УДК 004.6

ПРОБЛЕМА ИНТЕГРАЦИИ ДАННЫХ МОНИТОРИНГА ПОДЗЕМНЫХ ВОД

Шарапов Р.В.

Муромский институт ФГБОУ ВПО «Владимирский государственный университет им. А.Г. и Н.Г. Столетовых», Муром, e-mail: mivlgu@mail.ru

В работе рассматривается проблема интеграции данных мониторинга подземных вод. Рассматриваются проблемы сбора и обработки данных о состоянии подземных вод, получаемых из различных источников. Приводятся особенности интеграции данных на логическом и физическом уровнях. Интеграция на физическом уровне предполагает сбор всех данных в единое хранилище и перевод их в единую форму. Приводятся особенности интеграции на основе единой базы данных и XML хранилища. Разбираются варианты хранения данных на основе XML разметки. Интеграция на логическом уровне предполагает создание некоторой надстройки над уже существующими данными без их физического преобразования.

Ключевые слова: вода, подземные воды, мониторинг, интеграция

PROBLEM OF GROUNDWATER MONITORING DATA INTEGRATION

Sharapov R.V.

¹Murom Institute of Vladimir State University, Murom, e-mail: mivlgu@mail.ru

In this paper we consider the problem of data integration groundwater monitoring. We consider the problem of collecting and processing data on the state of groundwater from different sources. We are given especially data integration at the logical and physical levels. The integration of the physical layer involves the collection of data in a single repository and translates them into a single form. We give particular integration based on a single database and XML repository. We understand data storage options for XML-based markup. Integration at the logical level, involves the creation of a superstructure over existing data without requiring a physical transformation.

Keywords: water, ground water, monitoring, integration

Подземные воды – важный компонент нашей планеты. Несмотря на то, что они располагаются в толщах литосферы и не видны человеку, они оказывают существенное влияние на многие процессы, протекающие на Земле [2, 4]. В связи со значительной ролью подземных вод важное значение приобретает необходимость проведения их мониторинга. В настоящее время непосредственный мониторинг за состоянием подземных вод проводится Центром государственного мониторинга состояния недр ФГУГП «Гидроспецгеология». Центр использует наблюдательную сеть, развернутую на территории России. Кроме того, наблюдение за подземными водами осуществляется различными службами и потребителями, осуществляющими недропользование. Информация мониторинга подземных вод используется различными потребителями (как государственными службами, так и коммерческими организациями).

Разнородность данных. В ходе исследований, проводимых на базе Муромского института (филиала) Владимирского государственного университета, автор столкнулся с проблемой сбора и обработки данных о состоянии подземных вод, получаемых из различных источников [6]:

Каждая служба, так или иначе проводящая наблюдения за подземными водами, использует свои формы представления данных: консолидированные данные в таблицах MS Excel или базах данных, отдельных отчетах в MS Word, журналах, заполненных от руки и т.д.;

Различные службы могут проводить контроль различного набора параметров;

Сбор данных производится различными службами с разной периодичностью;

Точность измерения данных различными службами может существенно различаться (например, число знаков после запятой при вычислении концентрации (мг/дм^3), показатель цветности, изменяемый с шагом 1 или 5 в разных исследованиях);

Полученные при наблюдении данные могут иметь разные единицы измерения.

Все это сильно затрудняет возможность совместного использования полученных таким образом данных [8, 10]. Возникает задача интеграции данных мониторинга подземных вод с целью их совместного использования [9].

Интеграция данных. Существует несколько направлений интеграции данных мониторинга подземных вод. Интеграция на физическом уровне предполагает сбор всех данных в единое хранилище и перевод

их в единую форму. При этом вся информация о наблюдениях того или иного процесса преобразуется к единому типу и размерности (например, г/м³ для концентрации) [7].

В качестве среды хранения в данном случае можно предложить несколько решений:

Единая база данных

XML-хранилище.

Единая база данных может представлять собой набор таблиц, позволяющих хранить в себе данные, полученные при проведении различных исследований и наблюдений [3]. В качестве СУБД может использоваться MS SQL Server, Oracle, Firebird, MySQL и т.д. Преимуществом такого решения является полная унификация хранения данных в едином месте. Недостатком является жесткость структуры и сложность ее модификации. Кроме того, обмен информацией между различными службами будет требовать поддержки конкретной СУБД или экспорта данных в какой-то промежуточный формат.

XML-хранилище представляет собой набор из одного или нескольких структурированных файлов в формате XML (eXtensible Markup Language) представляет собой текстовый формат, который предназначен для хранения структурированных данных и обмена информацией между программами. Также XML использу-

ется для создания на его основе более сложных специализированных языков разметки.

XML-разметка за последнее десятилетие получила широчайшее распространение [5]. Причин для этого множество. XML позволяет описывать различные структуры, в том числе записи, списки и деревья. Он является самодокументируемым, то есть способен сам описывать структуры данных, имена полей и их значения. В отличие от бинарных форматов, XML содержит метаданные об именах, типах и классах описываемых объектов, по которым программы могут обрабатывать документы неизвестной структуры. XML не зависит от платформы и позволяет работать с различными кодировками, в том числе с Юникодом. Кроме того, стандарт XML основан на международных стандартах, что позволяет использовать его во всем мире без каких-либо проблем [1]. Ну и последним его преимуществом является то, что XML представляет данные в текстовом виде, что позволяет их так или иначе просмотреть без использования специальных средств. Таким образом, XML стал мощным средством представления информации.

Представление данных мониторинга при использовании XML-разметки будет иметь следующий вид:

```
<?xmlversion=»1.0»encoding=»UTF-8»?>
<research name=»1» date=»2013-06-01»>
<field name=»chromaticity» type=»int» length=»2» precision=»1»>11</field>
<field name=»hardness» type=»float» length=»16» precision=»0.1»>5.7</field>
...
</research>
```

Рис. 1. Пример представления данных с использованием XML

Особенностью такой организации является возможность добавлять в файл дополнительную информацию. Например, при необходимости добавить какие-то коммен-

тарии к изображению или сведения о диагнозе, в файл достаточно добавить строку следующего вида:

```
<field name=»Comment» researchid=»1» type=»text»>Следует провести повторный сбор проб.</field>
```

Рис. 2. Добавление комментариев к данным мониторинга

Конечно, XML позволяет заменить эту конструкцию и более простой конструкцией:

```
<comment> Следует провести повторный сбор проб.</comment>
```

Рис.3. Добавление комментариев с использованием конструкции comment

Тем не менее, использование конструкций подобного вида требует их предварительного описания. Это необходимо, в пер-

вую очередь, для того, чтобы правильно интерпретировать передаваемую информацию. Использование же конструкции field

(рис. 2) такого предварительного описания не требует.

Применение XML хранилища позволяет обеспечить хранение данных различной структуры, гибкость в модификации данных, возможность добавлять новые структурные элементы без изменения других данных и т.д. Кроме того, XML хранилище предоставляет наиболее широкие возможности по организации обмена информацией между различными системами.

Интеграция на логическом уровне предполагает создание некоторой надстройки над уже существующими данными без их физического преобразования. Эта надстройка представляет собой набор шаблонов и правил преобразования данных в структуры и форматы, используемые той или иной службой. Преимуществом интеграции на логическом уровне является то, что каждая служба работает со своими данными в привычном для них формате, а данные из других служб с помощью набора шаблонов преобразуются в форму, удобную каждой службе. В случае необходимости изменения формы представления данных в конкретной службе, меняется только нужный шаблон.

С другой стороны, интеграция на логическом уровне требует наличия правил для каждого потребителя информации. В случае изменения структуры данных одного из источников, необходимо изменение шаблонов для каждого потребителя данных мониторинга подземных вод.

Заключение

Интеграция данных мониторинга подземных вод имеет важное значение для большого количества служб. Интеграция возможна как на физическом, так и на ло-

гическом уровнях. Наиболее удобной для обработки является интеграция на физическом уровне в виде единой базы данных или XML хранилища, содержащих данные наблюдений из различных источников в единой форме.

Работа выполнена при поддержке гранта РФФИ № 13-07-97510.

СПИСОК ЛИТЕРАТУРЫ

1. Спенсер П. XML. Проектирование и реализация. Программирование с помощью XML, ASP и IE5. – М.: Лори, 2001. – 510 с.
2. Шарапов Р.В., Димакова Н.А. Проблема загрязнения подземных вод // Современные наукоемкие технологии, 2013. № 2, С. 79-82.
3. Шарапов Р.В. Аппаратные средства хранения больших объёмов данных // Инженерный вестник Дона, 2012, № 4 часть 2, С.20-23.
4. Шарапов Р.В. Глобальные экологические катастрофы: миф или реальность? // Машиностроение и безопасность жизнедеятельности, № 1, 2011. – С. 14-16.
5. Шарапов Р.В. Организация обмена графическими данными между медицинскими системами // Биотехносфера, 2012, № 5-6, С.88-92.
6. Шарапов Р.В. Принципы мониторинга подземных вод // Машиностроение и безопасность жизнедеятельности, 2012, № 3, С. 27-30.
7. Шарапов Р.В. Структура системы мониторинга подземных вод // Машиностроение и безопасность жизнедеятельности, 2012, № 4, С. 20-23.
8. Шарапов Р.В., Шарапова Е.В. Некоторые вопросы применения новых информационных технологий при моделировании чрезвычайных ситуаций // Машиностроение и безопасность жизнедеятельности. 2008. № 5. С. 62-66.
9. Шарапов Р.В., Шарапова Е.В. Проблема интеграции электронных коллекций состояний экосистем // Машиностроение и безопасность жизнедеятельности. 2009. № 6. С. 75-78.
10. Шарапов Р.В. Применение информационных технологий в задачах моделирования чрезвычайных ситуаций // Вестник Пермского национального исследовательского политехнического университета. Охрана окружающей среды, транспорт, безопасность жизнедеятельности. 2011. № 2. С. 162-167.