса часто выполняется с помощью готовых продуктов (MFC, WPF, Qt, GTK+, AWT, SWT, JavaFX и др.) [Эккель, 2013], представляющих собой наиболее удобные и широко распространенные пакеты программ. Это не позволяет учитывать внешние вводные разрабатываемых систем, такие как языки программирования, операционные системы в целом и т.п. Субъективность выбора конкретным разработчиком принципов, архитектуры и инструментов, исходя из прошлого опыта, квалификации и личных предпочтений, приводит к неоднозначному набору решений. В свою очередь, это влияет на детализацию решения задач ИАС, определяет её важнейшие свойства при реализации с помощью уже готовых программных решений, по сути, принимая и обеспечивая только программную совместимость, которая и становится главным фактором подбора.

4. ПРОГРАММНО-АРХИТЕКТУРНЫЕ НАПРАВЛЕНИЯ ПРИ СОЗДАНИИ ИАС

Проведение анализа программных решений при разработке архитектуры ИАС чаще всего учитывает взаимную связь функциональных частей системы. В зависимости от выбора принципов построения, система становится монолитной, крупноблочной или модульной. Монолитная система не позволяет независимое от других модулей использование части системы для решения части задач. Крупноблочная система позволяет выделять крупные задачи и решать их без использования системы в целом. Модульная архитектура системы позволяет обеспечить решение крупных, средних и мелких задач с обеспечением возможности максимально произвольного их выделения и замены (при необходимости).

Как правило, готовые программные решения закрыты и недоступны для глубокого изучения. Возможность их изучения заключается в анализе открытых характеристик и второстепенных (косвенных) факторов: опыта эксплуатации, отзывов сообщества о нём, полноты документации, обеспечения стандартов и т.п.

Выделим несколько основных путей программно-архитектурных направлений при создании ИАС:

- «путь Windows» – яркое направление монолитных систем, характеризующееся ориентацией на конечного пользователя, приоритетным решением прикладных задач, практическим отсутствием выделения и использования отдельных компонентов для включения во внешние системы, фактическим запретом использования вне объединяющего интерфейса, сложностями с переносимостью приложений [Рихтер, 2001; Гвоздев, 2016];
- «путь UNIX» подразумевает агрегирование модульной и крупноблочной архитектуры, что заметно снижает зависимость монолитного подхода, позволяет выбрать готовое или авторское решение для упрощения выполнения каждой отдельной функции и встроить его в стандартизированное общее пространство для выполнения конкретных задач/задачи. Данный подход обеспечивает максимальные возможности интеграции, компоновки и

построения систем, но требует высокой квалификации разработчиков, а также создаёт сложности поддержки при обновлениях, невозможность использования разработанных решений долгое время без актуализации [FlatBuffers...];

- «путь Java», основанный на языке программирования Java, а в настоящее время реализуем на других языках, например, Питон (Python), в системе языков C (Си), C++ (Си++). Он позволяет отстраниться от среды разработки, унифицировать решения под любую среду, и поэтому часто используется в виртуальных средах, что позволяет снизить последствия ошибок и угроз вирусов, но приводит к увеличенному потреблению ресурсов и сложности взаимодействия с внешними средами [Эккель, 2013].

4.1. Анализ типов и схем проектирования архитектур систем обработки пространственно-распределенных данных

Пространственно-распределенные данные – это большие массивы сходных по структуре данных, имеющих при этом ограниченное количество типов, что приводит к необходимости решения множества однотипных задач с разной периодичностью. В зависимости от целей и задач сбора пространственно-распределенных данных можно классифицировать типы архитектур СОГД, например, по методам обработки, схемам обработки и т.п.

На практике источниками получения гелиогеофизических данных в основном являются измерительные приборы и сторонние базы данных. Приборы для научных исследований разных производителей характеризуются разнородными принципами:

- получения тематических данных (типов, видов сенсоров);
- преобразования (АЦП, ЦАП);
- сбора, подбора форматов хранения данных;
- обеспечения пространственно-распределенного поля сбора данных;
- поддержки доступности каналов и протоколов передачи;
- и т.д.

В разнородных условиях планирование унифицированного подхода к схеме проектирования архитектуры системы сбора и обработки гелиогеофизических пространственно-распределенных данных не представляется возможным.

Сторонние базы данных (БД) не представляют сложных задач для сбора и обработки, т.к. практически во всех случаях позволяют получить массив данных в нужном виде, формате, частоте и пр., при условии профессионально созданной БД.

4.2. Принципы сбора и обработки пространственно-распределенных данных

Сбор и обработка пространственно-распределенных данных реализуется по принципам:

1. получения или генерирования собственных данных;
2. использования открытого исходного кода поставщика данных;
3. использования внешнего сервиса выдачи данных;
4. добавления источника данных в виде библиотеки;
5. выполнения внешнего фреймворка (сторонней программы).

Все варианты, кроме первого, предполагают очевидные плюсы и минусы при реализации.

1-ый вариант является эталонным, т.к. все необходимые и важные характеристики и параметры задаются самостоятельно: процесс разработки, развития и поддержки, стандарты и технологии взаимодействия, среда разработки и выполнения, совместимость, язык программирования.

2-й вариант предполагает использование открытого исходного кода фактически близкого к 1 варианту. При необходимости он позволяет изменить код, что является преимуществом. Минусом является заложенная исходная концепция функционирования, которая может коррелировать, но частично или полностью не будет совпадать с характеристиками и параметрами 1 варианта.

3,4,5-ый варианты подразумевают большие риски, заметно упрощая задачу. Использование стороннего сервиса, библиотеки и/или фреймворков позволяет переложить задачи по разработке архитектурно-технологических решений на стороннего разработчика, но при этом возникает максимальная зависимость от стороннего разработчика, что приводит к ограничениям использования, специфическим решениям, рискам неподконтрольности процессов и решений, отказу или приостановке взаимодействия по разным причинам.

4.3. Схемы проектирования топологии поступления данных

Широко распространены, имея очевидные плюсы и минусы, такие схемы, как линейная (полносвязанная), «шина», «кольцо», «звезда» (табл. 2) и комбинации из перечисленных схем. Актуальными являются специальные решения: «MapReduce», «Scatter-Process-Gather» [Богданов, 2004; Pavlo et al., 2009] и другие.

Таблица 2. Основные схемы поступления данных

Схема	Плюсы	Минусы
Звезда	Надёжность	Пространственная ограниченность
Шина	Простота и масштабируемость	Ограниченность пропускной способности
Кольцо	Высокая пропускная способность, масштабируемость	Ненадежность, трудоемкость обслуживания

Специфичность и разнородность входных гелиогеофизических данных, а также необходимость в оперативном использовании набора поступающих данных, при наличии данных с частотой обновления 1 раз в сутки и реже, не позволяет остановиться на одной схеме поступления данных, а требуют разнородных комбинаций в зависимости от задач.

4.4 Язык программирования и другие технологические факторы влияния на архитектуру

Планируемый к использованию при реализации функционала язык программирования является важнейшим фактором, влияющим на архитектуру и технологические решения. Выбранный язык программирования предопределяет полноту и сочетание идей и понятий, которые задают подход и стиль написания программ, т.е.:

- парадигму программирования;
- комплекс правил описания комбинаций символов языка, которые придают правильность структуры фрагменту или программе в целом;
- семантику и конструкцию элементов языка [Гвоздев, 2016];
- систему/системы компоновки типов данных и их производных [Гвоздев, 2016];
- способы внутреннего и взаимного взаимодействия модулей программы, комплексов программ и т.д.

В целях решения задач СОПД сейчас широко распространены следующие языки программирования (с примерами реализации в готовых решениях):

1. C (GRASS, GDAL, PostGIS);
2. C++ (ArcGIS, QuantumGIS, Mapnik);
3. Java (ArcGIS Engine, geoTools, gvSIG);
4. Python (ArcGIS, QuantumGIS, Mapnik);
5. Node (Mapnik).

В настоящее время является редкостью коммуникация комплексов программ и их модулей в одном физическом пространстве, поэтому обязательно обеспечение доступности работы в условиях удаленного доступа к общей памяти с использованием эффективных транспортных технологий (технологий взаимодействия):

1. межпроцессорного взаимодействия именованных каналов;
2. межпроцессорного взаимодействия неименованных каналов;
3. сетевого взаимодействия (API, Winsock, Berkeley Sockets API);
4. представления данных;
5. удаленного вызова процедур – RPC.

Особое место в рамках выбора и использования технологий взаимодействия приобретает программная совместимость и совместимость сред. Категория форматов – метаформаты (XML, XAML, YAML, BSON, JSON, GeoJSON и прочие), в которых определен синтаксический уровень взаимодействия, позволяет избежать ограничений на семантическом уровне, т.к. они предполагают необходимость разработать слой интерпретации данных с использованием стандартных средств кодировки для их внедрения.

Все программные решения и протоколы взаимодействия характеризуются понятием состояния данных, используемых поставщиком и пользователем. Состояние данных обеспечивает

интерпретирование потока сообщений, если без него невозможна корректная обработка потока сообщений.

Основным постулатом при подборе и разработке компонентов ИАС можно считать недопущение явной и неявной программной несовместимости, которая характеризуется аппаратной платформой, операционной системой и программной средой. Важнейшей задачей при построении информационно-аналитических систем является максимальная универсальность трёх вышеперечисленных факторов. Это предопределяет универсальность системы и позволит избежать сложных надстроек, преобразований и абстракций.

5. ВЫВОДЫ

Монолитные системы – готовые решения (такие как продукты ESRI ArcGIS, MapInfo, GRASS GIS, QGIS, gvSIG, ГИС Панорама, TerraLib и прочие), используемые в сочетании с СУБД собственных или внешних разработок (Oracle, PostgreSQL, Microsoft SQL, MySQL, MariaDB и другие), имеют высокий потенциал повторного использования. Их архитектуры, раскрываясь, постепенно меняются т.к. монолитность сильно ограничивает возможности решения каждой конкретной задачи. Наиболее явные шаги в этом направлении делаются в ESRI с продуктом ArcGIS, разработчиками QGIS и отчасти КБ Панорама с их продуктом GIS Toolkit. Открытость, многокомпонентность, которая делегирует пользователям-разработчикам возможности использования качественных программных компонентов и комплексов для более точного и гибкого решения задач.

Важнейшими недостатками применения готовых решений ИАС и ГИС с учётом современных условий являются закрытость исходного кода, локация разработчиков и зависимость от ежегодных обновлений.

Готовые программные решения имеют ограниченный набор схем организации данных, их жизненного цикла, видов, типов. Несмотря на заметную ориентированность в сторону прикладной автоматизации, создание и сохранение функционала межпрограммного взаимодействия, интеграция в другие системы, одновременное использование функций нескольких систем для решения конкретной задачи исключены. Это обусловлено коммерческим характером продуктов. Разработчиками часто исключается возможность использования функций их продуктов в связке со сторонними продуктами.

Решения с открытым исходным кодом отчасти компенсируют минусы монолитных систем, однако весь спектр готовых продуктов решает ограниченные классы задач, в основном после появления решения в коммерчески успешных системах, например, ESRI ArcGIS (попутно собирая все плюсы и минусы архитектуры «Unix-way»).

Сочетания решений с открытым исходным кодом требуют участия специалистов высокой квалификации на всём протяжении работы от формирования требований к ИАС или ГИС до обеспечения бесперебойной работы. Такие системы ориентированы на наиболее подготовленных конечных пользователей.

Несмотря на существование стандартов OGC, интерфейсы готовых программных продуктов различны и несовместимы. Часто аналогичные операции имеют разные названия и расположения в рамках интерфейса, что обусловлено попытками удержать разработчиков и пользователей в своей так называемой экосистеме, несмотря на возникающие сложности использования продукта и проблемы с решением необходимых задач. Даже с используемыми низкоуровневыми языками программирования интерфейсные решения обеспечивают хорошее качество продукта, но при этом имеют заметные ограничения функциональности, низкую степень выразительности и требуют использования таких же специфичных сред выполнения.

Важнейшим отличием решений с открытым исходным кодом от монолитных коммерческих продуктов является отсутствие системного подхода в совершенствовании программного продукта. По сути, его развитие сводится к запоздалому копированию новых решений в коммерческих продуктах. Созданные в разные периоды времени интерфейсы готовых программных продуктов представляют собой, как правило, разные решения, стили, расположения функций. При этом может встречаться дублирование функций, где каждая из них имеет свои ограничения и зависимости.

Перечисленные выше сложности при решении задач увеличивают трудо- и времязатраты, направленные на решение технологических и программных проблем совместимости, часто приводят к невозможности решения поставленных задач или делая принципиально невозможным рациональное решение прикладных задач, например, таких как поэтапное сопровождение пространственных данных в течение жизненного цикла, создание (распределенных, неоднородных, параллельных и др.) СОХА

пространственных данных [Гвоздев, 2016]. Архитектурно-технологические ограничения, заложенные в готовые программные продукты, наиболее ярко проявляются при высокой специфичности предметной области, решаемых задач, необходимого набора программных решений, технологий интеграции и прочего.

При комбинированном подходе сочетаются положительные факторы встроенных частей универсальных продуктов, таких как, например, wget [GNU Wget...], cron [Сороченко и др., 2018], cesiumjs [Стародубцев и др., 2016; Cesium...] с авторскими решениями для СОХА. Такой подход обеспечивает широкий спектр возможностей тонкой настройки под каждого пользователя. При этом заданная структура, имеющая свою физическую особенность предмета исследования, по сути, обеспечивает универсальность применения в тематическом сегменте. Такой подход удовлетворяет требованиям «быстрой разработки».

Разработка модели требований на принципах физичности предмета исследования (перфектный подход) позволяет сформировать методологическую основу разработки ИАС, которая избавляет от «ахиллесовой пяты» разработчиков – неопределенности окончательных требований к программной платформе.

6. ЗАКЛЮЧЕНИЕ

В рамках особенностей работ [Кузьмин и др., 2021a; Кузьмин и др., 2021b; Петрукович и др., 2017; Никифоров и др., 2018; Умрихин и др., 2019], а также проведенного выше анализа методик разработки информационных систем сбора, обработки и хранения гелиогеофизических пространственно-распределенных данных в целом и для ИАС ИКИ ГГФ в частности, сформулированы четыре принципа в иерархическом порядке, которыми целесообразно руководствоваться при их проектировании и разработке:

- приоритетность физически обоснованных решений конкретных тематических задач/задачи с предельно корректной привязкой, максимально описанным предметным областям, согласно перфектному подходу, из которых вытекают алгоритмы, структуры данных, структуры потоков данных и прочее;
- выбор комбинированного подхода при разработке СОХА с добавлением специального при необходимости позволит обеспечить совместимость с существующими шаблонами архитектур ПО, а также с вновь создаваемыми, обеспечивая универсальность, при этом легко подстраиваясь под требования перфектного подхода, обеспечивая его;
- осознанный выбор программного кода, с учётом обеспечения его доминирования при последующем выборе с планируемыми к использованию вычислительными средами, аппаратными и программными платформами;
- обеспечение применения прикладного программирования в целях создания решений для обеспечения, в первую очередь, перфектной логики работы ИАС (перфектного подхода) с безусловными используемыми стандартами системного программного обеспечения (вычислительными средами, аппаратными и программными платформами).

ЛИТЕРАТУРА

- Богданов А.В. Архитектуры и топологии многопроцессорных вычислительных систем // Учеб. пособие для студентов вузов, обучающихся по специальности 351400 "Прикладная математика и информатика" - Москва : Интернет-Ун-т Информ. Технологий, 2004.
- Власов А.И. Системный подход к проектированию при каскадной и итеративной модели жизненного цикла // Труды международного симпозиума "Надежность и качество" Том 1. С. 96-100. 2015.
- Гвоздев О.Г. Исследование принципов построения и разработка архитектуры обобщенной открытой программной платформы для обработки и хранения пространственных данных // Москва: МИИГАиК, 2016. 140 с.
- Железняков В.А. Применение международных стандартов OGC WMS и WFS для формирования, обмена и предоставления пространственных данных // Инженерные изыскания. Вып. 9, 2011. С. 76-78.
- Капралов Е.Г. Кошкарёв А.В., Тикунов В.С. и др. Геоинформатика // Москва: Издательский центр «Академия» - 3-е изд., перер. и доп., 2010. 432 с.

- Кнут Д. Искусство программирования // перев. Тригуб С.Г., Гордиенко Ю.Г., Красиков И.В – Москва: Вильямс Т. 1, 2006.
- Кузьмин А.К. Мерзлый А.М., Моисеев П.П., Никифоров О.В., Петрукович А.А., Щукин Ю.А., Янаков А.Т. Перспективная технология диагностики полярной ионосферы на основе МКА в интересах решения задач связи и радиолокации // Сборник статей I Всероссийской научно-технической конференции : «Состояние и перспективы развития современной науки по направлению «Малые космические аппараты». —Анапа : Федеральное государственное автономное учреждение "Военный инновационный технополис "ЭРА", 2021а. С. 37-47.
- Кузьмин А.К. Мерзлый А.М., Никифоров О.В. и др. Основы перспективной методики комплексных исследований влияния авроральных характеристик полярной ионосферы на условия распространения трансполярных сигналов. Обзор зарубежных экспериментов и результаты некоторых моделей // Гелиогеофизические исследования, вып. 32, 2021b. С. 3-60.
- Никифоров О.В. Мерзлый А.М., Петрукович А.А., Янаков А.Т., Моисеев И.А., Филатов Г.С., Уваров И.А. Использование информационно-аналитической системы ИКИ РАН «Гелиогеофизика» (ИАС ИКИ ГГФ) в экспериментальных работах // Девятнадцатая ежегодная конференция "Физика плазмы в солнечной системе" ред. Струминский А.Б. Нечаева А.Б. Москва, 2024. 303 с.
- Никифоров О.В. ОТЧЕТ О НИР "Большие данные в космических исследованиях: астрофизика, солнечная система, геосфера" по теме: КОСМОС-Д 0024-2021-0011 (промежуточный, этап 2021/1) // Институт космических исследований Российской академии наук, 2021.
- Никифоров О.В. Петрукович А.А., Уваров И.А. Информационная система "Аврора-Арктика". Задачи и перспективы развития // Сборник тезисов докладов шестнадцатой Всероссийской открытой конференции "Современные проблемы дистанционного зондирования Земли из космоса". Москва: Институт космических исследований Российской академии наук, 2018. 477 с.
- Паутов А.В. MySQL: руководство профессионала // Wikireading.ru 2023 [Электронный ресурс]. URL: <https://it.wikireading.ru/31847> (дата обращения 03.03.2023).
- Первухин Д.В. Исаев Е.А., Рытиков Г.О. и др. Сравнительный анализ теоретических моделей каскадных, итеративных и гибридных подходов к управлению жизненным циклом ИТ-проекта // Бизнес-информатика. Вып. 14, 2020. С. 32-40.
- Петрукович А.А. Никифоров О.В., Уваров И.А., Подладчикова Т.В. Информационная система мониторинга и прогноза аврорального овала // Сборник тезисов докладов пятнадцатой Всероссийской открытой конференции "Современные проблемы дистанционного зондирования Земли из космоса", Москва: Институт космических исследований Российской академии наук, 2017. 429 с.
- Рихтер Дж. Windows для профессионалов: создание эффективных Win32-приложений с учетом специфики 64-разрядной версии Windows // СПб: Питер. Издательско-торговый дом «Русская редакция», 2001. 752 с.
- Сороченко Р.А. Крылов Е.В., Аникин В.К. Использование программы sgon для выполнения сложных и трудоемких задач веб-приложений // Актуальные научные исследования в современном мире вып. 43, 2018. С. 66-70.
- Стародубцев А.В. Слепенков В.О. Фреймворк CesiumJS // Молодежный научно-технический вестник, вып. 12, 2016. С. 58-71.
- Умрихин А. Д. Никифоров О. В., Петрукович А. А., Уваров И. А. Интеграция моделей прогноза границ аврорального овала в информационную систему мониторинга ионосферы "Аврора-Арктика" // Материалы 17-й Всероссийской открытой конференции "Современные проблемы дистанционного зондирования Земли из космоса", Москва: Институт космических исследований Российской академии наук, 2019. 499 с.
- Флейс М.Э. Математическая основа геоинформационных систем [специальность 25.00.33 "Картография" : диссертация на соискание ученой степени кандидата технических наук] // Москва, 2002. 122 с.
- Шикин Е. В. Боресков А. В. Компьютерная графика: Учебник и практикум // Москва: Юрайт, 2020.
- Эккель Б. Философия Java // СПб: Питер, 2013.
- Cesium. The Platform for 3D Geospatial [Электронный ресурс]. URL: <https://cesium.com/platform/cesiumjs/> (дата обращения 03.03.2023).
- Cohn M. Succeeding with Agile: Software Development Using Scrum // Pearson, 2011. 567 p.

- ESRI Shapefile Technical Description // Environmental Systems Research Institute, Inc., 1998. 30 p.
- FlatBuffers. Cross platform serialization library architected for maximum memory efficiency [Электронный ресурс]. URL: <https://google.github.io/flatbuffers/> (дата обращения 03.03.2023).
- GISLAB. Введение в геоинформационные системы. Системы координат и проекции [Электронный ресурс]. URL: <https://gis-lab.info/docs/giscourse/08-coords.html> (дата обращения 03.03.2023).
- GNU Wget. free software package for retrieving files using HTTP, HTTPS, FTP and FTPS, the most widely used Internet protocols [Электронный ресурс]. URL: <https://www.gnu.org/software/wget/> (дата обращения 03.03.2023).
- GRASS. Powerful computational engine for raster, vector, and geospatial processing [Электронный ресурс]. URL: <https://grass.osgeo.org> (дата обращения 03.03.2023).
- gvSIG. The largest professionals network of open source geomatics [Электронный ресурс]. URL: <http://www.gvsig.org/> (дата обращения 03.03.2023).
- *Larman C.* Applying UML and Patterns : An Introduction to Object-Oriented Analysis and Design and Iterative Development // Pearson, 2006. - Т. 3-е издание. 736 с.
- OGC. Simple Feature Access. Common Architecture [Электронный ресурс]. URL: <https://www.ogc.org/standard/sfa/> (дата обращения 03.03.2023).
- ORACLE. MySQL 8.0 Reference Manual [Электронный ресурс]. URL: <https://dev.mysql.com/doc/refman/8.0/en/spatial-type-overview.html> (дата обращения 03.03.2023).
- *Pavlo A., Paulson E., Rasin A., Abadi D.J., DeWitt D.J., Madden S., and Stonebraker M.* Proceedings of the 2009 ACM SIGMOD International Conference on Management of data (SIGMOD '09) // A comparison of approaches to large-scale data analysis, New York, NY, USA, 2009. pp. 165–178.
- PostGIS. Spatial and Geographic Objects for PostgreSQL [Электронный ресурс]. URL: <http://postgis.net/> (дата обращения 03.03.2023).
- QGIS Project, Geographic Information System. [Электронный ресурс]. URL: <https://www.qgis.org/> (дата обращения 03.03.2023).
- *Rigaux P., Scholl M., Voisard A.* Spatial Databases - With Application to GIS // Morgan-Kaufman Publishers, 2002. 410 p.
- *Shannon C.E.* Claude Elwood Shannon: Collected Papers // Wiley-IEEE Press, 1993. 968 p.
- *Warmerdam F., Rouault E. and others.* GDAL — Geospatial Data Abstraction Library [Электронный ресурс]. URL: <http://www.gdal.org/> (дата обращения 03.03.2023).

PRINCIPLES FOR THE DEVELOPMENT OF INFORMATION SYSTEMS FOR THE COLLECTION, PROCESSING AND STORAGE OF HELIOGEOPHYSICAL GEOSPATIAL DATA

Nikiforov O. V., Kuzmin A. K., Merzliy A. M., Popov V. Yu., Shagurin I. A., Yanakov A. T.

As part of a series of works on the creation of an IAS (information and analytical system) in the interests of creating a system for monitoring the conditions of propagation of short-wave radio signals and systematic monitoring, evaluation and forecasting of the operating conditions of radio equipment in the polar ionosphere, the IAS "Heliogeophysics" and the geographic information system (GIS) "Aurora-Arctic" are being created. The article presents approaches to the development of a system in the field of collection, processing, storage and analysis of spatially distributed data, presents existing solutions for these tasks and describes a number of important factors affecting the software compatibility of such systems. The technological features of spatial data processing tasks, as well as approaches to the design of the IAS structure, are considered. The paper considers software and architectural directions in the creation of IAS, including the analysis of types and schemes for designing system architectures, principles of collecting and processing spatially distributed data, schemes for designing the topology of data receipt, programming languages and other technological factors affecting the architecture of IAS. At the end of the work, conclusions are drawn, and the principles of designing and developing IAS using heliogeophysical spatially distributed data are formulated, i.e. information, indicators, facts obtained from scientific and other sensors/devices, expressed both numerically and in any other form.

KEYWORDS: INFORMATION AND ANALYTICAL SYSTEMS, GEO-INFORMATION SYSTEMS, COLLECTION, PROCESSING, STORAGE, ANALYSIS, GEOSPATIAL DATA