

Подходы по визуализации данных ЕСИМО

Е.Д. Вязилов, Н.Н. Михайлов, Д.А. Мельников, Н.А. Чуняев

Аннотация

Представлено описание подходов по реализации и функций ЕСИМО по визуализации данных. Дано краткое описание таких функциональных возможностей как поиск и визуализация метаданных, данных, выявление индикаторной информации для поддержки решений.

Ключевые слова: функции ЕСИМО, поиск, визуализация метаданных, данных, выявление индикаторной информации.

1. Введение

В 2013 г. полнофункциональная версия Единой государственной системы информации об обстановке в Мировом океане (ЕСИМО) вводится в постоянную эксплуатацию. Первая очередь системы функционирует с 2008 г. Главной целью первой очереди системы была интеграция данных. За эти годы было интегрировано более 3500 информационных ресурсов, представляющий собой распределенные и разнородные данные, представленные в виде баз данных (БД), структурированных файлов, объектных файлов с изображениями, документами, ссылками на программные приложения. В настоящее время функционирует 34 узла системы, имеется три региональных узла, обеспечивающих поддержку региональной подсети узлов и информационное обеспечение пользователей.

При разработке ЕСИМО использованы следующие принципы. Каждый локальный узел системы выступает в качестве независимого, автономного узла системы. Каждый узел несет ответственность за безопасность, резервное копирование и восстановление данных. Все узлы потенциально имеют одни те же возможности. Работа системы не зависит от сбоев конкретных узлов. Система находится в непрерывной работе в случае отказа узла или расширения сети. Пользователю не нужно знать расположение данных, чтобы получить необходимые данные. Каждый пользователь видит только одну логическую БД. Пользователю не нужно знать имя БД и другие свойства объектов БД. Пользователь может выбирать любой информационный ресурс системы для доступа. Распределенный запрос может быть выполнен из нескольких узлов системы одновременно. Распределенная обработка данных выполняется на нескольких узлах системы. Система работает на любой аппаратной платформе, операционной и сетевой платформах.

Ресурсы ЕСИМО отличаются большим объемом данных (производится ежесуточная переработка до 10 Гбайт). Требуется высокая скорость обновления данных (до 30% ресурсов имеют частоту обновления от трех часов до недели). Наблюдается вариативность данных (структуры данных в информационных ресурсах изменяется - появляются новые параметры данных, добавляются новые показатели в существующие ресурсы, создаются аналитические представления из нескольких ресурсов или на основе агрегированных данных). Данные имеют высокую ценность для пользователей (информационные ресурсы используются для оповещения руководителей и населения об обстановке в Мировом океане, ресурсы являются уникальными, на их основе можно рассчитывать климатические характеристики и давать оценку риска опасных природных явлений).

Чтобы принимаемые решения на основе информации о состоянии морской среды были более эффективными важно наличие возможности обеспечить пользователей информацией в любое время, в нужном формате, минимально необходимом объеме информации и правильно представленном виде (на карте, графике, диаграмме и текстом). При этом пользователь работает в едином интерфейсе через портал системы (<http://portal.esimo.ru>, раздел «Данные»).

Пользователями ЕСИМО являются федеральные органы исполнительной власти, администрации субъектов РФ, руководители предприятий. Логика развития систем подобного класса требует максимальной персонализации. Поэтому кроме портала должны быть созданы подпорталы и автоматизированные рабочие места (АРМы). Подпорталы и АРМы ЕСИМО строятся по одному принципу, подпортал имеет свой URL адрес, а АРМ располагается на порталах или подпорталах. Назначением ЕСИМО является интеграция данных из распределенных, неоднородных, ведомственных информационных систем; обеспечение контроля обновления и комплексного анализа с помощью единого инструмента; сокращение трудозатрат на обработку и создание аналитических представлений; поддержка принятия решений; контроль отклонений показателей морской среды; комплексное информационное обеспечение пользователей; визуализация и привлечение внимания к проблемным ситуациям, зарегистрированным в морской среде. Оператором ЕСИМО является ФГБУ «ВНИГМИ-МЦД».

В традиционном подходе к получению аналитических материалов источники данных должны быть заранее определены IT-специалистами. Для поддержки решений данные загружаются в хранилище в виде куба, а потом - в аналитическую систему [4-6]. При любом изменении данных необходимо пройти весь процесс заново. Главными недостатками существующих систем являются их сложность применения для большинства пользователей - чужеродность и непривычность инструментария мешают сотрудникам взять его на вооружение; функциональность размазана по

нескольким приложениям [1-3].

К недостаткам первой очереди ЕСИМО можно отнести следующее. Есть задачи, для которых требуется получать информацию в режиме реального времени. По некоторым ресурсам представлены только агрегированные данные, без возможности их рассмотрения на более детальном уровне или информация ограничена определенным временным периодом (например, последние пять лет) без возможности увидеть всю историю конкретного показателя. Для большинства пользователей особенно важны средства визуализации, автоматические рассылки обновлений информационных ресурсов. Пользователи часто не имеют полного доступа к данным, на которые они опираются при подготовке решения. Часть данных доступна только с помощью отдельных запросов, не связанных между собой.

2. Основные подходы по разработке

Требованиями к обработке и визуализации являются наследование метаданных при создании агрегированных таблиц, использование метаданных при отображении информации, быстрое построение представлений на больших выборках данных, on-line анализ, наличие средств публикации карт показателей, сохранение результатов расчетов для дальнейшего использования, интеграция с геоинформационными сервисами. При разработке полнофункциональной ЕСИМО используются следующие подходы.

Подключение новых источников – любой участник ЕСИМО может подключить новый информационный ресурс и без доработки программного обеспечения визуализировать или доставить его по новому адресу (электронная почта или ftp – сервер).

Высокая скорость обработки

– информационные ресурсы в объеме более одного гигабайта обрабатываются в течение одного часа после его обновления на узле поставщика данных. Данные обновляются в информационных ресурсах ЕСИМО в соответствии с регламентом, представленном в описании ресурса.

Интерактивность

– пользователь задает множество запросов за один сеанс работы с данными, система отвечает и представляет результаты обработки данных.

Богатая визуализация

– данные представляются в виде таблиц, графиков, карт, текста, «приборов», показывающих состояние показателя в виде «светофора» (зеленый, желтый, красный). При этом применяется подход по предварительной фильтрации данных из ресурсов за счет синхронизации значений поисковых атрибутов, получаемых из базы интегрированных данных в процессе поиска данных.

Компонентный подход

– разделение программного обеспечения на независимые, программные компоненты, выполняющие определенные функции. В настоящий момент выделено 13 компонент.

1) Программное обеспечение поставщик данных – предназначено для организации описания интегрируемых данных, маппинга имен атрибутов данных и кодовых таблиц классификаторов, подготовки данных к передаче на сервер интеграции.

2) Сервер интеграции – в режиме реального времени по событию (изменению или замещению данных) реплицирует данные от поставщиков данных в базу интегрированных данных (БИД). При этом возможна фильтрация и преобразование структур данных.

3) Сервисная шина – собирает логи от программных компонент и приложений.

4) БИД под управлением системы управления БД PostgreSQL – представляет собой хранилище данных, содержит данные от всех поставщиков данных, построенные на единой модели данных, где для каждого информационного ресурса выделено свое семейство плоских таблиц.

5) ГИС сервер – предоставляет возможности создания картографических слоев и WMS-сервисов.

6) Картографический визуализатор – дает возможность искать картографические слои, визуализировать их, увеличивать или уменьшать масштаб карт, наносить дополнительные слои, проводить расчеты расстояний, площадей, др.

7) Аналитический комплекс – позволяет создавать прикладные задачи с несколькими аналитическими представлениями,

состоящими из таблиц, карт, графиков, диаграмм, приборов.

8) Портал – интегрирует картографические и аналитические представления, а также отдельные приложения (портлеты) для комплексного информационного обеспечения пользователей.

9) Централизованная база метаданных – включает сведения обо всех технологиях, используемых и разработанных в ЕСИМО с включением информации о таких объектах метаданных, как описание массивов и БД, форматов, организаций, сетей наблюдений, судов, нормативно-методических документов и др.

10) Общие коды и классификаторы – предназначены для систематизации, стандартизации используемых классификаторов и их поиска.

11) Мониторинг ресурсов и сервисов – позволяет обеспечить контроль работоспособности аппаратно-программных комплексов во всех узлах системы.

12) Информационная безопасность – предназначена для обеспечения защиты данных и сервисов информационно-технологических узлов от несанкционированного доступа и реализации единой политики доступа к данным и сервисам ЕСИМО.

13) Электронная карта-основа – обеспечивает ведение и распространение слоев и геосервисов единой карты-основы.

Операторы ЕСИМО производят:

- описание информационных ресурсов, маппинг имен атрибутов и кодовых таблиц классификаторов, подготовку данных к передаче на сервер интеграции;
- доставку данных в БИД и другие компоненты;
- интеграцию данных в БИД и ее использование различными приложениями;
- встраивание полученных представлений в порталы (АРМы);
- использование данных в других информационных системах;
- сохранение результатов работы в виде совокупности ресурсов, слоев, представлений.

Основными функциями пользователей ЕСИМО являются:

- поиск метаданных - просмотр списка объектов метаданных и визуализация отдельных экземпляров;
- поиск данных - просмотр списка доступных ресурсов СРБД с возможностью визуализации данных;
- агрегация данных – получение обобщенных свойств объектов для различных пространственно-временных масштабов;
- визуализация данных и формирование представлений средствами аналитического комплекса, ГИС-вьюера или портала;
- просмотр отдельных ресурсов в виде таблицы, графика, карты в формате GeoJSON для плоской карты, галереи изображений, формате RSS для последних поступлений данных, бегущей строки.

Общая схема представлений с помощью портлета визуализации дана на рис.1.

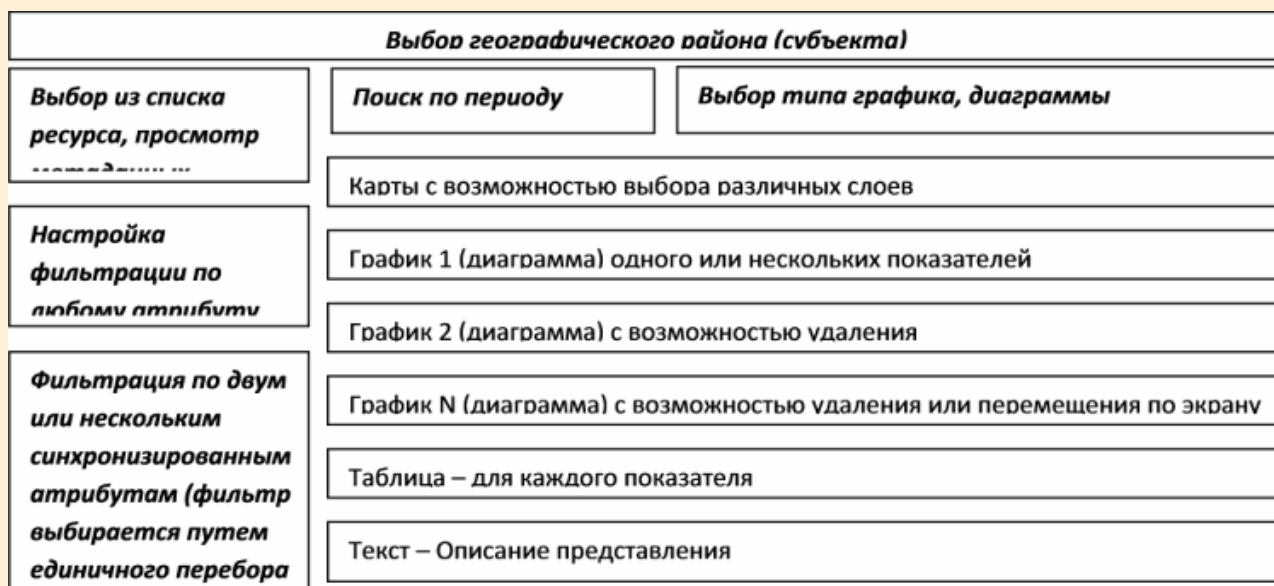


Рисунок 1 – Обобщенный шаблон по визуализации данных

Администратор портала выполняет следующие функции:

- получение статистики доступа к данным, слоям и аналитическим представлениям;
- создание и просмотр профиля ресурсов, слоев, представлений пользователей и удаление их;
- просмотр и удаление подписки пользователей;
- создание АРМов пользователей (именование, назначение, описание, выбор стиля) и их каталога (просмотр списка АРМов, редактирование и удаление из каталога);
- создание, настройка, просмотр, редактирование, удаление экземпляров портлетов;
- получение сведений о системных элементах;
- фильтрация экземпляров данных внутри ресурса по выбранным критериям;
- получение доступа к архиву новостей, штормовых оповещений;
- поиск прикладных задач, аналитических представлений, информеров по каталогу и их визуализация;
- получение данных в виде карты, графика (диаграммы);
- отображение различных типов хранимых данных (БД, структурированных файлов, каталогов объектных файлов) в виде таблиц;
- экспорт данных в форматы CSV, EXCEL, XML;
- доставка информационной продукции на регулярной основе.

3. Описание некоторых функций визуализации данных

Для выполнения вышеперечисленных функций используются портлеты «Метаданные», «Данные», «Бегущая строка», «Geo-RSS», «Информеры», «Подписка», «Обратная связь с пользователями», «Прикладные задачи».

Портлет «**Бегущая строка**» визуализирует новостную информацию и штормовые сообщения. У портлета есть возможность ввода новых сообщений. Для работы со штормовыми сообщениями добавлена возможность указать атрибут «Важность сообщения», по которому можно сортировать их. Оператор имеет возможность ввести с экрана новое событие или штормовое сообщение. В портлете есть возможность получения доступа к архиву новостей, штормовых оповещений.

Портлет «**Geo-RSS**» передает пространственные данные ресурсов БИД для отображения в виде тематического слоя карты. Портлет настраивается на географический район интересов, параметр (ы), состав атрибутивной информации, визуализируемой на карте. При выводе карты появляются минимальные возможности работы с ней (увеличение, уменьшение масштабов, перетаскивание карты).

Портлет «**Информеры**» выполняет функции поиска информеров по каталогу на основе таких поисковых атрибутов, как прикладная задача, показатель, форма представления, географический район, и визуализировать их.

Портлет «**Прикладные задачи**» усваивает ресурсы с прикладными задачами, настраивается на состав аналитических представлений, показателей, географический район интересов. Результатом работы портлета является автоматически генерируемое меню прикладных задач для региональных и ведомственных подпорталов. Портлет имеет каталог прикладных задач с поиском, аналогичным по составу полей поиску ресурсов. При этом производится просмотр отдельной прикладной задачи с навигацией между аналитическими представлениями посредством вкладок и кнопок перехода к предыдущему / следующему представлению; поиск представлений внутри прикладной задачи по заданной совокупности ресурсов и параметров; перетаскивание элементов представления внутри рабочей области; растягивание элементов представления внутри рабочей области; открытие элементов представления в отдельном окне. Кроме того можно произвести сортировку, фильтрацию, изменение состава колонок; включение, выключение отдельных графиков, масштабирование, получение информации о данных в точке; осуществить навигацию, масштабирование, получение данных в точке для отдельного слоя; визуализировать данные в виде «прибора» в соответствии с условиями, заданными аналитиком; экспорт прикладной задачи и/или ее представления и/или его элементов в формат doc и pdf, экспорт карт, графиков и приборов в форматы jpeg, tiff; нанесение элементов «текст» или «рисунок» на представление перед экспортом; получение метаданных по используемым в представлении ресурсам.

Поиск и визуализация метаданных. Список объектов включает нормативно-правовые документы; концепции, доктрины, кодексы; объекты морской деятельности (мореведческие организации; порты мира; суда; научные исследования; проекты и программы; системы наблюдений (прибрежные станции, пункты мониторинга загрязнения морей, научно-исследовательские суда, буи, спутники, приборы, методы и модели расчетов); сведения об экспедиционных исследованиях (заявки, планы, выполненные экспедиции); информационные технологии (сведения о массивах и БД, программных средствах, методах и алгоритмах обработки, форматах данных, международных и российских стандартах); словари терминов и определений (океанографические термины, информационные термины, список сокращений, единый словарь параметров); технические спецификации и технологии ЕСИМО. Поиск метаданных осуществляется для каждого объекта отдельно. Пользователь сначала выбирает объект метаданных, а затем заказывает значения поисковых атрибутов. Результаты работы портлета представляются в виде таблицы с основными полями описания объекта, по идентификатору экземпляра объекта метаданных пользователь получает полное описание объекта.

Портлет «**Визуализация данных**» имеет гибкую настройку на состав требуемых информационных ресурсов, параметры, географический район. Для типизации представлений используется универсальный шаблон для визуализации данных, хранимых в виде временных рядов, сеток, профилей, объектных файлов, ресурсов, не привязанных ко времени, а также результатов объединения текстовых и структурированных данных. Ресурсы выделяются по таким атрибутам метаданных как пространственный масштаб (фиксированная точка, траектория, сетка); временной масштаб (регулярные - срок, сутки, декада, месяц, квартал, год и нерегулярные измерения); платформа (прибрежная станция, научно-исследовательское судно, буй). Возможность построения карты определяется по наличию в данных атрибутов широта и долгота. Если широты и долготы нет, но имеются идентификаторы геообъектов (идентификатор станции - координатный или синоптический номер, идентификатор порта, предприятия, др.), то автоматически добавляется широта, долгота из базового ресурса. Если для визуализированного ресурса имеется заранее построенный картографический сервис, то его идентификатор заранее настраивается в матрице соответствия.

Возможные варианты поиска данных в ресурсах включают поиск по слову или фразе, на основе рубрикатора, по дате, географии, поставщикам ресурсов, уровню доступа, уровню обработки, типу платформы, параметрам. Для доступа к данным пользователь запрашивает разрешение на их использование (если ресурс закрыт для пользователя), используя механизм заявок на ресурсы. Можно визуализировать исходные данные (если ресурс разрешен и данные есть в БИД узла) в отдельном окне; скачать данные; подписаться на данные (для зарегистрированных пользователей); просматривать его на карте; использовать в аналитическом представлении.

Есть возможность осуществить экспорт-импорт в различные форматы хранения данных. В результате представления превращаются в интерактивные документы с графиками, картами, оповещениями, гиперссылками, т.е. выдаются с насыщенной информацией и презентационным качеством. Данные могут загружаться на ftp-сервер или передаваться по электронной почте.

Содержимое представления дается с минимумом информации (объем выдаваемой информации уменьшается до одной страницы) и при этом есть удобная навигация и подача информации. Если в традиционных приложениях таблицы, графики, карты выводятся отдельно и их можно только смотреть как фотографии, то с помощью портлета все визуальные компоненты объединяются в одном представлении и становятся интерактивными. При этом пользователь может сам фильтровать данные, сворачивать представление и углубляться в него, менять масштаб карты и фильтровать через нее данные, включать и выключать слои.

Поддержка решений. Информация, с помощью которой можно прогнозировать развитие событий, называется индикаторной. Эта информация постоянно помещается в БД и непрерывно анализируется. Этот компонент отслеживает изменения в данных в фоновом режиме и отправляет сообщение пользователю. Создание сообщений активируется состоянием данных, заданным пользователем (например, превышение критических значений). Функция предварительного оповещения конечных пользователей повышает производительность их труда.

Портлет «**Подписка**» позволяет организовать на регулярной основе доставку информационной продукции (ресурсов, слоев, информеров, аналитических представлений). Портлет настраивается на географический район интересов, состав параметров, форму представления, др.

Портлет «**Обратная связь с пользователями**» дает возможность пользователям (администраторам, операторам или конечным пользователям) послать информацию об ошибке, просьбу о помощи и т.п. Информация об ошибке включает название компонента (приложения), где произошла ошибка, тип ошибки, скриншот экрана с ошибкой, описание предусловий, которые привели к ошибке. При ответе пользователю предлагается заполнение опросника по удовлетворенности ресурсами и сервисами.

4. Заключение

Технологическими преимуществами решений ЕСИМО являются гибкое построение представлений и диаграмм «на лету», исследование данных с разной степенью детализации, гибкость настройки, адаптации к новым ресурсам, средствам доставки информации, наличие единого источника интегрированной информации, стандартизованные методы обработки, отказоустойчивость и безопасность работы компонент системы.

ЕСИМО обеспечивает новые возможности анализа, позволяя пользователям получать доступ и интегрировать данные практически из любого источника, создавать эффективные представления и аналитические приложения, а также обмениваться данными анализа и работать над ними с помощью привычных инструментов.

За счет доступности больших объемов данных экономится время на поиск и обработку информации, появляется возможность в режиме on-line видеть результаты обновления данных. К интегрированным данным можно применить различные методы обработки (выявление критических значений, получения климатических значений). Любой пользователь может создать новое представление, преобразовать форму представления данных одним щелчком мыши, добавить анимацию, позволяющую быстро выявить тенденции или аномалии.

Список литературы

[1] Бычков В. ВІ с самообслуживанием: обзор платформ // Ж-л «Computerra». 05.03.2013. [Электронный ресурс]. – Режим доступа: <http://www.computerra.ru/cio/2601>, свободный. – Загл. с экрана.

[2] Вейн Раш. ВІ в режиме самообслуживания: всё ближе к цели // PC Week/RE. 31 августа — 6 сентября 2010. - №31-32 (733-734).

[3] Вязилов Е.Д, Мельников Д.А., Чуняев Н.В. Аналитический комплекс ЕСИМО // Интернет и современное общество. Материалы пятнадцатой всероссийской объединенной конференции «Информационные системы для научных исследований». Сб. научных статей. С-Петербург. 10-12.10.2012. - С.33-37.

[4] Коптелов А. Управление бизнес-процессами: разбираемся с программным обеспечением // С-News. [Электронный ресурс]. – Режим доступа: http://www.cnews.ru/reviews/index.shtml?2012/10/01/505034_2, свободный. – Загл. с экрана.

[5] Обзор архитектуры QlikView Информационные материалы по технологиям QlikView. – QlickTech. - Октябрь, 2010. – 12 с. <http://www.qliksolutions.ru/files/qliksolutions/documents/pdf/QlikView%20Architectural%20Overview.pdf>

[6] Солдатов О. Аналитика: путь к интеллектуальному бизнесу // С-News. 2013. №3. [Электронный ресурс]. – Режим доступа: <http://www.cnews.ru/reviews/?2013/03/28/523883>, свободный. – Загл. с экрана.

Труды XVI Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2013), Санкт-Петербург, Россия, 2013.

Об авторах

Вязилов Евгений Дмитриевич – заведующий лабораторией, доктор технических наук, ФГБУ «Всероссийский научно-исследовательский институт гидрометеорологической информации – Мировой центр данных». e-mail: vjaz@meteo.ru

Чуняев Никита Владимирович – аспирант, инженер-программист, ФГБУ «Всероссийский научно-исследовательский институт гидрометеорологической информации – Мировой центр данных». e-mail: chunyaev@meteo.ru

Михайлов Николай Николаевич – зав. отделом, ФГБУ «Всероссийский научно-исследовательский институт гидрометеорологической информации – Мировой центр данных». e-mail: nodc@meteo.ru

Мельников Денис Андреевич – инженер-программист, ФГБУ «Всероссийский научно-исследовательский институт гидрометеорологической информации – Мировой центр данных». e-mail: melnikov@meteo.ru
