

*10-ая Всероссийская научная конференция
«Электронные библиотеки: перспективные методы и технологии,
электронные коллекции»*

RCDL'2008, Дубна, Россия, 2008

Унификация структур данных в области изучения, освоения и использования ресурсов Мирового океана

Е.Д. Вязилов, А.А. Федорцов, А.Е. Кобелев
Всероссийский научно-исследовательский институт
гидрометеорологической информации –
Мировой центр данных

vjaz@meteo.ru

Актуальность

Создаваемые в разных организациях базы данных не имеют общей концептуальной основы и по большому счету не совместимы друг с другом.

Интеграция разнородных, распределенных информационных ресурсов и совместное использование заложенной в них информации становится трудоемким процессом

Задачи

- Стандартизация моделей, схем и структур данных
- Использование стандарта ISO 19115, спецификаций SensorML, TML, MODBUS
- Развитие широкого спектра объектов метаданных – кроме сведений о массивах и БД, создание описаний форматов, проектов, организаций, наблюдательных платформ, рейсов научно – исследовательских судов и атрибутов, др.
- Учет жизненного цикла объектов при создании моделей данных

Некоторые определения

- **Тип данных** - Точка, профиль, сетка, объектный файл
- **Подтип данных** – Точка случайная, временной ряд, сетка регулярная и нерегулярная и т.д.
- **Экземпляр объекта** - Запись, характеризующая единицу объекта (временной ряд, отходы, судно и т.п.), минимальная единица хранения
- **Унифицированный элемент данных** - это поименованный атрибут данных, находящийся под управлением информационной системы
- **Атрибут метаданных** - Спецификация характеристик получения данных
- **Атрибут данных** - Спецификация измеряемых характеристик объекта
- **Значение атрибута** - Значение, присвоенное характеристике конкретного экземпляра объекта или факта
- **Жизненный цикл объекта** - Совокупность взаимосвязанных процессов изменения состояния объекта при его создании и использовании
- **Стадия жизненного цикла** - Часть жизненного цикла объекта, которая характеризуется спецификой направленности работ, производимых на этой стадии, и конечными результатами
- **Факты** - Совокупность событий, которые могут происходить с объектами в соответствии со стадиями ЖЦ

Словарь атрибутов

Раздел «Метаданные» :

- общей идентификации данных;
- принадлежности стране, организации, автору и др.;
- спецификации производства (получения) данных – платформа, метод, прибор и др.;
- географические характеристики – районы, высота над уровнем моря, глубина, толщина слоя, расстояние, местоположение платформы (широта, долгота), направление движения платформы;
- временные характеристики (год, месяц, день, время начала и окончания события).

Раздел «Данные» :

- морская среда;
- морская деятельность;
- социально – экономическая информация (СЭИ)

Структура имени унифицированного элемента: ANNNN_SS_I (M, Q)

- где, **A** – раздел словаря (P, M или R),
- **NNNN** – числовой код, определяющий атрибут наблюдений (измерений, вычислений);
- **SS** – числовой код, отражающий название статистической характеристики значения атрибута (измеренное значение, минимум, максимум, среднее, повторяемость, вероятность, аномалию, продолжительность явления, число случаев, сумму, тенденцию);
- **I** – интенсивность явления, выраженная в значении атрибута;
- **M** – метод получения атрибута;
- **Q** – признак качества значения атрибута.

Примеры унифицированных элементов представлены ниже: P0295_00 - температура воды, измеренная; P0295_01_Q – признак качества температуры воды, минимальной

Хранение множественных значений атрибутов

Проблема. Один объект имеет несколько значений для одного атрибута:

- страна – автор - проектировщик, производитель, владелец, разработчик, администратор;
- организация – производитель, владелец, разработчик (застройщик), оператор;
- персона - владелец, разработчик;
- проект – автор, исполнитель, участник

Структура хранения множественных значений атрибутов:

- **Объект: идентификатор** (прибор, судно, организация, др.)
- **Объект: ЖЦ** (проектирование, разработка, эксплуатация, др.)
- **Дата: значение**
- **Атрибут: название** (страна, организация, персона, проект, др.)
- **Атрибут: значение** (идентификатор страны, организации, персоны, проекта)

Для жизненного цикла документа используются другие стадии: дата создания, описания, модификации (редактирования), рассмотрения, согласования, принятия, ратификации, утверждения (принятия) с одной и другой стороны, подписания, вступления в силу документа, передачи в печать, издания (публикации), срок действия, дата доставки, уничтожения.

- **Документ: идентификатор**
- **Дата: ЖЦ**
- **Дата: значение**

Унифицированная модель данных

Подходы:

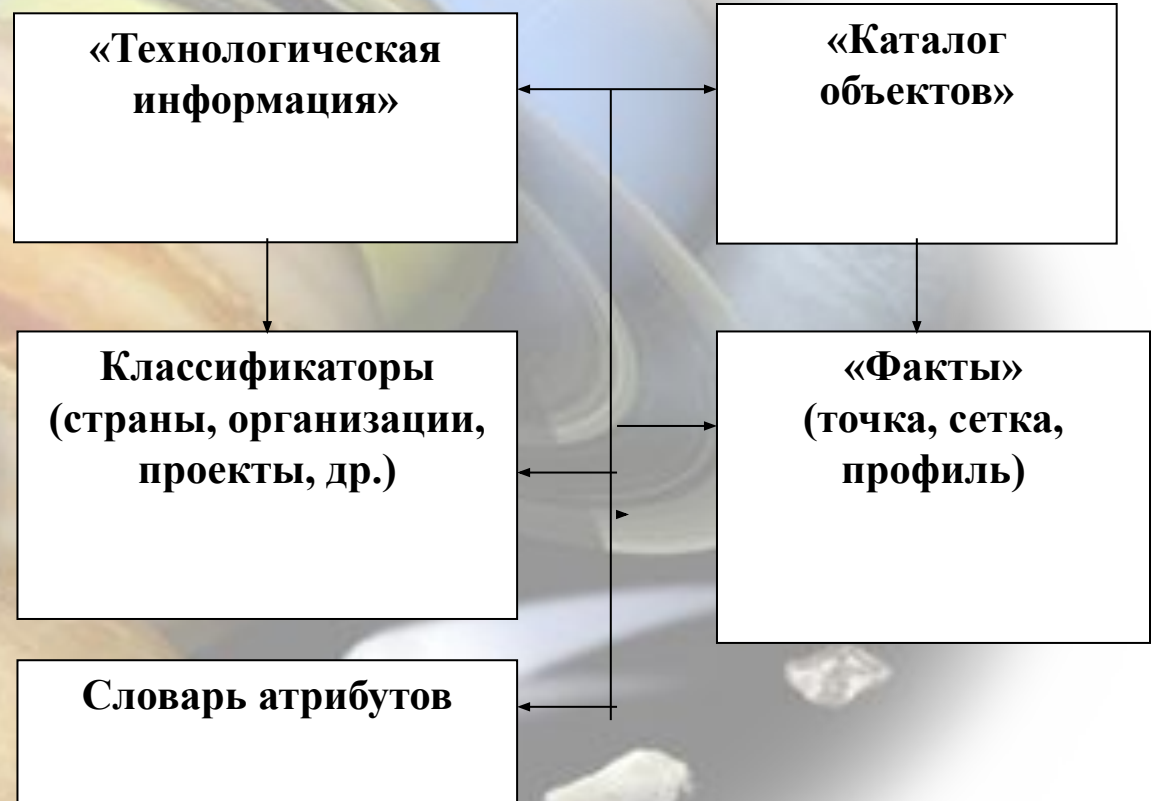
Каждый объект имеет две таблицы: сведения об объекте и факты ЖЦ объекта

Все таблицы хранятся в многомерной модели данных

Все классификаторы хранятся в одной таблице по структуре (ID, код, значение)

Выделен отдельный объект «Сведения об состоянии объектов предметной области в БД» - Технологическая информация

Словарь атрибутов – это фактически тоже отдельный объект – пока храниться отдельно



Каталог объектов

Каталоги конкретных объектов могут очень сильно отличаться по составу атрибутов, поэтому при использовании многомерной модели хранения данных таблица «Каталог объектов» имеет следующую структуру данных:

- **Идентификатор экземпляра объекта**
- **Идентификатор категории объекта (каталога)**
- **Имя атрибута (унифицированный элемент каталога)**
- **Значение атрибута (каталога)**

Факты



Таблица «Факты» отражает пространственно – временные координаты объектов, находящихся на различных этапах жизненного цикла. Данные могут быть представлены в виде:

- точки (временной ряд)

- сетки

- профиля

Таблица «Факты» должна включать атрибуты:

- идентификатор объекта

- жизненный цикл объекта

- дата ЖЦ

- имя атрибута

- значение

В зависимости от ситуации могут использоваться дополнительные атрибуты ((если значения этих атрибутов изменяются):

- страна

- организация

- персона

- проект

Сведения об унифицированных элементах

- Идентификатор
- Имя
- Временной, пространственный масштабы
- Полное и краткое наименование атрибута на русском и английском языках
- Точность наблюдений или рекомендуемая точность расчета
- Единицы измерения
- Диапазон изменчивости (мин, макс)
- Вертикальное разрешение (мин и макс высота/глубина)
- Используемый классификатор для значений атрибутов (международный, национальный)
- Метод определения атрибута
- Международный код атрибута
- Описание

Технологическая информация

Предназначена для мониторинга состояния таблиц «Каталог объектов» и «Факты» и включает:

- дату ввода
- все даты редактирования
- кто вводил и редактировал
- показатели использования данных
- другую информацию

Примеры реализации многомерных структур

- **Временные ряды.** Таблицы Сведения о временном ряде и Матрицы значений.
 - Сведения о временных рядах: идентификатор ряда, станции, широту, долготу, дату начала и окончания наблюдений, горизонт, дискретность - месяц, масштаб пространственного осреднения, кто создал, когда, дату последней редакции, название исходного массива.
 - Матрица временного ряда значений включает ID ряда, год, месяц, день, время, значение.
- **Сеточные данные.** Таблицы Сведения о сетке, поле, значения в узлах сетки
 - Сведения о сетке включают уникальный код, тип сетки, ссылку на источник данных, ID модели, размерность сетки по X и Y, шаг сетки в град по X и Y, координаты района, временной масштаб осреднения (сутки, месяц, год), пространственный масштаб осреднения (квадрат, трапеция, точка), шаг при отрисовке изолиний.
 - Сведения о поле включают ID заголовка, ссылку на сетку, значения года, месяца, дня, часа, минут наблюдений, имя атрибута, тип уровня, значение уровня.
 - Данные в узлах сетки включают идентификатор точки, широту, долготу, значение атрибута.
- **Профиль.** Таблицы сведения о профиле и данные по профилю
 - Сведения о профиле включают идентификатор профиля и пространственно-временные координаты профиля.
 - Данные по профилю включают идентификатор профиля, уровень, имя атрибута и значение.
- **Объектные файлы.** Каталог со ссылками на файлы с документами, картами и рисунками, включает идентификатор, класс объекта, координаты района, дату, др.)

Заключение

- Унификация процессов управления данными возможна только через рассмотрение стадий жизненного цикла объектов, сведения о которых отражаются в БД
- Необходимо создавать каталоги для организации хранения и поиска объектных файлов (документов, графических файлов, др.), а также для любой информации, представленной в виде сведений о каких-либо объектах.
- В сущностях, для которых создаются БД, необходимо выделять общие поля атрибутов метаданных типа идентификатор, дата/время (создания, редактирования, хранения), георайон, платформа, др., в т. ч. поля для организации связи между объектами и т.д.
- Предложенные структуры данных могут быть использованы в других предметных областях, при этом не надо добавлять новые сущности и поля
- Подана заявка на изобретение «Универсальный способ организации данных»

Литература

- Есин В.И., Пергаменцев Ю.А. Технология проектирования модели предприятия на основе универсальной модели данных. 2008. <http://www.citforum.ru/database/articles/udm/>
- Кузнецов С. Будущие направления исследований в области баз данных: десять лет спустя». - 1999. http://www.citforum.ru/database/articles/future_01.shtml. - 5с.
- Кузнецов С. Кризис технологии СУБД и пути его преодоления. Конференция «Корпоративные базы данных 2008». М. 24-25 апреля 2008 г. http://www.citforum.ru/seminars/cbd2008/2_8.ppt.
- Марков Б. Проектирование систем регистрации и анализа данных. www.citforum.ru/database/articles/reg_data.shtml
- Пергаменцев Ю. Проектирование БД на основе универсальной модели данных. Конференции "Корпоративные базы данных 2002". <http://www.citforum.ru/seminars/cbd2002/111.shtml>
- Стоунбрейкер Майкл, Кетинтемел Угур. Один размер пригоден для всех»: идея, время которой пришло и ушло. Перевод: С.Кузнецов http://www.citforum.ru/database/articles/one_size_fits_all/
- Bernstein Phil, Brodie Michael, Ceri Stefano, DeWitt David, Franklin Mike, Garcia-Molina Hector, Gray Jim, Held Jerry, Hellerstein Joe, Jagadish H.V., Lesk Michael, Maier Dave, Naughton Jeff, Piranesh Hamid, Stonebraker Mike, and Ullman Jeff. The Asilomar Report on Database Research. http://www.citforum.ru/database/digest/asil_01.shtml. September 1998. - 12 с. (перевод).
- Botts Mike. OpenGIS® Sensor Model Language (SensorML), Implementation Specification, Version 1.0 // Open Geospatial Consortium Inc., 2006. – 118 с.
- Havens Steve. OpenGIS® Transducer Markup Language, Implementation Specification, Version 1.0.0 // Open Geospatial Consortium Inc., 2006. 258 pp. <http://xml.coverpages.org/OGC-06-010r6-TransducerMarkupLanguage-TML.pdf>