

ИНТЕГРАЦИЯ ГЕТЕРОГЕННЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ В ОБЛАСТИ МОРСКОЙ ДЕЯТЕЛЬНОСТИ*

Е. Д. Вязилов, Н. Н. Михайлов

Всероссийский НИИ гидрометеорологической информации —

Мировой центр данных, Обнинск, Россия

e-mail: vjaz@meteo.ru, nodc@meteo.ru

The ESIMO portal is developed, which includes current, prognostic and climatic information, allowing access to the data at any moment, from any point, on any area, in the digital, graphic and textual forms. Approaches dealing with integration of information resources in the field of marine activity are presented.

Введение

Единая система информации об обстановке в Мировом океане (ЕСИМО) представляет собой сеть информационных систем различных организаций, министерств и ведомств России, которые связаны с исследованиями морской среды или организацией морской деятельности. Ведомственные системы, как правило, реализованы на разных принципах, вычислительных платформах и программных инструментах. Здесь главная информационная проблема — это гетерогенность (разнородность) созданных информационных ресурсов (ИР). Огромные объемы информации хранятся в различных системах управления базами данных (СУБД), имеют разнообразные схемы баз данных (БД), что не всегда позволяет создать из них однородную информацию. К тому же эти БД созданы фрагментарно, т. е. не имеют 100 % полноты. Повышение оперативности доступа к такой информации стоит очень дорого. В то же время имеется насущная необходимость создания приложений, которые будут использовать данные, хранящиеся в разных базах данных. Разрастание несогласованных структур данных влечет за собой много организационных и технических проблем. Выходом из этого является интеграция гетерогенных информационных ресурсов.

Возможности интеграции данных рассматривались уже на самых первых этапах создания массивов и БД. Например, в области океанографии еще в конце шестидесятых годов предлагалась единая система занесения на перфокарты данных всех наблюдений, проводимых в научно-исследовательском рейсе. К сожалению, из-за сложности задачи в то время она не была реализована. В последующем похожую интеграцию данных произвели в системе сбора данных на магнитных лентах. Здесь в одном файле хранились океа-

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант № 04-01-00678-а).

© Институт вычислительных технологий Сибирского отделения Российской академии наук, 2005.

нографические и метеорологические данные, данные наблюдений за загрязнением воды, результаты измерений различными приборами (всего восемь типов записей).

При создании БД также производится объединение данных, но уже требуется интеграция данных, собранных с различных наблюдательных платформ (гидрометеорологических станций, научно-исследовательских судов, буев, др.). Эта задача достаточно хорошо решается за счет использования современных СУБД при правильной организации структуры БД (типизации схем БД, путем применения многомерного подхода и других методов).

В последние годы в связи с развитием Интернет значительно увеличилось разнообразие используемых форм представления данных. Кроме БД и файловых систем появились страницы Web-сайтов с текстовой, графической и мультимедийной информацией, файлы на ftp-серверах. Чтобы найти такие данные во Всемирной паутине и осуществить к ним доступ, требуются определенные усилия, необходима специальная система поиска. Обычные каталоги ссылок на Web-сайты не дают хороших результатов, так как они либо быстро разрастаются, либо не дают точного адреса необходимого ресурса. Данные, хранящиеся в БД, индексируются не всеми поисковыми машинами. Наиболее приемлемым методом интеграции здесь служит порталная технология [1].

В настоящее время разработано достаточно много инструментов для создания Web-порталов. Практически все известные компьютерные фирмы предложили свои инструменты для создания порталов (.Net — Microsoft, WebSphere — IBM, Oracle-portal, Sun ONE и др.) [2]. К сожалению, эти средства очень дорогие, а разрабатываемые приложения в рамках предлагаемых этими фирмами технологий жестко привязаны к этим инструментам. В 1999 году, когда начиналось создание ЕСИМО, порталные средства были в разработке, поэтому выбран вариант собственного инструмента для создания портала ЕСИМО.

Для описания ИР уже разработано несколько моделей, например, в области образовательных ресурсов (Learning Object Metadata Core), научно-технической информации (Dublin Core). Но эти модели не учитывают специфику данных о состоянии природной среды: наличие прогностической и климатической информации; необходимость регулярного пополнения ИР; географическую привязку данных; большое разнообразие параметров наблюдений и их пространственно-временные масштабы обобщения. Поэтому была разработана оригинальная модель описания ИР в области исследований природной среды и морской деятельности. Последняя версия этой модели дана на сайте <http://data.meteo.ru/e2edm/index.php?section=1> для представления в качестве единой модели в рамках Межправительственной океанографической комиссии ЮНЕСКО и Всемирной метеорологической организации.

1. Подходы к интеграции

Единая система информации об обстановке в Мировом океане включает информационные ресурсы, размещенные в организациях-участницах ЕСИМО и других родственных организациях, в том числе зарубежных, в виде баз данных и массивов текущих, оперативных, исторических данных и климатической (обобщенной), диагностической и прогностической информации, а также функциональных приложений, реализуемых в многоаспектной (языковой, программной, лингвистической и др.) среде разработки. Компоненты ЕСИМО представлены на рис. 1. Основные подходы к построению и реализации портала ЕСИМО опубликованы в [3–6].

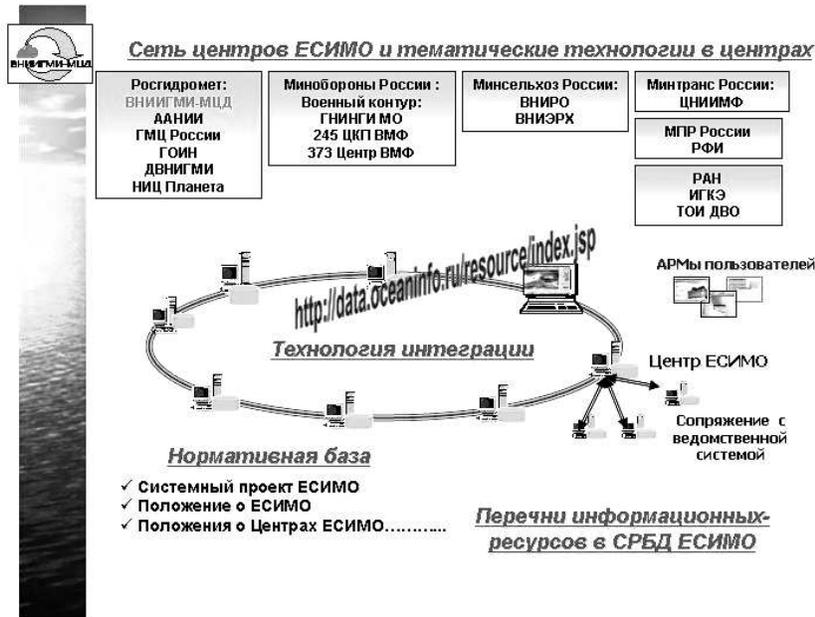


Рис. 1. Компоненты первой очереди ЕСИМО.

Интеграция гетерогенных информационных ресурсов производится:

- применением единой модели описания ИР;
- путем создания и ведения широкого комплекса объектов метаданных (сведения об ИР, организациях, экспертах, форматах, программных средствах и др.), единого словаря параметров;
- использованием единого словаря параметров (<http://data.oceaninfo.ru/udopweb/index.jsp>);
- использованием многомерного подхода к созданию БД (рис. 2) (выделено пять под-схем данных: исходные данные, сетки, матрицы, геоданные, текст, графика);



Рис. 2. Модель базы данных.

— созданием и хранением электронных копий нормативно-правовых и научно-технических документов и изображений в виде файловых систем.

Эти подходы обеспечивают решение задач управления разнородными и слабоструктурированными информационными ресурсами посредством:

— уникальной идентификации источников данных;

— единообразного доступа ко всем ИР и использования специфических поисковых атрибутов для разных форм представления информации (текстовой, фактографической, графической, пространственной), типов данных (наблюдений, анализов, прогнозов, климатических обобщений);

— применения нескольких методов доступа в зависимости от специфики источника данных (HTML-страницы, ftp-файлы, БД, приложения).

Примеры ИР представлены на рис. 3. В системе обеспечивается удаленный ввод описаний ИР и их экземпляров в БД Web-портала ЕСИМО. На основе описаний ИР осуществляются автоматизированный поиск, обращение к информационному ресурсу и его экземплярам (временной ряд, поле, станция и др.). Поиск возможен по классификации (сфере, процессу, параметрам), географическим районам, организации — автору ИР. Мониторинг посещений ИР, а также управление регистрацией и доступом к ИР осуществляются административным сервисом Web-портала ЕСИМО.

Для ведения и использования портала разработан многокомпонентный, кросс-платформенный информационно-программный комплекс. При этом используются Web-сервер Apache, сервер приложений JBOSS, сервер БД Oracle 9i, геоинформационная система Mapserver для визуализации картографической информации в режиме on-line, геоинформационная система ArcInfo для подготовки тематических карт, электронных справочных пособий в режиме off-line. Программное ядро комплекса обеспечивает управление и хранение описаний ИР, метаданных, исходных и аналитических данных в СУБД Oracle 9i. Пользовательский интерфейс реализован на базе технологии J2EE. При разработке приложений широко применяется метод создания java-классов.

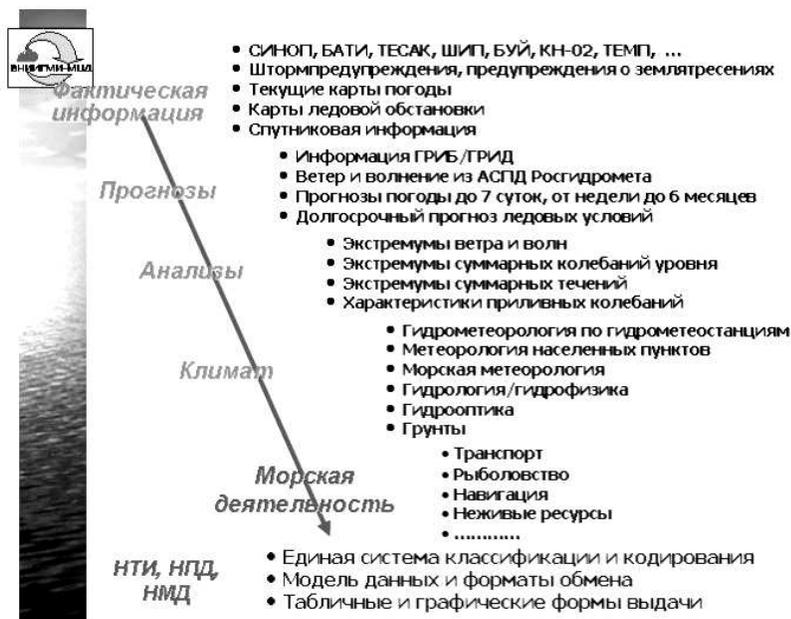


Рис. 3. Информационные ресурсы ЕСИМО.



Рис. 4. Формы представления информации в различных АРМ.

Для повышения эффективности работы специализированных пользователей в портале ЕСИМО создаются автоматизированные рабочие места (АРМ), для которых:

- виртуально сегментируются источники данных (создаются профили ИР) по различным признакам (типу информации, географической области, параметрам наблюдений и др.);

- генерируются персонализированные страницы для отдельных или типовых классов пользователей;

- разработано три типа генерации профилей ИР (подписка пользователя АРМ на определенные информационные ресурсы; генерация информационных ресурсов под стандартное в зависимости от географического района меню; генерация ИР под любой, заданный пользователем состав меню).

Созданы следующие опытные версии АРМ ЕСИМО:

- для общего пользования;

- для федеральных органов власти (Росгидромета, Минтранса, Агентства по рыболовству, МЧС, МПР, ГУНиО, РАН);

- региональные АРМ “Арктика”, “Антарктика”, “Каспий”, “Черное и Азовское моря”.

Ведутся работы по созданию АРМ для лица, принимающего решения на объектах экономики (морской порт, судно, местная администрация и др.). Формы представления информации в различных АРМ даны на рис. 4.

Публикация данных на портале — это всего лишь небольшое, но очень важное звено процесса эффективной доставки информации конечным пользователям. Поэтому при формировании отчетов разработчики уменьшили объем выдаваемой информации до 1–2 страниц. Пользователям сообщается, сколько страниц выбрано, и отображается только, например, первые десять выбранных экземпляров ИР или перед поиском задается количество представляемых на странице документов. Это позволяет, не дожидаясь загрузки всего файла, увидеть результат на экране. Часть информационных ресурсов представляется в виде списков и передается по электронной почте по технологии PUSH, что также экономит время пользователя при их просмотре, так как загружается только один ИР, выбранный пользователем.

2. Перспективы развития

Перспективными направлениями развития портала ЕСИМО являются:

- создание виртуального центра;

- применение концепции семантической сети [7];

- мониторинг вычислительных, информационных и сетевых ресурсов.

Консолидация ресурсов и интеграция информационных систем превращают портал ЕСИМО в виртуальный центр (рис. 5). Функции обычного центра данных (ввод описаний ИР, каталогизация, хранение, прикладная обработка, обмен, использование данных) перемещаются в распределенную среду. Управление ЕСИМО осуществляется в координационном центре, а выполнение функций — децентрализовано.

Внутри виртуального центра данных приложения строятся на базе компонентов и служб, которые отвечают потребностям пользователей, взаимодействующих с множеством информационных ресурсов и пользователей. Системы хранения, вычислительные и сетевые ресурсы приобретают виртуальный характер. При этом более широко должны использоваться Web-сервисы, прообразом которых сейчас являются java-классы (интерфей-

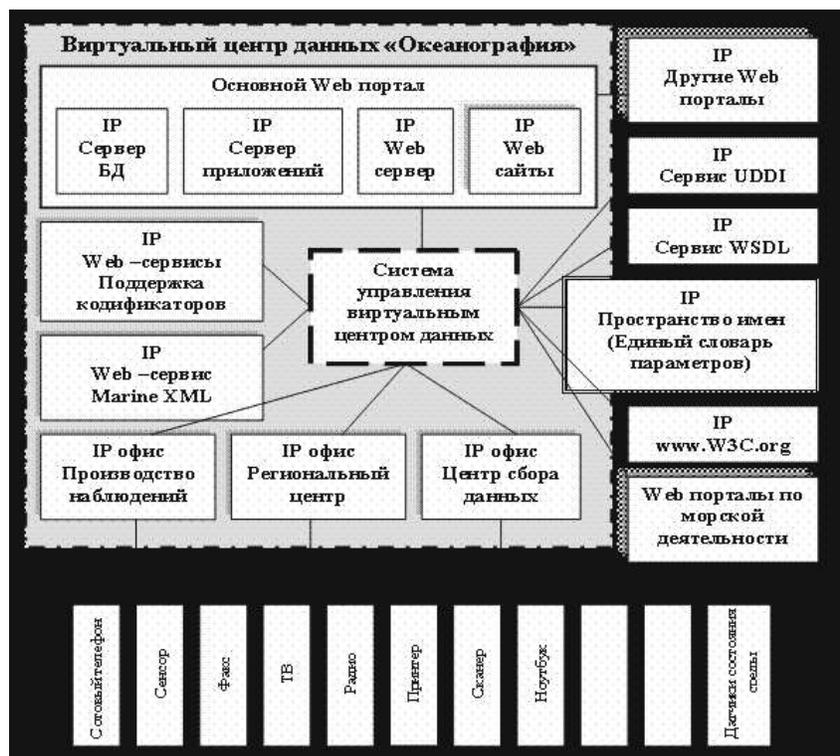


Рис. 5. Схема функционирования виртуального центра данных.

сы для визуализации отдельных объектов метаданных). Сервисами могут быть службы ведения кодификаторов, получения климатических характеристик, обмена данными различных типов и видов. Web-сервисы упрощают обмен информацией между различными приложениями, в том числе удаленными, за счет передачи данных в виде XML-файлов с использованием согласованной схемы и стандартов W3C (Universal Description, Discovery, and Integration — UDDI, Web services Description Language — WSDL, Simple Object Application Protocol — SOAP).

Создание семантической сети предусматривает объединение разных видов информации в единую глобальную XML-схему, где каждому параметру наблюдений и атрибуту описания метаданных будет присвоен свой тэг. Часть параметров на основе классификационных признаков будет объединена в группы, например, группа “волнение” включает параметры — высоту, длину, период волн. Некоторые группы могут включаться в другую — более общую группу. Все тэги составят единую иерархическую структуру. Для обеспечения единого понимания применяемых кодификаторов в рамках такой схемы создаются разделы описания кодификаторов (название, уровень стандартизации и применяемые значения кодов). В зависимости от названия и применяемого уровня стандартизации данные автоматически будут перекодироваться. Таким образом, на основе такой схемы все приложения смогут понимать смысл представляемой им информации.

На портале ЕСИМО имеются средства мониторинга (получение сведений о состоянии и пополнении ИР, посещаемости различных разделов портала в различных временных масштабах обобщения, актуальности ссылок и др.). Мониторинг информационных и вычислительных ресурсов позволяет создавать механизмы управления ИР и приложениями. При этом необходимы инструменты не только для оперативного сбора статистики, но и для идентификации узких мест, анализа тенденций в производительности сервера.

Для прогнозирования развития событий в виртуальном центре должна использоваться индикаторная информация: объем оперативной памяти, требуемый для выполнения приложений; актуальность ИР в сети; загруженность каналов связи и др. Эти индикаторы должны отслеживаться, помещаться в БД и непрерывно анализироваться на предмет выявления их критических значений.

Создание системы управления работой виртуального центра позволит упростить работу администратора портала по поддержке информационно-коммуникационных технологий портала ЕСИМО. После разработки такой системы активное участие администратора портала потребуется только в случае принятия критически важных решений. Это позволит предотвращать повторение проблемных ситуаций, а также обеспечить управление функционированием портала с упреждающим характером.

Портал ЕСИМО в силу большого объема данных, значительного числа интегрированных ресурсов, необходимости оперативного сбора и передачи данных по телекоммуникационным каналам, нахождения в стадии эксплуатации может быть эффективной областью применения GRID-технологий [8].

Заключение

Создана технология интеграции информационных ресурсов по морской деятельности. Разработана первая очередь портала ЕСИМО, включающая текущую, прогностическую и климатическую, социально-экономическую информацию. Информационные ресурсы стали доступными пользователю в любой момент, в любой точке нахождения пользователя, по любому району, в цифровой, графической или текстовой формах. В ЕСИМО наряду с гидрометеорологическими данными представлены ИР по другим аспектам обстановки в Мировом океане — биоресурсы, геология-геофизика, батиметрия, право и др. Доступ к информационным ресурсам ЕСИМО осуществляется по адресу <http://data.oceaninfo.ru/resource/index.jsp>.

Основными задачами развития технологии интеграции гетерогенных информационных ресурсов ЕСИМО являются:

- существенное увеличение полноты описания ИР и баз метаданных;
- комплексное информационное обеспечение морской деятельности на основе требований морской национальной политики, изложенной в Морской доктрине России до 2020 года, на основе автоматизированных рабочих мест;
- реализация сквозной технологической схемы потоков информации от наблюдения до принятия решения.

Список литературы

- [1] Вавилов К., Щербина С. Web-интеграция // Электронный журнал “Открытые системы”. 2001. Вып. 1. 12 с. <http://www.osp.ru/os/2001/01/043.htm>
- [2] Елманова Н. Web-порталы: назначение, преимущества, особенности и средства // КомпьютерПресс. 2002. Вып. 6. http://www.interface.ru/fset.asp?Url=/misc/mnogoe_1.htm
- [3] Вязилов Е.Д., Михайлов Н.Н., Белов С.В., Сухоносков С.В. Технологические аспекты организации доступа к разнородным информационным ресурсам ЕСИМО // Пятая

Рос. научно-техн. конф. “Современное состояние и проблемы навигации и океанографии” (НО-2004), 10–12 марта 2004. Т. 2. С. 178–180.

- [4] Вязилов Е.Д., Михайлов Н.Н., Чепурнов В.М. Web-портал Единой системы информации об обстановке в Мировом океане (ЕСИМО): методы построения и реализации // Информационные технологии Государственного фонда данных о состоянии окружающей природной среды. Росгидромет. ВНИИГМИ-МЦД. СПб.: Гидрометеиздат, 2002. Вып. 170. С. 123–144.
- [5] Шаймарданов М.З., Михайлов Н.Н., Воронцов А.А. и др. Методы и решения построения региональных интегрированных систем информационного обеспечения морской деятельности на примере Баренцева моря // Тр. РАО-03. СПб., 2003. С. 308–311.
- [6] VYAZILOV E., MIKHAILOV N., BELOV S. ET AL. Technological aspects of heterogeneous information resources integration // 19th Intern. CODATA Conf. “The Information Society: New Horizons for Science”. Berlin. 7–10 Nov. 2004.
<http://www.codata.org/04conf/abstracts/Interop/Vyazilov-TechnologicalAspects.htm>
- [7] Башмаков А.И., Башмаков И.А. Интеллектуальные информационные технологии: Учеб. пособие. М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. 304 с.
- [8] Черняк Л. Web-сервисы, grid-сервисы и другие // Открытые системы. 2004. Вып. 12.
<http://www.osp.ru/os/2004/12/020.htm>

Поступила в редакцию 2 июня 2005 г.