

# КОНСОЛИДАЦИЯ МЕТАДААННЫХ В ОБЛАСТИ НАУК ОБ ОКРУЖАЮЩЕЙ СРЕДЕ

Е. Д. ВЯЗИЛОВ

*Всероссийский НИИ гидрометеорологической информации —*

*Мировой центр данных, Обнинск, Россия*

e-mail: [vjaz@meteo.ru](mailto:vjaz@meteo.ru)

The sources of metadata are presented. Metadata development in Russia and abroad is reviewed. The place and the role of metadata at various stages of data processing are determined. The structure of metadata organization is provided.

## Введение

За последние десятилетия созданы тысячи исходных, инвертированных, расчетных, глобальных, региональных и локальных массивов и баз данных. Собрана информация о десятках тысяч экспедиций. Разработаны сотни различных программных средств, на основе которых можно получить тысячи расчетных характеристик.

В этом огромном потоке информации о состоянии природной среды необходимо очень быстро ориентироваться. И здесь имеющиеся бумажные каталоги, списки, справки, полученные с технических носителей, уже не помогают. Во-первых, информация очень быстро изменяется, массивы пополняются, корректируются и пересчитываются. Во-вторых, объем такой информации также относительно большой. Так, базы сведений об изученности океанографическими наблюдениями того или иного района оцениваются в десятки мегабайт. В-третьих, справочная информация — это основа для перехода на безбумажную технологию обработки информации о состоянии среды. Для фильтрации данных иногда требуются такие атрибуты, которых нет в исходных массивах данных. Для успешного поиска качественных данных необходимы сведения о массивах данных, форматах их обмена, программных средствах их обработки, организациях, собирающих и хранящих данные и др. Все эти сведения называются метаданными.

Необходимость получения различного рода метаданных определяется:

— распределенным характером исследуемого объекта и, как следствие, рассредоточенностью пунктов наблюдений на огромной территории, что требует знания характеристик сети наблюдений;

— многообразием наблюдательных платформ, измеряемых параметров, методов и способов их получения;

— разнообразием методов регистрации наблюдений (книги, журналы, графики на лентах самописцев, технические носители — магнитные, оптические и др.), что определяет особенности поиска данных.

Прежде чем рассматривать проблемы создания и развития баз метаданных, необходимо уточнить их место в системе обработки данных. Если в файловых системах поиск данных велся предварительно на основе метаданных, то при работе с базами данных организуется автоматический переход от метаданных к данным. Например, по характеристикам прибрежной станции (дата открытия станции и виды выполняемых наблюдений) производится выборка временных рядов по соответствующим параметрам. Когда поиск данных осуществляют пользователи, работающие в одной предметной области, то им чаще всего достаточно адреса организации, где хранятся интересующие их данные. Более сложная ситуация у лиц, работающих в разных предметных областях. Здесь нужны более полные и подробные метаданные не только о самом массиве данных, но и с каких платформ данные получены, в рамках каких проектов они созданы, с использованием каких программных средств.

Метаданные — это сведения о данных, являющиеся вспомогательными, справочными при обработке данных. Метаданные описывают свойства данных, определяющие их струк-

Т а б л и ц а 1

Место возникновения метаданных

Технологический этап	Метаданные
Производство наблюдений	Сведения о сетях и методах наблюдений, методы определения гидрохимических и загрязняющих параметров, сведения об измерительных средствах, сведения о способах и местах поверки приборов
Средства производства наблюдений	НИС, прибрежные станции, спутники, самолеты, буи
Сбор данных	Сведения о технологиях сбора данных, форматах передачи данных, описания передаваемых комплектов данных, стандарты представления и передачи данных и метаданных в Интернет
Накопление данных	Описание массивов данных, организаций-поставщиков, владельцев, пользователей, форматов сбора, хранения и обмена данными, наблюдательных проектов, рейсов НИС, изученность наблюдениями, сведения о параметрах. Единый словарь параметров, кодификаторы. Сведения о технологиях, методах контроля данных
Межведомственный и международный обмен данными	Сведения о технологиях и форматах данных, описание комплектов данных, сведения о наблюдательных проектах и программах
Моделирование	Сведения о моделях, методах, форматах выходных данных
Распространение данных	Сведения о продукции (анализы, бюллетени, ежемесячники, ежегодники, климатические справочники, прогнозы), формах выдачи (таблицы, графики, карты, текст, звук, издания), пространственно-временных масштабах представления данных, программно-технических средствах, изданиях (справочники, атласы, в том числе электронные), Web-адресах
Охрана природы	Сведения о методиках, воздействиях природной среды, чрезвычайных ситуациях, объектах, подверженных влиянию среды, опасных объектах, способах предотвращения воздействий

туру, допустимые значения и способы их представления, взаимосвязи с другими данными, размещение и другие характеристики данных, которые помогают правильно их интерпретировать и использовать.

Состав метаданных и место их возникновения представлены в табл. 1. Анализ таблицы показывает, что метаданные возникают уже в момент наблюдений, а далее они обобщаются, дополняются новыми атрибутами. В первых системах сбора данных часть метаданных включалась в основные записи на технических носителях. Например, для океанографических данных сведения о рейсе научно-исследовательских судов — НИС (страна, судно) заносились на каждую перфокарту, отражающую данные по температуре, солености и другим параметрам наблюдений для одного горизонта океанографической станции.

Основными объектами метаданных являются:

- источники данных (сведения о прибрежных станциях, судах, спутниках и др.);
- информационные ресурсы (описания массивов данных, сведения о картах, методах наблюдений и т. п.);
- информация для управления данными (форматы, программные средства, кодификаторы, модели);
- сведения о продукции (атласы, электронные справочники, прогнозы, Web-сайты).

## 1. Обзор имеющихся систем метаданных

Первые попытки создания массивов справочной информации на технических носителях относятся еще к концу шестидесятых годов [1]. К сожалению, технические возможности того времени не позволили создать полноценные справочные массивы. Была выработана только идея накопления, в первую очередь, справочных сведений о массивах данных в различных организациях страны. В конце семидесятых годов вновь вернулись к идее создания массивов справочных сведений. В России была разработана система “Каталог” [2], которая позволяла настраиваться на различные описания данных. На основе этой программной системы были разработаны две подсистемы — общие сведения о массивах гидрометеорологических данных и автоматизированный каталог океанографических данных (АКОД), в котором предлагалось отразить как сведения об источниках данных (рейсах НИС и прибрежных станциях), так и детальные сведения о данных на носителях. Эти системы отражали не весь объем справочной информации.

Более широкий подход к созданию баз метаданных был предпринят в России в начале восьмидесятых годов. Так, Институтом кибернетики Национальной академии наук Украины совместно с Всероссийским НИИ гидрометеорологической информации — Мировым центром данных (ВНИИГМИ — МЦД) была разработана система управления океанографическими данными. В состав этой системы кроме сведений о данных была включена информация о пользователях, запросах, учреждениях, программных средствах и др. То есть состав справочных данных значительно расширился.

Список справочных систем приведен в табл. 2. Более полный список систем метаданных дан в книге [3]. Анализ этой таблицы показывает, что вопросам создания автоматизированных систем каталогизации данных уделяется большое внимание во многих странах, но только небольшая часть систем доведена до практической реализации. Так, в области гидрометеорологии реализованы и функционируют такие справочные системы, как АКОД, European Directory of Marine Environmental Data (EDMED), Global Change Master Directory (GCMD), Arctic Data Directory (ADD).

Т а б л и ц а 2

Сведения об информационно-справочных системах с метаданными

Учреждение-разработчик	Система	Объекты метаданных	Источник
ВНИИГМИ — МЦД	Сведения о гидрометеорологических данных на базе локальной десятичной классификации	Общие сведения о гидрометеорологических данных	[1]
ICES	ROSCOP	Сведения о рейсах	1970, <a href="http://www.ices.dk">http://www.ices.dk</a>
ВНИИГМИ — МЦД	Сводная таблица результатов океанологических наблюдений	Сведения о рейсах, пространственно-временных координатах станций, измеренные значения параметров наблюдений	[5]
МОК ЮНЕСКО	МЕДИ	Сведения об организациях, каталог станций океанографических данных, общие сведения о массивах данных, находящихся в различных странах, детальное описание файлов данных	<a href="http://ioc.unesco.org/medi/">http://ioc.unesco.org/medi/</a>
ВНИИморгео	Система сбора, обработки и хранения морских геологических данных	Сведения об экспедициях, профилях, станциях, пробах	[6]
ВНИИГМИ — МЦД	АИСС КАТАЛОГ	Общие сведения о массивах данных	[2]
NASA	GCMD, Генеральный каталог	Общие сведения о данных по природной среде	1989, <a href="http://gcmd.nasa.gov">http://gcmd.nasa.gov</a>
ВНИИГМИ — МЦД	АКОД	Общие сведения о данных, сведения об источниках данных (каталог рейсов прибрежных станций), сведения о файлах, словарь параметров	<a href="http://data.oceaninfo.ru/cruisecat/index.jsp">http://data.oceaninfo.ru/cruisecat/index.jsp</a>
BODC	EDMED, Sea-Search Net	Сведения о массивах данных	1993, <a href="http://www.sea-search.net/">http://www.sea-search.net/</a>
BODC		Наблюдения за течениями 22 500 серий из 97 организаций 15 стран	1990, <a href="http://www.bodc.ac.uk">http://www.bodc.ac.uk</a>
ВНИИГМИ — МЦД	Информационные ресурсы	Метаданные	1999, <a href="http://www.oceaninfo.ru">http://www.oceaninfo.ru</a>
IASC	ADD	Сведения об организациях	<a href="http://www.grida.no/prog/polar/add/">http://www.grida.no/prog/polar/add/</a>
NASA, Earth Observing System Data Gateway	EOSDIS	Сведения о массивах данных	<a href="http://harp.gsfc.nasa.gov/~imswww/pub/imswelcome/plain.html">http://harp.gsfc.nasa.gov/~imswww/pub/imswelcome/plain.html</a>

Система GCMD по информационному содержанию очень похожа на системы MEDI и EDMED. Эта информационно-поисковая система предоставляет пользователям краткую обзорную информацию по космическим данным и данным о Земле, собираемым как в NASA, так и вне ее, о различных системах, а также о проектах и кампаниях, в рамках которых проводятся сбор, обработка и хранение данных.

Система EDMED разработана Британским центром океанографических данных (BODC) для стран Европейского экономического сообщества. На каждый набор данных в этой системе заполняется специальная карточка, которая включает сведения о массивах и базах данных. В 1996 году разработано программное обеспечение, которое позволяет вводить данные в каталог и осуществлять поиск в нем по различным критериям через Интернет.

**Сведения об организациях, связанных с исследованиями природной среды.** На сайте IODE МОК (<http://iode.org>) созданы страницы, на которых находится информация о мировых центрах данных, национальных центрах океанографических данных.

**Сведения о программных средствах** (<http://data.oceaninfo.ru/programs/index.jsp>). Количество компьютерных моделей настолько велико, что поиск и использование таких моделей без систематизации и описания сведений об этих моделях уже невозможны. По проекту SAMWAT создана база данных в среде системы управления базами данных (СУБД) Dbase III [4].

**Интернет-системы.** В сети Интернет огромное число Web-страниц в области исследования окружающей среды, их большой объем создает трудности для поиска и просмотра необходимых пользователю данных и информации. Многие серверы уже сейчас имеют страницы со списками адресов, составители которых считают их полезными для потенциальных пользователей. Наиболее систематизированным сервером является Web-сайт ИОС OceanPortal.org.

**Сведения о гидрометеорологических станциях** (<http://data.oceaninfo.ru/resource/objects/gms/index.jsp>) широко используются при сборе и обработке данных для получения краткой информации о станции, выборки кода и имени из словаря станций, сведений о видах наблюдений, справочных сведений о станциях (номер и название станции, разряд, количество сроков, год открытия), объемов наблюдений по станциям, карты расположения станций.

Справочными массивами общего пользования являются **словари и кодификаторы** (<http://data.oceaninfo.ru/codes/>), которые позволяют облегчить операции по группировке, сортировке и выборке справочной и исходной информации. Словари обеспечивают также единое понимание терминов. В гидрометеорологии применяются сотни кодификаторов (коды облачности, видимости, волнения, ледовых характеристик и др.), большинство из них стандартизировано на национальном и международном уровне.

## 2. Место и роль метаданных в управлении данными

Состав и точность описания характеристик объектов метаданных зависят от уровня управления данными (локальный — экспедиция, региональный — проект, национальный — страна, земной шар — международный), масштаба системы, этапа обработки данных.

Справочные сведения одного класса различаются масштабами обобщения информации. Так, наивысший масштаб обобщения требуется для международного уровня, здесь отражаются сведения о массивах данных, которые страна может передать в международ-

ный обмен или массивы международных данных. На национальном уровне необходимо иметь сведения обо всех имеющихся в стране массивах данных о состоянии среды независимо от носителя. В корпоративной системе необходимы сведения о каждом наборе данных, единице сбора, учета и обмена данными (рейс, месячный поток данных от прибрежной станции). Такие сведения должны позволять получать справки для организационного управления данными. Например, количество рейсов, станций, представленных в международный обмен каждой страной, по районам, ведомствам и др. Описание массивов данных позволяет оценить значимость тех или иных данных, выбрать наиболее подходящие для равноценного обмена с зарубежными организациями. Годовые планы экспедиций служат для контроля за их выполнением и поступлением сведений о рейсах и самих данных.

### 3. Создание и использование баз метаданных

Во ВНИИГМИ — МЦД используется большое число объектов метаданных (табл. 3). Для их создания и хранения применяется СУБД Oracle 9i. В настоящее время основным методом принят удаленный ввод метаданных в базах данных, например, ввод сведений о массивах и базах данных можно найти по адресу <http://data.oceaninfo.ru/resource/objects/dataset/backend.jsp?id=3771>, сведения о программных средствах — <http://data.oceaninfo.ru/programs/index.jsp>, сведения о Web-ресурсах — [http://data.oceaninfo.ru/resource/links/index\\_all\\_ref.jsp?ir=90\\_RINMI-WDC\\_212](http://data.oceaninfo.ru/resource/links/index_all_ref.jsp?ir=90_RINMI-WDC_212), сведения об информационных ресурсах — <http://data.oceaninfo.ru/resource/addprod/index.jsp>.

Для поиска метаданных разработаны соответствующие приложения для каждого объекта. На странице <http://data.oceaninfo.ru/resource/index.jsp> через раздел “Метаданные” можно запустить приложения по поиску экземпляров объектов метаданных.

Сведения о массивах данных нужны не тем, кто занимается созданием и использованием баз данных. Об их существовании и характеристиках знает большинство разработчи-

Т а б л и ц а 3

Сведения о базах метаданных

Название	Количество	
	документов	атрибутов
Сведения о массивах и базах данных, имеющихся в различных организациях России	470	35
Описания форматов файлов данных	15	20
Сведения об организациях	1 300	12
Сведения о НИС за весь период сбора данных	10 000	13
Сведения о действующих НИС	122	5
Сведения о прибрежных станциях	750	65
Сведения о спутниках	25	10
Сведения о сетях наблюдений	10	17
Сведения об экспертах	170	12
Сведения о наблюдательных проектах	150	15
Сведения о рейсах НИС	32 500	21
Сведения о моделях и других программных средствах	10	13
Сведения об океанографических приборах	150	17
Сведения о морских картах	100	15

ков. Эта информация нужна тем, кто работает в смежных науках и не знает, что имеется в интересующих их областях. Поэтому системы сбора метаданных должны быть на первом этапе централизованы, как это было в шестидесятых-семидесятых годах с исходными данными. Чем полнее метаданные, тем они эффективнее могут быть использованы для поиска сведений о данных.

Среднее время жизненного цикла программной системы составляет 3–5 лет, а для создания некоторых баз данных в области наук о Земле с полнотой до 95 % требуется иногда до десяти лет и более. Например, в России для создания базы метаданных “Сведения о рейсах НИС” в объеме 33 тыс. рейсов потребовалось 10 лет (1981–1990 годы). За это время сменились две операционные системы, поколение ЭВМ. Главное — не разработка программных средств поиска метаданных, а хорошо проработанные политика, технология и организация сбора метаданных для управления данными на уровне центров данных, проектов.

Рост объемов метаданных увеличивает их полезность. Чтобы ускорить создание баз метаданных, необходимо обязать исполнителей каждого проекта, программы, экспедиции, связанных с получением данных, заполнять описания массивов данных.

Главной характеристикой системы с метаданными является полнота описаний по центрам, видам наблюдений, т. е. нужно создавать метаданные, касающиеся всей природной среды — метеорологии, океанографии, геологии, гидрологии рек и др. Число атрибутов для первого уровня описаний может быть минимальным. Главное — как можно шире представить информационные ресурсы. За счет этого можно существенно сократить время ввода описания одного массива данных.

## 4. Методы и примеры консолидации метаданных

В настоящее время в мире в основном создаются отдельные системы для различных объектов метаданных (сведений о массивах и базах данных, рейсах НИС, прибрежных станциях и др.). Наибольшее число систем создано для таких объектов, как сведения о массивах и базах данных, рейсах НИС. Наступил момент, когда необходимо не только стандартизировать состав атрибутов объектов метаданных, но и создать единую систему метаданных, включающую все объекты (сведения о массивах данных, организациях, экспертах, форматах, проектах и др.). Одной из задач развития метаданных может быть объединение их в одну структуру (систему), например объединение баз данных таких систем, как EDMED, GCMD, ADD и др., как это делается с исходными данными в мировых центрах данных. Это процесс сложный, и реализовать его в ближайшие годы очень трудно по организационным и техническим причинам.

Перспективой развития метаданных является их консолидация. Например, это может быть объединение описаний массивов данных различных стран, организаций, систем с подключением дополнительной информации — форматов, библиографии по массивам данных, карт изученности, таблиц содержания и др., как это сделано в системе EOSDIS (NASA USA). Объединение можно также произвести путем создания общего списка массивов и баз данных и соответствующих ссылок на их описания, находящиеся на сайтах организаций—владельцев этих данных. Такая схема организации метаданных реализована в России (<http://www.oceaninfo.ru/data/russ.htm>).

Другим методом консолидации метаданных можно рекомендовать подключение к каждому объекту метаданных ссылок на аналогичные системы, работающие в других орга-

низациях и странах. Например, при поиске сведений о массивах и базах данных на портале ЕСИМО (<http://data.oceaninfo.ru/resource/objects/dataset/index.jsp>) выдается дополнительный список адресов систем, где можно осуществить поиск метаданных. При этом в некоторые системы можно организовать автоматизированную передачу поисковых критериев пользователя.

Еще одним методом консолидации метаданных на уровне разработки приложений можно предложить создание java-классов для визуализации каждого объекта метаданных. Таким образом, при работе с любым объектом метаданных можно получить дополнительную информацию из других объектов, например, сведения об организациях, экспертах подключаются к таким объектам, как сведения о массивах и базах данных, информационных ресурсах, сведения о рейсах НИС и др.

Наиболее перспективным направлением консолидации метаданных является создание единой схемы метаданных и ее поддержка различными организациями. Каждая организация может внести новое описание в любой объект метаданных. Такая консолидация возможна только после согласования структур описания для всех объектов метаданных.

## 5. Характеристики информационных ресурсов как результат обработки баз метаданных

Кроме получения справочных сведений в том виде, в каком они были введены в ЭВМ, на основе метаданных можно выполнять аналитические запросы и получать агрегированные характеристики, т. е. проводить анализ поступления данных от различных организаций. Например, можно получить количество массивов данных по организациям, регионам; количество рейсов НИС, поступивших в 1990–2000 годы от различных стран, организаций России; количество рейсов НИС, поступивших в 1990–2000 годы по видам наблюдений; количество станций по квадратам, периодам, параметрам и т. п.

Для эффективного управления данными необходимо знать состояние информационных ресурсов и продукции, которое можно отразить: состоянием сетей наблюдений, с которых получается информация; оценкой информационных потоков; распределением информации по носителям; обобщенными характеристиками баз данных; количеством выполненных запросов с указанием тенденции в информационных потребностях пользователей и др.

Характеристиками потоков информации являются количество источников данных (например, рейсов), объем (число наблюдений), прирост объемов данных в среднем за год. Характеристиками потоков информации в системе переработки океанографических данных являются:

- на этапе производства наблюдений — количество источников информации (НИС, прибрежных станций, буев и др.), объемы получаемой информации от одного источника (срочной, ежедневной, ежемесячной, годовой);
- на уровне объединения информации во времени — объемы поступающей информации на обработку в центр (ежедневной, ежемесячной, годовой) в байтах;
- на уровне объединения информации в пространстве — станция, рейс, трасса, территория, бассейн, регион;
- на этапе обработки данных — объем обрабатываемой информации, время обработки информации;
- на этапе выдачи информации — объем выходной информации, периодичность представления (срок, сутки, неделя, декада, месяц, год), пространственное объединение дан-



ных (регион, бассейн), сортировка данных (во времени, в пространстве, в пространстве и во времени).

Сведения об используемых приборах, измерительных системах являются важным показателем качества сети наблюдений, поэтому получение сведений об оснащенности сети различными типами приборов — важная функция системы обработки баз метаданных.

Распределение данных океанографических наблюдений по основным регионам Мирового океана с указанием организаций—поставщиков данных отражает изученность наблюдениями того или иного района, поэтому такая таблица полезна для потенциального пользователя. Распределение информации по носителям показывает готовность данных к обработке.

## Заключение

Определен состав метаданных, необходимых для обслуживания экономики страны информацией о состоянии морской среды. Проведено выделение объектов метаданных, что позволяет более четко разделить ответственность за создание такой информации и планомерно создавать их в зависимости от важности и необходимости. Разработана структура базы метаданных, метаданные включены в несколько разрабатываемых систем (интегрированную базу данных “Океанография”, систему “Информационные ресурсы России по морской среде”, портал ЕСИМО, <http://www.oceaninfo.ru>).

## Список литературы

- [1] Клюкин Н.К. и др. Система классификации гидрометеорологической информации для хранения и поиска, в том числе автоматизированного // Обнинск: ВНИИГМИ — МЦД. Экспресс-информация. 1969. Вып. 5. 36 с.
- [2] ВЕСЕЛОВ В.М. и др. Автоматизированная информационно-справочная система КАТАЛОГ // Сб. науч. тр. / Росгидромет. ВНИИГМИ — МЦД. 1978. Вып. 43. С. 91–104.
- [3] Вязилов Е.Д. Информационные ресурсы по окружающей среде. М.: УРПС, 2001. 312 с.
- [4] The SAMWAT Database for Computer Models in Water Management. The Hague. SAMWAT. 1988. Report N. 2. 55 p.
- [5] ИНСТРУКЦИЯ по заполнению сводной таблицы результатов океанографических наблюдений. Обнинск: ВНИИГМИ — МЦД. 1971. 59 с.
- [6] ВРЕМЕННАЯ ИНСТРУКЦИЯ о порядке и форме представления организациями МИНГЕО СССР справочных сведений о морских геолого-геофизических работах и результатах морских геологических исследований в отраслевой центр геологических данных. М.: ВНИИМОРГЕО, 1977. 38 с.

*Поступила в редакцию 2 июня 2005 г.*