

*Шестая Всероссийская Открытая конференция  
«СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО  
ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА  
(Физические основы, методы и технологии мониторинга  
окружающей среды, потенциально опасных явлений и  
объектов)  
(Москва, ИКИ РАН, 10-14 ноября 2008 г.)*

## **УНИВЕРСАЛЬНАЯ МОДЕЛЬ ХРАНЕНИЯ ДАННЫХ С УЧЕТОМ ЖИЗНЕННОГО ЦИКЛА ОБЪЕКТОВ**

Е.Д. Вязилов, А.А. Федорцов  
Всероссийский научно-исследовательский институт  
гидрометеорологической информации –  
Мировой центр данных  
[vjaz@meteo.ru](mailto:vjaz@meteo.ru)

## **Актуальность**

- ✦ Уже имеются БД, включающие тысячи таблиц
- ✦ Создаваемые в разных организациях БД 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**

- ✦ Развить универсальную модель, схемы и структуры данных и провести их стандартизацию
- ✦ Максимально использовать серию стандартов 19100, включая ISO 19115, спецификации SensorML, TML, MODBUS и др.
- ✦ Развивать широкий спектр объектов метаданных
- ✦ Учитывать жизненный цикл объектов при создании моделей данных

## Предыдущие исследования

### 1988 и 1998 гг. были составлены прогнозы будущих направлений исследований в области развития БД

- ✦ Имеет смысл работать над стандартной моделью данных следующего поколения", «стандартную модель многие хотели бы видеть, но пока шансов немного
- ✦ Отмечалась потребность в стандартном представлении данных, но для этого требуются более широкие справочники данных (метаданные)
- ✦ Несмотря на существенные продвижения в области создания метаданных их стандартизация все еще остается проблемой
- ✦ Название одного из разделов прогноза «Умные данные должны унифицировать процессы и данные в системе БД» достаточно красноречиво говорит о том, что УМД должна основываться на метаданных, которые должны не только идентифицировать данные, но и управлять процессами их обработки
- ✦ За прошедшие годы после выдвижения идеи создания машин БД (реализация ряда функций управления данными на аппаратном уровне) продвижений очень мало

## Недостатками существующих подходов по созданию УМД

- ✦ Отсутствие словаря описаний атрибутов
- ✦ Нет единой схемы хранения классификаторов
- ✦ Слабое использование унифицированных атрибутов метаданных при описании объектов и событий
- ✦ Трудности представления данных в виде временных рядов, профилей и сеток
- ✦ Существующие технологии:
  - ориентированы на проектирование БД для одной предметной области
  - не обеспечивают расширение состава таблиц и атрибутов в БД,
  - не достаточно способствуют использованию типовых проектных решений по созданию БД
  - значительно увеличивают число таблиц в интегрированных БД
- ✦ Есин В.И., Пергаменцев Ю.А. Технология проектирования модели предприятия на основе универсальной модели данных. 2008.  
<http://www.citforum.ru/database/articles/udm/>
- ✦ 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/](http://www.citforum.ru/database/articles/one_size_fits_all/)
- ✦ Хранилище данных и аналитические системы: технология и инструментальные средства корпорации Oracle. 2007. с.9

## 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
- ✦ **УМД** - это стандартная для разных предметных областей схема БД, созданная с помощью средств описания концептуальной схемы БД,

## Совокупность существенных признаков УМД

- ✦ Создание **единого словаря описаний** унифицированных **атрибутов**, в т.ч. атрибутов метаданных для описания объектов и событий
- ✦ Использование минимальной **атомарной единицы** хранения - значения атрибута данных с помощью многомерной схемы представления как для данных, так и метаданных. Многомерная модель данных позволяет хранить данные в виде **трипла** (идентификатор записи, имя атрибута значение атрибута)
- ✦ Применение единой схемы хранения всех **классификаторов** в одной таблице (**ID, код, значение**), что позволяет легко добавлять новые значения кодов в существующие классификаторы и новые классификаторы при настройке на предметную область
- ✦ Выделение простых сущностей (объекты метаданных, товары и т.п.) в виде отдельных объектов УМД
- ✦ Выделение этапов жизненного цикла для всех выделяемых сущностей (объектов), предназначенных для хранения в УМД;
- ✦ Выделение статической справочной (**Каталог**) и динамической (**Факты**) информации для каждого объекта в виде отдельных таблиц, связанных между собой и описывающих один объект
- ✦ Сохранение связей между таблицами «Каталог» и «Факты», экземплярами различных объектов в виде отдельной таблицы
- ✦ Хранение информации об объектах БД в виде **«технологической информации»** и их связях как один из объектов БД - **«Сведения об состоянии объектов предметной области в БД»**

# Схема универсальной модели данных



# Факты

- **Факты** возникают как результат реализации некоторой программы измерений
- Атомарной единицей хранения факта является экземпляр данных (совокупность данных), соответствующих однократной реализации программы измерений, вычислений или обобщений в какой-то момент (срок) или интервал (сутки, месяц) времени
- **Каталоги** – это статическая информация о сущностях
- Заранее неограниченное многообразие объектов распределяется по следующим сущностям - **каталог объектов**, включающий технологическую информацию и определяющий состояние всех компонент модели данных; **каталог экземпляров объектов**; таблицы **фактов** состояния объекта; классификаторы
- При любом расширении набора объектов, атрибутов объектов (метаданных, фактов, связей между ними) не надо добавлять новую структуру данных, поле
- Для каждого объекта создаются свои таблицы «Каталог» и «Факты»

# Словарь унифицированных элементов (атрибутов)

## Раздел «Метаданные» :

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

## Раздел «Данные» :

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

## Структура унифицированного элемента: ANNNN\_SS\_I (M, Q)

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

# Жизненный цикл

## Стандарт ИСО 19115 определяет элемент «Роль»

- ✦ **страна** – автор - проектировщик, производитель, владелец, разработчик, администратор
- ✦ **организация** – производитель, владелец, разработчик
- ✦ **персона** - владелец, разработчик
- ✦ **проект** – автор, исполнитель, участник

Структура данных с учетом ЖЦ объекта:

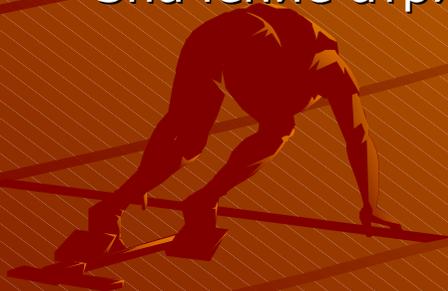
- ✦ **Объект: идентификатор** (прибор, судно, организация, др.)
- ✦ **Объект: ЖЦ** (проектирование, разработка, эксплуатация, др.)
- ✦ **Дата: значение**
- ✦ **Атрибут: название** (страна, организация, персона, проект, др.)
- ✦ **Атрибут: значение** (идентификатор страны, организации, персоны, проекта)

**Для жизненного цикла документа используются другие стадии ЖЦ:** 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**
- ✦ **Документ: ЖЦ**
- ✦ **Дата: значение**

## Таблица связей

- ✦ Идентификатор связи
- ✦ Идентификатор объекта, который связывается
- ✦ Идентификатор атрибута, по которому идет связь
- ✦ Идентификатор экземпляра объекта
- ✦ Значение атрибута
- ✦ Идентификатор объекта, с которым связываются
- ✦ Идентификатор экземпляра связи объекта, с которым связываются
- ✦ Значение атрибута





## Каталог объектов - технологическая информация

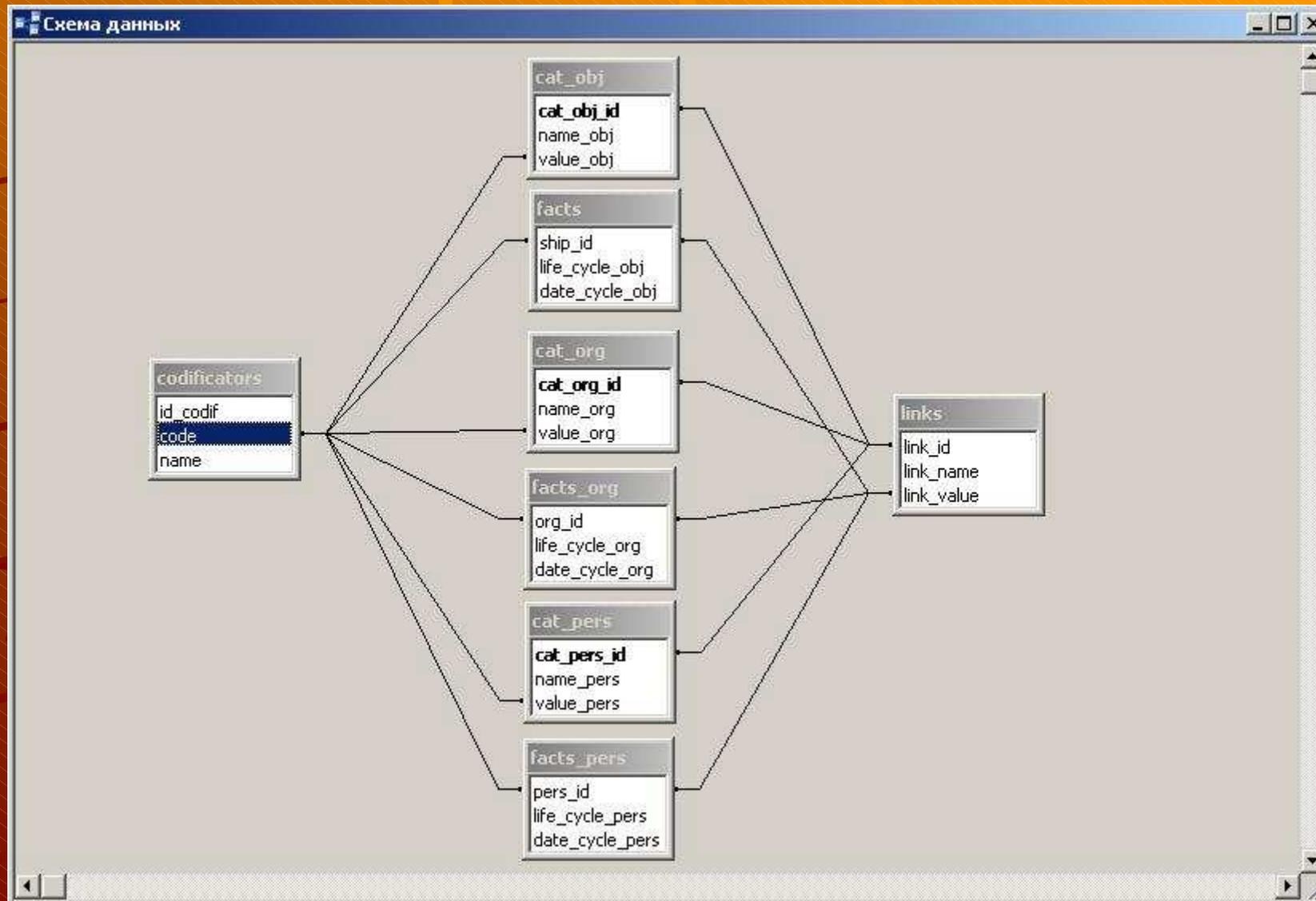
Предназначена для мониторинга состояния таблиц «Каталог объектов» и «Факты» и включает:

- ◆ Дату ввода
- ◆ Все даты редактирования
- ◆ Кто вводил и редактировал
- ◆ Показатели использования данных
- ◆ Другую информацию

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

# Пример схемы УМД в реляционной БД



# Примеры таблиц БД на основе УМД

SQL Window - Edit data of table YMD.CAT\_OBJ@WO\_NEW

select t.\*, t.rowid from ymd.cat\_obj t

## Каталог судов

CAT_OBJ_ID	NAME_OBJ	VALUE_OBJ	ROWID
1	10100002	vladelec	cat_org
2	10100002	value	10200001
3	10100002	captain	cat_pers
4	10100002	ship_id	74
5	10100001	ship_id	25372
6	10100001	name_eng	DLGA
7	10100001	length	100
8	10100001	breadth	30
9	10100001	displacement	6000
10	10100002	ship_id	1991
11	10100002	name_eng	STORQJEVOT
12	10100002	length	315
13	10100002	breadth	47
14	10100002	displacement	12125

1:1 14 rows selected in 0,046 seconds

SQL Window - Edit data of table YMD.FACTS@WO\_NEW

select t.\*, t.rowid from ymd.facts t

## ЖЦ судов

SHIP_ID	LIFE_CYCLE_OBJ	DATE_CYCLE_OBJ	ROWID
1	25372	дата спуска на воду	20.10.1983
2	25372	дата списания	15.12.2003
3	25372	выход в рейс 1	12.07.2001
4	25372	приход с рейса	05.09.2002
5	1991	дата спуска на воду	03.01.1989
6	1991	дата списания	
7	1991	выход в рейс АФМ	10.08.2009
8	1991	приход с рейса АФМ	29.01.2010

1:1 8 rows selected in 0,14 seconds

SQL Window - Edit data of table YMD.CAT\_ORG@WO\_NEW

select t.\*, t.rowid from ymd.cat\_org t

## Каталог организаций

CAT_ORG_ID	NAME_ORG	VALUE_ORG	ROWID
1	10200001	org_id	9
2	10200001	name_eng	Государственный океанографический институт
3	10200001	addres_id	9
4	10200001	country_id	149

1:1 4 rows selected in 0,031 seconds

SQL Window - Edit data of table YMD.CAT\_PERS@WO\_NEW

select t.\*, t.rowid from ymd.cat\_pers t

## Каталог персон

CAT_PERS_ID	NAME_PERS	VALUE_PERS	ROWID
1	10300001	pers_id	74
2	10300001	name	Левин И.М.
3	10300001	org_id	16
4	10300001	country_id	149
5	10300002	pers_id	253
6	10300002	name	Голубев В.А.
7	10300002	org_id	27
8	10300002	country_id	149

1:1 8 rows selected in 0,063 seconds

SQL Window - Edit data of table YMD.FACTS\_PERS@WO\_NEW

select t.\*, t.rowid from ymd.facts\_pers t

## ЖЦ персон

PERS_ID	LIFE_CYCLE_PERS	DATE_CYCLE_PERS	ROWID
1	74	birthdate	
2	253	birthdate	

1:1 2 rows selected in 0,078 seconds

SQL Window - Edit data of table YMD.FACTS\_ORG@WO\_NEW

select t.\*, t.rowid from ymd.facts\_org t

## ЖЦ организаций

ORG_ID	LIFE_CYCLE_ORG	DATE_CYCLE_ORG	ROWID
1	2829	birthdate	01.01.1989

1:1 1 row selected in 0,063 seconds

SQL Window - Edit data of table YMD.CODEFACTORS@WO\_NEW

select t.\*, t.rowid from ymd.codefacts t

## Классификаторы

ID_CODEIF	CODE	NAME
1	288	4087
2	288	4088
3	288	4086
4	288	4084
5	288	4085
6	117	24
7	117	9
8	117	16
9	334	3223
10	334	12326
11	334	24788
12	334	3212

1:1 12 rows selected in 0,031 seconds

SQL Window - Edit data of table YMD.LINKS@WO\_NEW

select t.\*, t.rowid from ymd.links t

## Связи

LINK_ID	LINK_NAME	LINK_VALUE
1	10200001	id_table_in
2	10200001	id_inst
3	10200001	id_strib
4	10200001	value
5	10200001	id_out
6	10200001	value
7	10200001	id_table_out
8	10200001	id_inst
9	10200001	id_strib
10	10200001	value
11	10200001	id_inst
12	10200001	id_strib
13	10200001	value
14	10200001	id_table_out
15	10200002	value
16	10200002	id_table_out
17	10100001	id_strib
18	10100001	id_inst
19	10100001	id_strib
20	10300002	id_out
21	10300002	value
22	10300002	id_table_in
23	10300002	id_inst
24	10300002	id_strib

1:1 24 rows selected in 0,157 seconds

## Этапы создания БД на основе УМД

- ✦ Анализ предметной области и составление полного списка атрибутов и их свойств для хранения в БД
- ✦ Выделение хранимых объектов (сущностей)
- ✦ Определение связей между выбранными объектами
- ✦ Создание каталога объектов
- ✦ Разработка каталога и таблиц фактов для выбранных объектов
- ✦ Составление словаря атрибутов, включаемых в выбранные объекты БД
- ✦ Маппинг имен, имеющих в локальных БД и УМД
- ✦ Анализ, выбор или разработка классификаторов для соответствующих атрибутов
- ✦ Разработка средств загрузки в созданные структуры данных
- ✦ Загрузка БД путем конвертирования данных в УМД
- ✦ Использование готовых или создание собственных приложений

## Преимущества УМД

- ✦ Может хранить данные в виде временных рядов, сеток, профилей, каталогов в стандартной структуре с фиксированным набором таблиц
- ✦ Сохраняет целостность при добавлении новых объектов и связей
- ✦ Не зависит от используемых СУБД и ОС
- ✦ Приложения становятся более универсальными
- ✦ Обладает расширяемостью по содержанию данных
- ✦ Независима от предметной области
- ✦ Позволяет хранить и манипулировать данными в терминах объектов, каталогов и фактов
- ✦ Добавлять по мере необходимости в БД новые объекты без изменения модели данных
- ✦ Разносит во времени и распределяет между разработчиками проектирование отдельных объектов БД и приложений
- ✦ Позволяет упростить проектирование, разработку, создание, эксплуатацию БД

**Недостаток** – усложняет понимание структуры данных, визуальный контроль полноты и качества 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](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](http://www.citforum.ru/seminars/cbd2008/2_8.ppt).
- ✦ Марков Б. Проектирование систем регистрации и анализа данных.  
[www.citforum.ru/database/articles/req\\_data.shtml](http://www.citforum.ru/database/articles/req_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/](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](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>