

## RDF СХЕМА ДЛЯ МЕТАДААННЫХ ПО МЕТЕОРОЛОГИИ И КЛИМАТУ

Титов А.Г.

*Сибирский центр климато-экологических исследований и образования и Институт мониторинга климатических и экологических систем СО РАН,  
634055, г. Томск, пр. Академический, 10/3, [titov@scert.ru](mailto:titov@scert.ru).*

*В рамках программы Интегрированные Региональные Исследования Сибири (ИРИС) планируется создать информационно-вычислительную инфраструктуру для поддержки мультидисциплинарных исследований окружающей среды данного региона. Одним из важных шагов для достижения этой цели является разработка современного стандарта метаданных для создания пригодных к автоматической обработке описаний наборов данных, содержащих наблюдения различных параметров атмосферы. Для реализации модели описания метаданных была выбрана технология RDF [1]. RDF схема создана на базе существующих широко используемых стандартов, таких как Dublin Core [2], Ecological Metadata Language [3] и др. Она может быть использована в качестве универсального инструмента для описания данных по окружающей среде с целью последующей организации доступа к ним через Интернет.*

### Введение

Одна из важных особенностей наук об окружающей среде заключается в том, что они оперируют огромными массивами разнородных данных. Хотя при выполнении фундаментальных исследований ученые часто считают конечным результатом публикацию в научной литературе, для науки все большее значение имеют наборы данных, и, соответственно, управление этими данными. Данные для исследователя действительно являются исходным материалом для публикации, но при этом они также являются самоценным ресурсом [4]. При соответствующей обработке, хранении они в будущем могут быть повторно использованы другими научными исследователями, а также для коммерческих и образовательных целей.

В рассматриваемом случае под наборами данных понимаются:

1. Цифровые массивы, содержащие измерения параметров окружающей среды, и, в частности, атмосферы. Они остаются после окончания научных проектов, а также находятся в распоряжении отдельных ученых как часть их исследований.
2. Данные, полученные в результате работы компьютерных прогнозирующих моделей, которые можно считать “данными об окружающей среде”, хотя они и не являются результатом измерений.

Для адекватного восприятия данных человеком, а также осмысленной машинной обработки они должны сопровождаться метаданными, описывающими эти данные. На основе метаданных осуществляется поиск ресурсов и вывод результатов поиска. Для разработки модели описания метаданных предлагается использовать средства технологии RDF, [1], которая является стандартом для определения и обработки метаданных Web-ресурсов.

Технология RDF (Resource Description Framework), основана на простой, но устойчивой модели данных, в которой ресурсы описываются через их свойства (атрибуты). Свойства могут принимать как атомарные значения, такие как строки текста и числа, так и представлять собой другие ресурсы, имеющие собственные свойства. Эта модель данных часто визуально представляется как ориентированный помеченный граф. Простой пример подобного описания приводится ниже [5]:

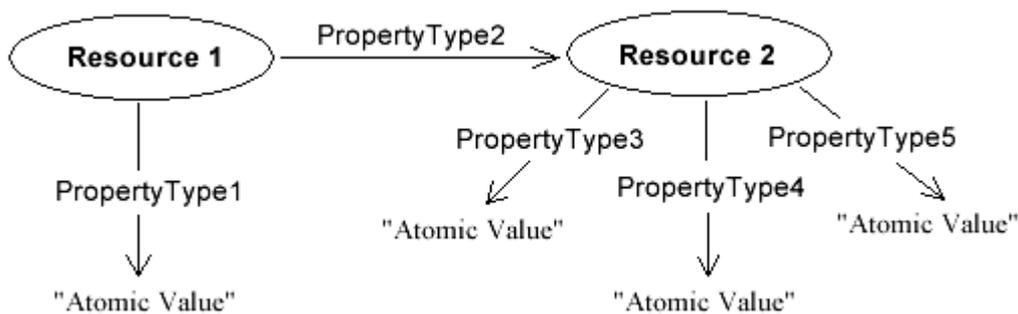


Рис. 1. Пример RDF-описания [5]

На Рис. 1 можно видеть, что Resource 1 обладает свойством типа PropertyType1, которое принимает атомарные значения. Также у него имеется свойство Resource 2 типа PropertyType2, который в свою очередь является ресурсом со своими собственными свойствами. Используя эту легко расширяемую и устойчивую логическую модель, можно создать структурированные метаданные для описания сетевых ресурсов.

Для обеспечения механизмов описания групп связанных ресурсов и отношений между ресурсами, используется язык описания словарей RDF, RDF Schema [6]. Язык RDF Schema определяет специализированные классы и свойства, которые могут быть использованы для описания классов ресурсов, свойств и других ресурсов. Система классов и свойств языка RDF Schema подобна системе типов таких объектно-ориентированных языков, как Java.

Данная работа является частью более обширной задачи по построению рабочей модели распределенной программной системы для эффективной обработки, хранения, поиска и организации доступа к наборам данных, содержащих результаты наблюдений и математического моделирования климата, произведенные различными моделями, а также обнаружения основных признаков изменений климата на региональном уровне [4, 7].

Кроме данных метеонаблюдений планируется использование данных реанализа NCEP/NCAR [8], и ECMWF [9], а также другие результаты моделирования климатических и атмосферных процессов, полученные как в России, так и за рубежом. В настоящее время выполняется разработка инструментария для предварительной обработки собранных массивов данных и их визуализации как части интегрированной системы. Реализация системы предполагается в виде специализированного веб-сайта, созданного на базе Java Servlet технологии, и обеспечивающего вышеупомянутую функциональность.

Разработка современного стандарта метаданных в виде RDF схемы является одним из шагов к решению поставленной задачи по созданию пригодных к автоматической обработке описаний наборов данных, содержащих наблюдения различных атмосферных параметров.

### Обзор существующих стандартов метаданных

Начиная разработку стандарта, следует иметь в виду, что предварительно необходимо четко определить, какого рода информацию должны представлять собой метаданные. Данная задача не является тривиальной, поскольку кроме стандартной информации о персонале, географическом местоположении мест наблюдения, и т.д., необходимо также представлять информацию о всевозможных параметрах окружающей среды, временном и пространственном масштабе данных, их формате, способах обработки, а также о приборах, используемых для измерения.

В настоящее время уже существует ряд широко используемых стандартов метаданных, в том числе для описания Web-ресурсов и наборов данных. Примерами могут служить следующие стандарты:

1. Набор Элементов Метаданных Дублинского Ядра (Dublin Core Metadata Set) [2], который используется для каталогизации электронных изданий.
2. Формат Обмена Директориями (DIF) [10], который является стандартом де-факто для создания директорий с записями в формате ASCII, описывающими наборы данных. DIF используется для обмена научными данными, включая данные по окружающей среде.
3. Ecological Metadata Language (EML) [3], спецификация метаданных, разработанная специально для экологии.
4. Content Standard for Digital Geospatial Metadata [11], стандарт Федерального Комитета по Географическим Данным, США

Перечисленные выше XML-стандарты с использованием XML-спецификации, разработанной в рамках сети NERIN [12] были задействованы в работе по созданию собственной RDF схемы, предназначенной для описания наборов данных относящихся к метеорологии, климату и переносу атмосферного загрязнения.

### Полученные результаты

В настоящее время разработана первая версия RDF схемы для описания данных по окружающей среде, фрагмент графического представления которой можно видеть на Рис. 2:

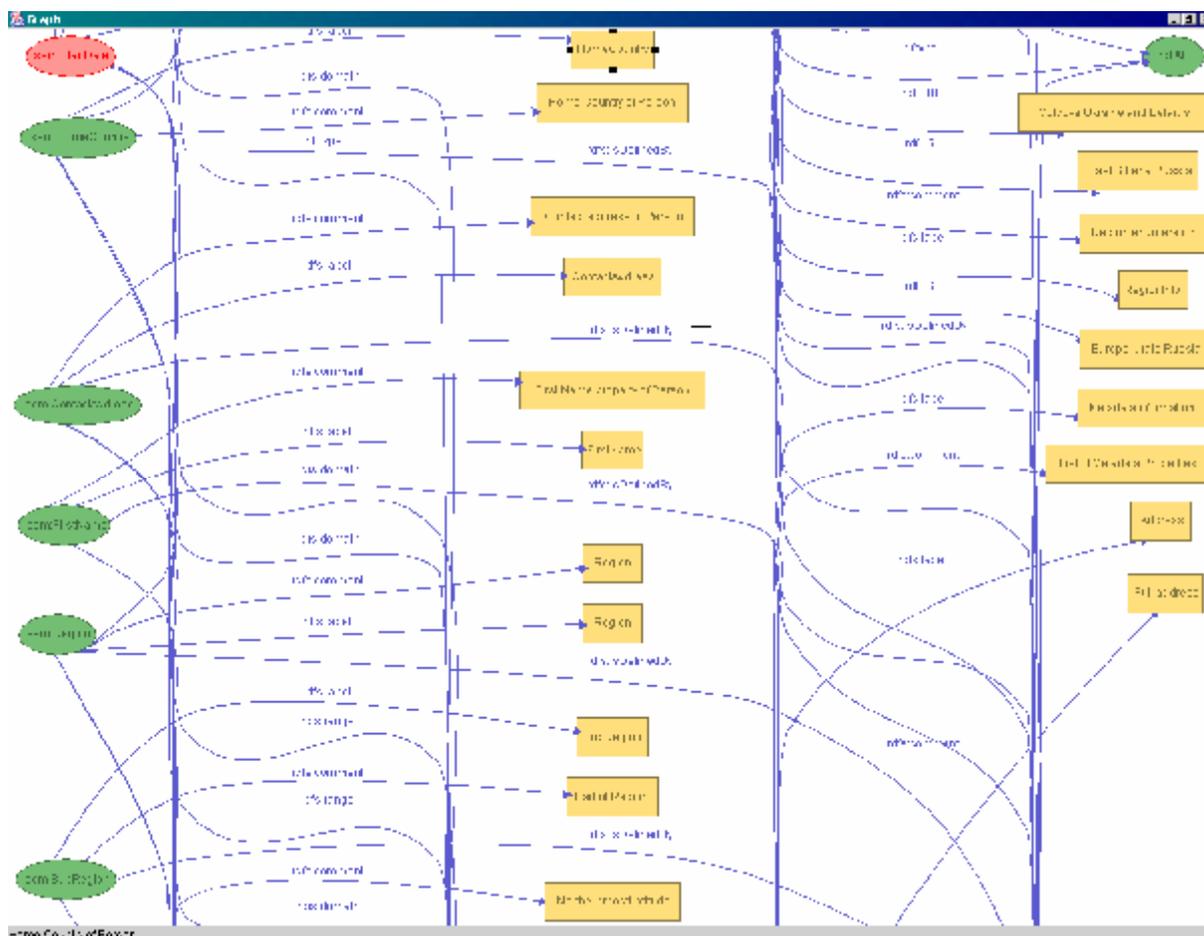


Рис. 2. Фрагмент разработанной RDF-схемы

Созданы примеры RDF документов на базе существующих метаданных из базы NERIN [12], иллюстрирующие применение разработанного стандарта. Фрагмент RDF документа, описывающего набор данных по переносу углекислого газа и энергии в Казахстане, приведен на Рис. 3.

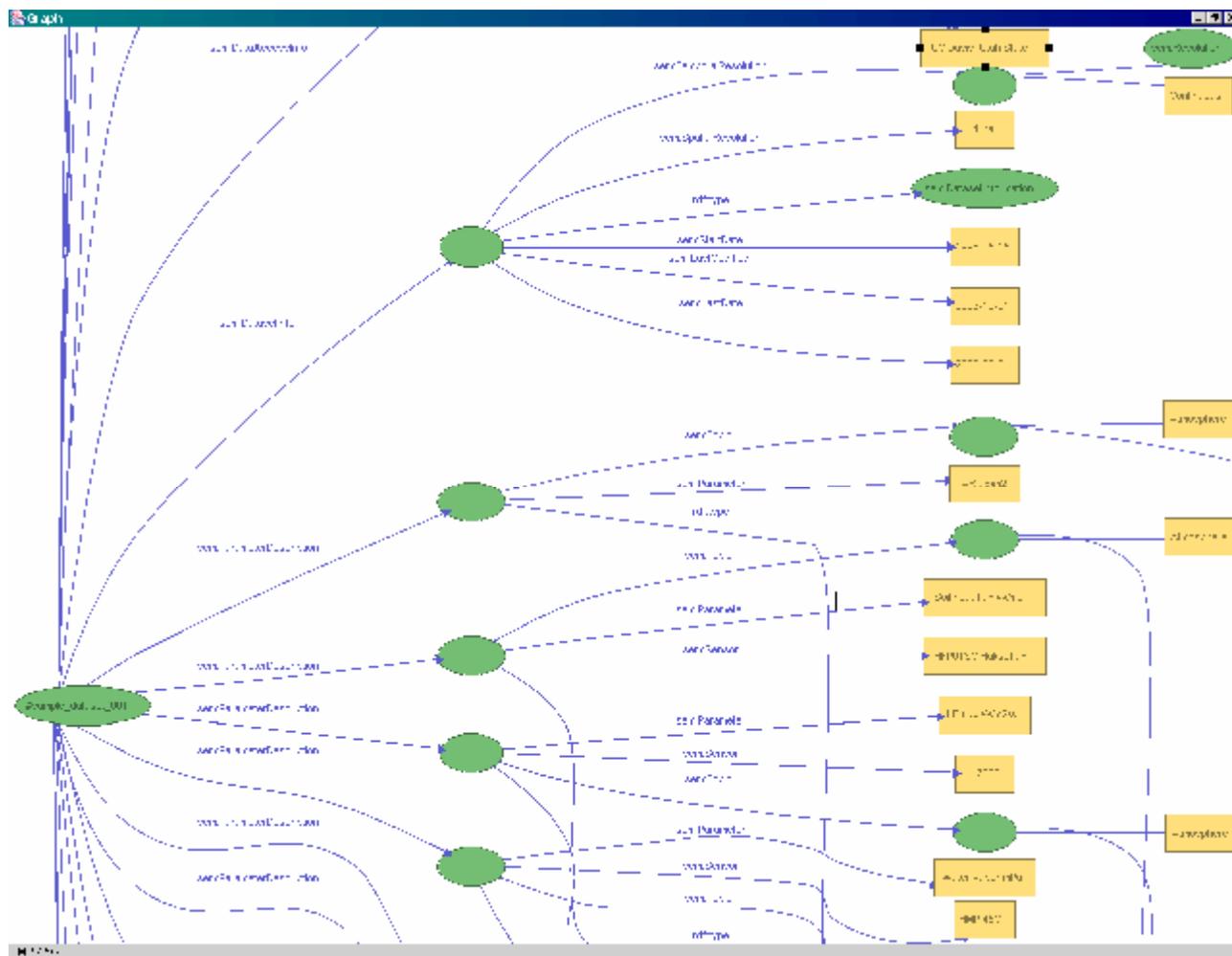


Рис. 3. Пример RDF документа, соответствующего RDF схеме

На Рис. 3 можно видеть ресурс `sample_dataset_001`, соответствующий описываемому набору данных, а также ряд его свойств (`DatasetInfo`, `ParameterDescription`), которые в свою очередь являются ресурсами. Например, некоторые свойства ресурса типа `ParameterDescription` принимают атомарные значения.

Полученные таким образом RDF документы являются синтаксически корректными, пригодными для обработки стандартными программными инструментами, такими как Jena [13]. Одним из тестов валидации является их занесение в RDF базу данных Sesame, [14], с целью последующей реализации на этой платформе семантического поиска. На Рис. 4 приведен пример веб-интерфейса Sesame после занесения полученной RDF схемы.

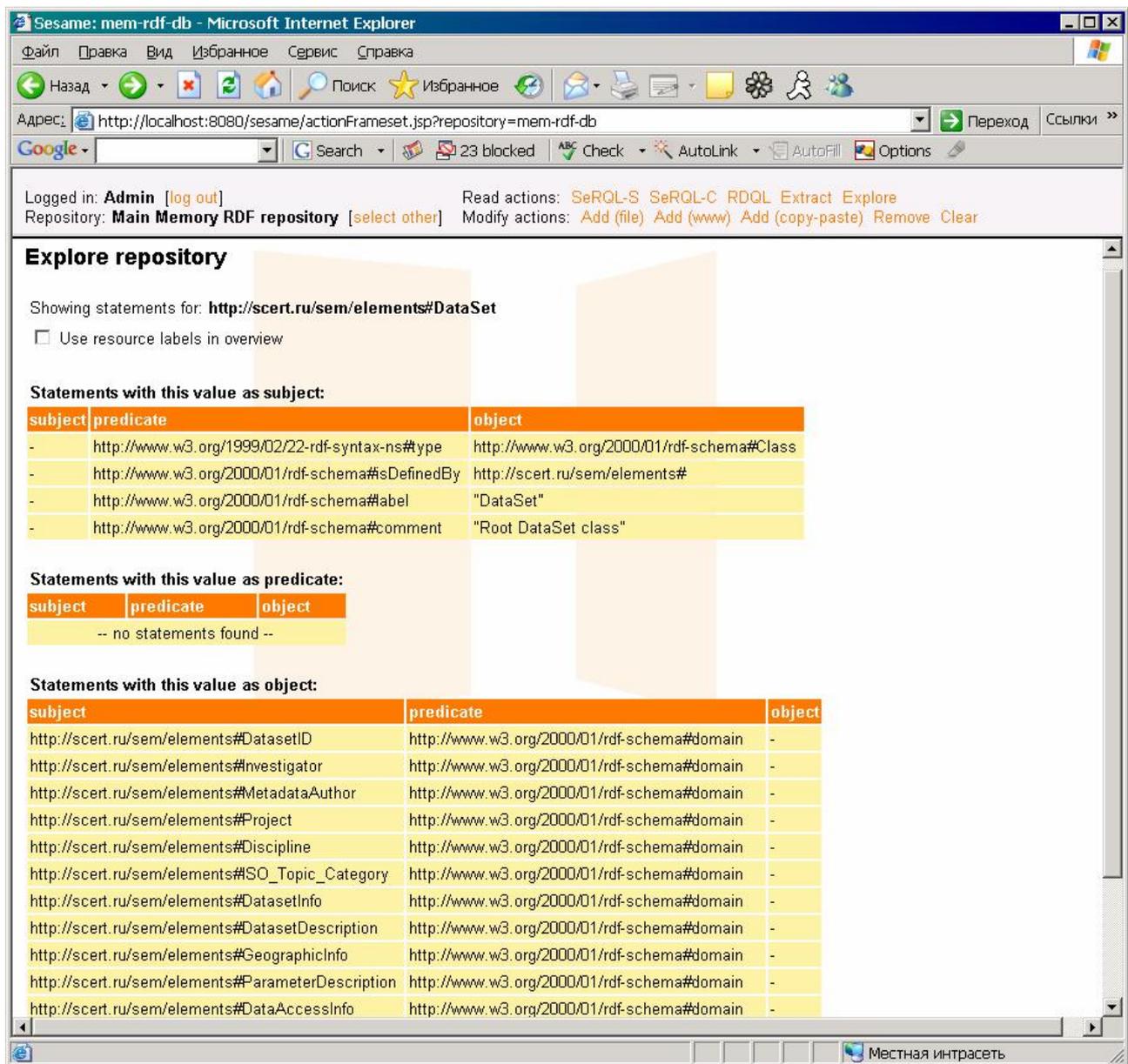


Рис. 4. Веб-интерфейс RDF-репозитория Sesame

На Рис. 5 приводится пример структуры базы данных для хранения RDF-высказываний:

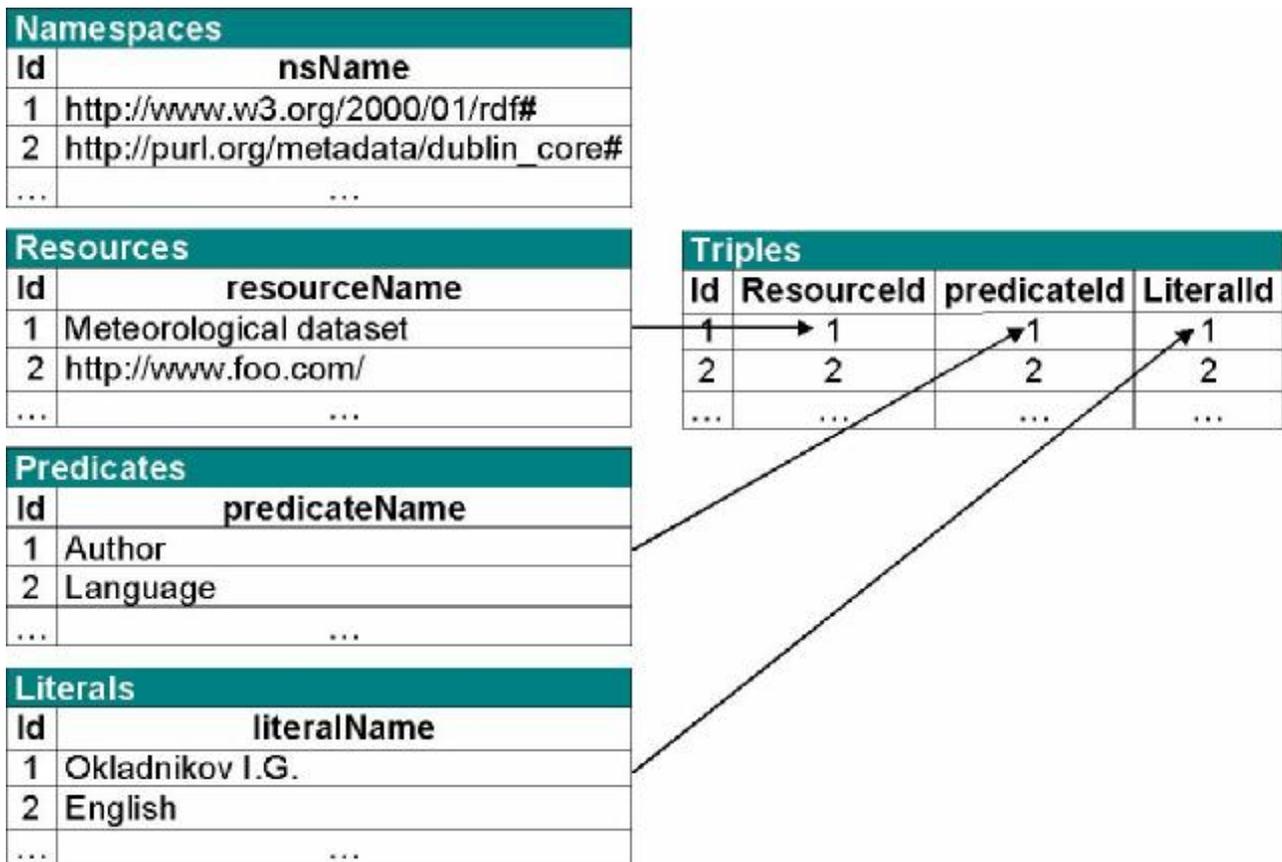


Рис. 5. Структура базы данных RDF

### Реализация распределенной программной системы

Как уже отмечалось выше, данная работа является частью задачи по построению специализированного веб-сайта для работы с данными по метеорологии и климату. Планируемая структура разрабатываемой модели системы представлена на Рис. 6.

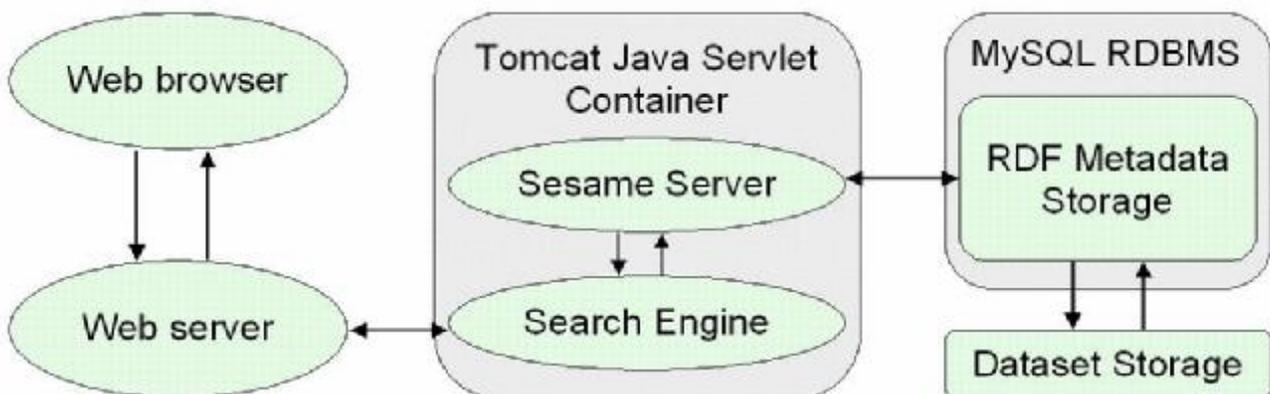


Рис. 6. Общая структура модели распределенной системы

Ядром системы является веб-сервер Apache [15], в связке с Java-сервлет контейнером Tomcat [16]. Сервер RDF-репозитория Sesame функционирует в рамках среды Tomcat и обеспечивает необходимый инструментарий для анализа, интерпретации, создания запросов и хранения RDF-документов. Поскольку Sesame может использовать внешнюю СУБД для хранения

RDF-данных, для реализации хранилища была выбрана хорошо зарекомендовавшая себя СУБД MySQL [17]. Следует отметить, что RDF-репозиторий связан с распределенным хранилищем наборов данных. С другой стороны, сервер Sesame обменивается информацией с Java-приложением Search Engine, осуществляющим алгоритм семантического поиска и представление полученных результатов.

## Заключение

Предложенная модель метаданных в виде RDF-схемы для описания наборов данных по метеорологии и климату позволяет использовать все преимущества технологии Semantic Web [18]. В частности, она может быть использована в качестве универсального инструмента для описания данных по окружающей среде с целью последующей организации доступа к ним через Интернет, а также в процессе разработки информационно-вычислительной инфраструктуры в рамках Интегрированных Региональных Исследований Сибири (ИРИС) для поддержки мультидисциплинарных исследований окружающей среды данного региона.

## Благодарности

Данная работа частично поддержана ИП №34 СО РАН, “Разработка распределенной информационно-вычислительной среды для исследования экологических систем”, грантом РФФИ №05-05-98010, а также проектами Enviro-RISKS (INCO-CT-2004-013427) и Enviromis-2 (031303).

Автор выражает благодарность профессору Гордову Е.П. и к.ф.-м.н. Фазлиеву А.З. за постановку задачи, моральную поддержку и ценные советы по созданию RDF схемы метаданных и соответствующих RDF описаний.

## Литература

1. Resource Description Framework (RDF), <http://www.w3.org/RDF/>
2. Dublin Core Metadata Element Set, <http://dublincore.org/>
3. Ecological Metadata Language (EML), <http://knb.ecoinformatics.org/software/eml/>
4. Гордов Е.П. Современные тенденции в региональных исследованиях окружающей среды. //География и природные ресурсы. – Труды международной конференции по измерениям, моделированию и информационным системам для изучения окружающей среды: ENVIROMIS-2004, Томск, 16-25 июля 2004 г. – С. 11.
5. Tony Gill. Metadata and the World Wide Web. - Murtha Baca (ed), *Introduction to Metadata: Pathways to Digital Information*. S.l.: Getty Information Institute, 1998, [http://www.getty.edu/research/conducting\\_research/standards/intrometadata/metadata.html](http://www.getty.edu/research/conducting_research/standards/intrometadata/metadata.html)
6. RDF Vocabulary Description Language 1.0: RDF Schema, <http://www.w3.org/TR/rdf-schema/>
7. Gordov E.P., Begni G. Siberia integrated regional study development // *Computational Technologies*, – 2005. – Vol. 10, Part 2, – P. 149-154.
8. NCEP/NCAR Reanalysis Project, <http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml>
9. European Centre for Medium-Range Weather Forecasts, <http://www.ecmwf.int/>
10. Directory Interchange Format (DIF), <http://gcmd.gsfc.nasa.gov/User/difguide/difman.html>
11. Content Standard for Digital Geospatial Metadata, [http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2\\_0698.pdf](http://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf)
12. <http://nerin.scert.ru/>
13. Jena – A Semantic Web Framework for Java, <http://jena.sourceforge.net/>

14. Jeen Broekstra, Arjohn Kampman, Frank van Harmelen Sesame: An Architecture for Storing and Querying RDF Data and Schema Information (2001) // In: Semantics for the WWW, MIT Press, 2001, D. Fensel, J. Hendler, H. Lieberman and W. Wahlster (eds).
15. Apache HTTP Server, <http://httpd.apache.org/>
16. Apache Tomcat, <http://tomcat.apache.org/>
17. MySQL, <http://www.mysql.org/>
18. Semantic Web, <http://www.w3.org/2001/sw/>

## **RDF SCHEMA FOR METADATA DESCRIBING METEOROLOGY AND CLIMATE DATASETS**

Titov A.G.

*Siberian Center for Environmental Research and Training and Institute of Monitoring of Climatic and Ecological Systems SB RAS,  
634055, Tomsk, Akademicheskii Ave., 10/3, [titov@scert.ru](mailto:titov@scert.ru).*

*Within the framework of Siberian Integrated Regional Study it is planned to create an information-computational infrastructure to support multidisciplinary environmental investigations of this region. One of the important steps to achieve this goal is the development of the modern metadata standard for machine-readable description of datasets containing observations of various atmospheric parameters. The Resource Description Framework (RDF) technology means have been chosen for metadata description model realization [1]. The RDF Schema is implemented on the base of the existing widely used standards, such as Dublin Core Metadata Element Set [2], Ecological Metadata Language [3] and others. It can be used as a universal tool for describing environmental data and making it available for researchers over the Internet.*