

Предложения по наборам метаданных для научных информационных ресурсов ЕНИИ РАН

Бездушный А.Н.
Кулагин М.В.

Жижченко А.Б.
Серебряков В.А.

Калёнов Н.Е.
Бездушный А.А.

ВЦ РАН, ЦНТК РАН, БЕН
bezdushn@ccas.ru, abz@ipsunras.ru, nek@benran.ru,
ql@ccas.ru, serebr@ccas.ru, alix@7ka.mipt.ru

Аннотация

Работа посвящена вопросам формирования наборов элементов метаданных и онтологий для научных информационных ресурсов РАН в рамках проекта Единого Научного Информационного Пространства (ЕНИИ) РАН. Рассматриваются потребности, цели и задачи организации ЕНИИ РАН, как среды взаимосвязанных распределённых гетерогенных систем. Дается представление о предметных областях и типах ресурсов, информацию о которых планируется представлять в ЕНИИ. Описывается методика, используемая для описания схем метаданных, приводится список проанализированных стандартов и предложений по схемам метаданных, использованных при разработке схем ЕНИИ.

1 Цели и задачи ЕНИИ

Российская Академия Наук имеет разветвленную структуру, которая объединяет большое число научно-исследовательских учреждений и коллективов, расположенных на всей территории России и вовлеченных во всё многообразие видов научной деятельности. Эти учреждения обладают уникальными научными информационными ресурсами. Среди них – опубликованные результаты научных исследований и экспериментов, библиографические и фактографические базы данных, сведения об ученых, их научной деятельности, публикациях, проектах и т.п. Эти ресурсы представляют значительный интерес для сотрудников РАН, членов мирового научного сообщества, для представителей промышленности и предпринимателей, которые заинтересованы во внедрении результатов научных исследований.

Однако в настоящий момент значительная часть информационных ресурсов РАН недоступна широкому кругу научной общественности, а ресурсы, представленные в Интернет, существенно разрознены, недостаточно систематизированы и структурированы. При создании их описаний

недостаточное внимание уделяется вопросам интероперабельности – слабо применяются соглашения по стандартизации электронного представления информационных ресурсов и соответствующие средства, призванные поддержать интеграцию информационных ресурсов, повышение точности поиска и т.п. Результатом этого является невозможность для пользователя получить полную и достоверную информацию о ресурсах, представляющих для него интерес. Очевидно, что каждая область науки, оперируя со своими специфическими данными, имеет потребности в собственных форматах их представления, обусловленных требованиями функциональности соответствующих систем обработки информации. Этим объясняется малая степень интеграции таких систем (например, по сравнению с системами обработки бизнес-данных). Тем не менее, необходимость обеспечения активных научных коммуникаций и эффективного использования научной информации делает актуальной задачу интеграции разнородных научных данных. В качестве первого шага необходимо обеспечить такую интеграцию на некотором «верхнем уровне», общем для всех отраслей фундаментальной науки. На решение этой задачи направлена разрабатываемая силами специалистов ВЦ РАН и ЦНТК РАН система [1,2], имеющая целью поддержку формирования Единого Научного Информационного Пространства РАН, интегрированного источника научной информации.

Система предусматривает объединение сведений о разнородных научных информационных ресурсах РАН, обеспечение актуальности этих сведений и широких возможностей для достаточно точного поиска научных ресурсов на основе этих сведений, поддержку средств научной коммуникации, сервисов, связанных с возможностью оперативного информирования пользователей о необходимых им ресурсах и т.п.

Такая система обеспечит пользователей актуальными данными о текущем состоянии и характеристиках информационно-научной базы институтов РАН и их подразделений, упростит анализ состояния и тенденций развития науки.

Облегченный доступ к информации изменит способы ведения научной деятельности, способы обучения.

В сложившейся ситуации, когда сведения о ресурсах представлены в виде слабоструктурированного текста, когда поисковые системы осуществляют полнотекстовый поиск нужных данных по запросам в свободной форме, пользователь получает огромное количество «шумовой» информации, среди которой очень трудно выбрать действительно полезные знания. Учитывая это обстоятельство, для представления сведений о ресурсах стали использовать структурное представление, выделять понятие метаданных, описывающих содержимое ресурса в виде набора именованных значений, в том числе связей с другими ресурсами. Метаданные используются для автоматизированного анализа содержимого ресурса, построения поисковых индексов и позволяют обеспечить достаточно высокую точность и эффективность поиска разнородной информации.

Очевидно, что глубина структуризации метаданных является одной из важнейших характеристик любой информационной системы. Слишком большая глубина усложняет процессы подготовки метаданных, слишком малая – ограничивает возможности системы. Глубина структуризации метаданных во многом определяется задачами конкретной системы. В узкопрофессиональных системах используют достаточно большую степень детализации и структуризации данных, что обусловлено необходимостью проведения специальных исследований и поддержки соответствующих процессов обработки информации. Однако во многих случаях при интеграции разнородных информационных ресурсов высокая степень структуризации не требуется, а также усложняет процессы формирования, поиска и представления данных. В связи с этим одной из первоочередных задач при разработке интегрированной системы представления научных данных является определение минимальной глубины структуризации метаданных о ресурсах. Предполагая постепенное наращивание функциональности системы, необходимо выработать методику развития схем метаданных - увеличения глубины описания той или иной предметной области, детализации представления информации.

Для обеспечения взаимодействия существующих разнородных научных систем на информационном уровне необходимо выработать корпоративные стандарты на интерфейсы взаимодействия, а также профили метаданных, что позволило бы реализовать инструментальные средства, обеспечивающие интеграцию данных в единую среду. Результатом решения этих первоочередных проблем должны явиться предложения ЕНИП по:

- типовым интерфейсам взаимодействия (форматы данных, протоколы обмена)

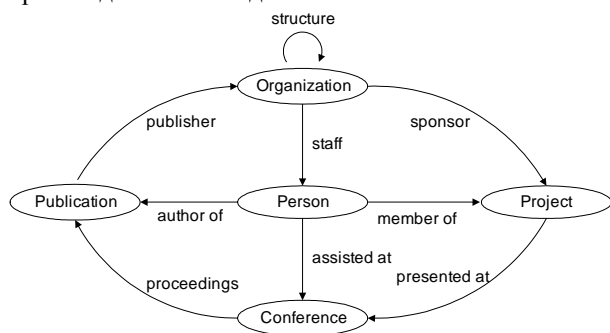
отдельных информационных источников (организаций РАН, поддерживающих собственные научные информационные ресурсы);

- профилям метаинформации, предоставляемой этими источниками. В частности, производится разработка набора элементов метаданных для научной информации общего характера, предложений по формированию элементов метаданных для отдельных областей науки и согласование их с научным сообществом и международными открытыми стандартами;
- справочникам и классификаторам ресурсов;
- реализации политики информационной безопасности и требований по разграничению прав доступа к цифровым ресурсам.

2 Информационное наполнение ЕНИП

Рассмотрим информационное наполнение Единого Научного Информационного Пространства на начальном этапе поддержки ЕНИП. Естественно, что научные учреждения заинтересованы в предоставлении доступа к данным о научных достижениях, научной деятельности сотрудников, административной информации об организации. Эта информация представляет интерес и для конечных пользователей системы, осуществляющих поиск и навигацию по информационному пространству РАН, позволяет сотрудникам получить информацию о смежных со своими работами в других коллективах.

Естественно, что основным «информационным ресурсом» науки являются люди, которые эту науку делают. Поэтому остальные ресурсы являются производными от их деятельности.



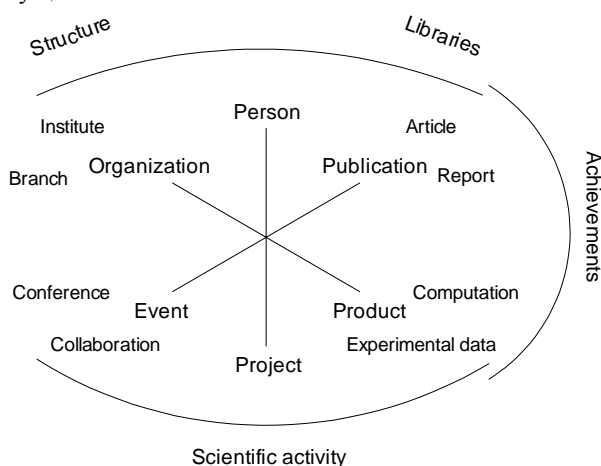
Следует выделить базовый набор профилей метаданных информационных ресурсов РАН, состоящий из нескольких основных типов информационных ресурсов, которые разбиваются на множество взаимосвязанных подтипов. Это, в первую очередь, поддержка следующих видов ресурсов:

- «Участники научной деятельности» - центральное звено, вся информация в РАН связана с научной деятельностью её сотрудников, «Персон», образующих разнообразные организационные объединения от формальных («Организации» и «Подразделения») до неформальных

(«Коллективы», «Сообщества» «Рабочие группы»). Примерами их уточнения могут служить «Читатели» библиотек, «Разработчики», «Эксперты» и «Инвесторы» инновационных систем и порталов и т.п.

- «Документы и публикации» - ресурсы этого типа представляют собой научные труды, статьи, отчёты сотрудников (научные «Публикации» и «Диссертации» сотрудников РАН, их «Коллекции»), возможно, административные «Постановления» и «Распоряжения». Примерами специализации публикации могут служить, например, «Тезисы конференций» и т.п.
- «Научная деятельность», в частности, «Проекты», отражающие процесс научной деятельности, информация о результатах проектов, патентах и т.п.
- «Научные события» - представляющие как разовые, так и повторяющиеся научные мероприятия, такие как «Конференции», «Семинары», «Симпозиумы».
- «Результаты научной деятельности», в которые могут входить
 - «Интернет системы» - Web-сайты и пр.,
 - «Базы данных», предоставляющие автономные коллекции информации с той или иной степенью интеграции с ЕНИП, и т.п.
 - «Экспериментальные данные» и их «Математические модели»,
 - «Программные системы», в частности, «Научные вычислительные приложения»,
 - «Экспериментальные установки», «Изобретения», «Технологии», и т.п.

Следует отметить, что на этом уровне тематическая специализация отраслей науки не существенна.



Приведённая диаграмма показывает не только упомянутые типы информационных ресурсов, но и направления специализации под конкретные прикладные области и приложения. Библиотечные системы имеют в основном дело с изданиями, статьями, авторами публикаций, они расширяют базовую атрибутику этих типов ресурсов, добавляя термины, специфичные для библиотечных

специалистов. Кроме изданий как таковых, выделяются экземпляры изданий, которые могут получить на руки читатели, ведётся реестр читателей библиотеки и фиксируется выдача книг читателям. С другой стороны, информация о проектах может широко использоваться инновационными системами. Инновационные системы расширяют базовую схему для обеспечения эффективной оценки проектов экспертами, поиска проектов по запросу инвестора и пр.

3 Semantic Web как основа методики

На основе нашего опыта в разработке информационных систем, мы пришли к решению использовать механизмы Semantic Web для описания метаданных Единого Научного Информационного Пространства РАН.

Модель данных Semantic Web, Resource Description Framework (RDF), была специально спроектирована для интеграции распределённых метаданных в Web. RDF позволяет описывать объекты, или «ресурсы», указывая их «свойства» и значения свойств. Ресурсы можно идентифицировать URI-идентификаторами, аналогично тому, как Web-страницы идентифицируются URL. Каждое свойство ресурса также указывается URI-идентификатором. Использование URI-идентификаторов вместо коротких имён – следствие «распределённости» информации, с которой мы сталкиваемся в Web. Другое следствие – это «децентрализация» информации: при описании некоторого ресурса в RDF-документе (скажем, человека) не обязательно описывать значения всякого его свойства (например, организацию, в которой он работает) – вместо этого можно указать это значение ссылкой по URI, аналогично тому, как в HTML мы указываем ссылки на другие Web-страницы. Кроме того, никакое описание ресурса не может считаться полным и окончательным – всегда есть возможность, что в Web содержится ещё какая-то дополнительная информация об этом ресурсе. Ведь ресурсы идентифицируются URI, и значит, могут быть описаны одновременно в нескольких RDF-документах. Это тоже следствие «децентрализации» информационной системы.

Язык RDF Schema позволяет описывать на RDF словари классов и свойств; можно описать и контролируемые словари вариантов значений свойств. Поскольку классы, свойства и экземпляры метаданных идентифицируются не просто именем, а URI, то это позволяет разделить их по «профилям», соответствующим разным «пространствам имён». RDFS служит базой для более сложного языка описания «онтологий» предметных областей, Web Ontology Language (OWL), который позволяет определить более сложные ограничения на применение классов и свойств, структуру метаданных.

4 Основы схем метаданных ЕНИП

В настоящее время заметна широкая тенденция по стандартизации RDF-словарей свойств метаданных для конкретных предметных областей – так называемых «обменных схем», или «профилей метаданных». Использование терминов (свойств, словарей значений и пр.), зафиксированных в стандартах, позволяет приложениям легко интегрироваться между собой, обмениваться информацией, понятной им всем. Например, при получении данных из сторонней системы, приложение может найти среди неизвестных ему свойств некоторые свойства, регламентированные стандартом, и соответственно будет уверено в их смысле, семантике, сможет правильно их проинтерпретировать. Это называется «семантической интероперабельностью», и считается одним из основных преимуществ Semantic Web.

Dublin Core Metadata Initiative (DCMI) определил минимальный набор свойств для описания цифровых ресурсов Web, а также их детализацию в рамках «общего профиля» [6]. Отдельные рабочие группы DCMI занимаются стандартизацией более специализированных профилей метаданных таких предметных областей, как библиотечная информация [7], образование [8], правительственная сфера [9], информация о людях [10] и пр.

Dublin Core стал базисом для других «стандартов обмена». В первую очередь, следует упомянуть стандарт Publishing Requirements for Industry Standard Metadata (PRISM) [11], разработанный издательскими организациями для обмена метаданными о публикациях (документах, журналах, книгах и пр.). Государственный архив Австралии выдвинул и стандартизовал основанный на Dublin Core набор профилей метаданных для описания государственной информации – AGLS Metadata Standard [12]. Заслуживают упоминания также проекты, делающие попытку спецификации схем для библиографической информации (BIBLINK [15], bibTeX [16]..), европейская инициатива по разработке схем для Math-Net [17], UKOLN CLD [18] профиль метаданных для описания цифровых коллекций и пр. Широкое применение нашли предложения по представлению информации стандарта VCard («визитная карточка») в RDF [13]. VCard определяет свойства для описания информации о людях, их контактной информации и пр. На описание информации о людях направлена также набирающая популярность открытая инициатива Friend of a Friend (FOAF) [14].

Помимо обменных «профилей метаданных», существуют инициативы по построению «онтологий» предметных областей, нацеленных больше на спецификацию большого количества классов и их взаимоотношений, нежели словарей свойств для обмена. Среди них есть весьма близкие к сложившейся в ИСИР РАН схеме минимальных научных метаданных [3]: это KA2 - Knowledge

Acquisition Community Ontology [20] и SWRC - Semantic Web Research Community Ontology [21]. Эти онтологии описывают персоналии, организации, проекты, публикации и пр. Из последних Semantic Web-разработок в этой области следует упомянуть онтологию портала Advanced Knowledge Technologies (AKT) - "AKTive Portal" [19].

При разработке предложений по наборам метаданных ЕНИП мы провели детальный анализ всех упомянутых и других (daml.org, protege.stanford.edu,...) стандартов и предложений, а также анализ различных не-RDF ориентированных предложений по стандартизации метаданных (CERIF 2000 [22], CIDOC [23], MARC и RUSMARC и др.), различных отечественных и международных систем классификации ресурсов.

Кроме того, мы основывались на нашем опыте в разработке и поддержке информационного портала ИСИР РАН (<http://isir.ras.ru>), а также нашего богатого опыта в разработке информационных систем, в частности, Инновационного портала РФФИ [24], Портала библиотеки диссертаций РГБ [25], информационного портала ГСНТИ, официального портала ВМиК МГУ (<http://cmc.msu.ru>), Портала Государственных Закупок МЭРиТ...

Следует упомянуть также ведущиеся исследования по библиотечной подсистеме [26], portalу mathnet.ru [27], системе ведения конференций [28], системе каталогизации экспериментальных данных научных исследований [29], каталогизации музейной информации [30], подсистеме ведения тезаурусов [31], каталогизации online вычислительных сервисов.

В плане взаимодействия с отечественными информационными системами, были рассмотрены варианты интеграции с Библиотекой Естественных Наук, системой Socionet Отделения общественных Наук, системой Informika ГСНТИ и пр.

5 Методика разработки и развития схем

Схемы метаданных играют в ЕНИП двоякую роль. С одной стороны, они служат «обменными схемами», с разными уровнями детализации, для обмена данными между системами, входящими в Единое Научное Информационное Пространство РАН. С другой стороны, в рамках ЕНИП стоит задача не только предложить обменные схемы, но и разработать конкретные типовые информационные системы для научных институтов, библиотек, издательских отделов и пр., которые дали бы стимул к информационному наполнению ЕНИП.

Различные информационные системы могут ориентироваться на различные предметные области. Например, одни имеют дело с научными публикациями, другие с проектами, третьи и с тем, и с другим. Соответственно, каждую конкретную предметную область предлагается описывать отдельной схемой (а точнее, набором схем),

возможно, опирающихся друг на друга. Это разбиение схемы по «минимальным предметным областям» мы называем разбиением на «модули». Модули рассматриваются не только как способ деления схемы, но и как способ деления функциональности реализуемых в рамках ЕНИП типовых информационных систем, порталов по отдельным компонентам.

Наряду с выделением профилей метаданных как таковых, делается попытка определиться со стратегиями, методиками *развития* схем – наращивания уровней, глубины описания той или иной предметной области, подходящих для разных систем. Необходимо не просто предложить схему для той или иной сущности или научной области, но и для каждой из них предложить несколько «уровней поддержки» схем, например:

- *минимальная* – необходимый разумный минимум, минимально достаточный для обмена метаданными, поддержки взаимосвязей ресурсов;
- *базовая* – объем достаточный для эффективной работы «дилетантов» в конкретной предметной области;
- *расширенная* – объем достаточный для эффективной работы «специалистов» предметной подобласти;
- *специализированная* – объем, ориентированный на «специалистов» предметной области, используется только в рамках подпространства, включающего специализированные системы.

Разбиение схемы метаданных на последовательно наращиваемые подсхемы становится возможным благодаря свойственной RDF «децентрализации» данных: каждая схема рассматривается как набор утверждений, а расширенная схема – как набор дополнительных утверждений, вдобавок к утверждениям базовой схемы. Web Ontology Language (OWL) позволяет указывать метаданные о схемах, и, в частности, их функциональную зависимость - «импорт» схем. При импорте все утверждения импортируемой схемы становятся частью импортирующей онтологии (которую мы будем называть подсхемой). Интересная особенность заключается в том, что подсхема может не только определять собственные классы и их свойства, но и указывать любую дополнительную информацию об импортированных классах и свойствах, в частности, *добавлять новые свойства к импортированным классам, уточнять тип значений и ограничения на импортированные свойства* и пр. Такая особенность, непривычная для традиционной объектной парадигмы, оказывается очень полезной для эффективного наращивания детализации схем метаданных, перехода от обменных схем к схемам конкретных информационных систем.

«Минимальные» подсхемы ориентированы в первую очередь на обеспечение максимальной гибкости обмена данными. Здесь не важна спецификация детальной и точной структуры

данных (например, разбиение почтового адреса по полям), но важно указать словарь свойств, терминов для обмена информацией в данной предметной области, а также отображения на стандартизованные и уже применяющиеся предложения по профилям метаданных. Рассмотрим методические приёмы, предоставляемые нам для этих целей языками RDF Schema и OWL:

- Импорт схем позволяет добавить в разрабатываемую схему термины других схем, в частности, стандартных профилей метаданных. Эти термины могут использоваться как непосредственно, так и специализироваться механизмами подклассов и подсвойств, если их семантика слишком абстрактна для рассматриваемого уровня детализации схемы.
- Традиционный *механизм подклассов* позволяет указывать специализацию классов, уточнение семантики термина и набора свойств. Пример: «диссертация» - подкласс «документа». Зная эту информацию, система, не работающая конкретно с диссертациями, получив данные из библиотеки диссертаций, сможет идентифицировать их как данные об абстрактном «документе» и воспользоваться такими свойствами как «автор», «издательство» и пр., проигнорировав информацию об оппонентах, дате защиты и пр.
- Механизм *подсвойств* позволяет указать специализацию свойств – для того чтобы, в первую очередь, уточнять их смысл. Приведём пример: «аннотация» - подсвойство «описания», а «альтернативное название» - подсвойство «названия» (Dublin Core). Этот нетрадиционный для объектно-ориентированных систем механизм играет ключевую роль в обеспечении *семантической interoperability* систем. Предположим, некоторая специализированная система использует понятие «официального названия» (my:legal) для именовании организаций и обменивается своими данными с другой системой, которая различает только простой термин «название» из Dublin Core (dc:title). Без дополнительной информации, вторая система не имела бы ни малейшего шанса догадаться, что же за информация идёт в текстовом поле my:legal. Теперь допустим, что вместе с данными специализированная система предоставляет также свою RDF-схему, описывающую используемые термины. В частности, в этой схеме указано, что my:legal – это подсвойство dc:title, то есть некоторая специализация стандартизованного в Dublin Core термина «название», и используется для именовании ресурса. Благодаря этой дополнительной информации вторая система сможет воспользоваться данными, указанными в поле my:legal. Естественно, она не сможет автоматически воспользоваться информацией о том, что это не просто название, а именно

«специализированное официальное название», но эта информация систему и не интересует в рамках её предметной области. Помимо уточнения смысла, подсвойство может уточнять характеристики суперсвойства. В частности, подсвойства могут иметь более специализированный тип значений (см. ниже пример с «отчётами по проекту»).

- OWL позволяет указывать *эквивалентность* классов, свойств, либо экземпляров (например, элементов различных словарей значений). Эти механизмы, наряду с механизмами подклассов и подсвойств, позволяют указать отображение схем на стандартные и широко применяющиеся профили метаданных, что гарантирует семантическую интероперабельность.

На этапе перехода от «минимальной» к «базовой» и более специализированным подсхемам встаёт вопрос о более чёткой спецификации структуры данных – в частности, чёткой *спецификации типов значений свойств*. Это возможно благодаря уже упомянутому механизму введения дополнительных утверждений об импортированных ресурсах, в частности, свойствах. Минимальная схема может не указывать явно тип данных свойства, если он потенциально может быть уточнён впоследствии, тогда более специализированная схема сможет указать специализацию этого типа. Если же тип значений с большой вероятностью подойдёт всем системам, то можно указать его уже в «минимальной» схеме, таким образом, накладывая некоторую резонную «строгость» на формат обмена. Например, можно указать, что «дата выпуска» издания имеет значения типа «дата» (xs:date), в чётко регламентированном формате (W3C-DTF). Это требование обязует все системы экспортировать данные о дате выпуска в этом формате, а не в виде произвольной строчки, и исключит ситуации непонимания формата при импорте данных. Рассмотрим ещё один пример. Если тип свойства – объект, то минимальная схема может указать тип значений как некоторый абстрактный класс, а специализированная схема – уточнить тип значений, указав его подкласс. Пусть мы имеем свойство «публикация выполнена по проекту», позволяющее указать литературу («публикации»), полезную для понимания проекта. В более специализированной схеме введём понятие «отчёта по проекту»: заведём соответствующий класс «Проектный Отчёт» (имеющий дополнительные метаданные, такие как номер отчёта) и свойство «отчёт по проекту», позволяющее сопоставлять проектам «отчёты». Это свойство мы будем считать подсвойством, частным случаем «публикации, выполненной по проекту», но с более специализированным типом значений.

В процессе детализации схемы возникает желание *структурировать* некоторые данные, которые до этого можно было считать строковыми. Например, резонно представлять телефонный номер строкой. Но зачастую возникает потребность

указать к нему комментарий, если указано несколько телефонов, например, «предпочтительный», «рабочий», «домашний», «факс», «мобильный телефон» и пр. В таком случае каждый телефон представляет собой структуру: «комментарий» плюс «собственно значение». Аналогичный пример – адрес может быть структурирован с выделением страны, почтового индекса, региона, города, улицы, адреса по улице и пр. Ещё пример – степень (кандидата, доктора наук по некоторой дисциплине) можно минимально представить контролируемым словарём. Однако на базовом уровне детализации нужна возможность указать также дату присуждения степени, специальность ВАК и пр. А в схеме библиотеки диссертаций указать диссертацию, включающую также метаданные о дате и месте защиты, научных руководителях, оппонентах, рецензентах и пр.

Очевидно, что *механизм наращивания структуризации* должен учитывать необходимость семантической интероперабельности систем. Например, система, которая представляет адрес текстом, и система, которая представляет его в структурированном виде, должны беспрепятственно обмениваться адресной информацией и «понимать» друг друга. Язык RDF содержит встроенный механизм, помогающий нам в этом – предопределённое свойство *rdf:value* [5]. Это свойство позволяет выделить «собственно значение» из структурированного описания (не обязательно строковое, в рассмотренном нами примере со степенями значением является словарный элемент). Если система получает информацию, в которой значением свойства является некоторый объект неизвестного системе типа, но имеет свойство *rdf:value*, то система может опустить этот неизвестный объект и попытаться рассмотреть значение *rdf:value* как значение обрабатываемого свойства. Аналогичное поведение можно предложить для подсвойств *rdf:value*.

Рекомендация по применению Dublin Core в RDF использует механизм *rdf:value* для определения «квалификаторов значений» (Encoding Scheme [5]). Эти квалификаторы указывают схемы, помогающие системам в интерпретации значения свойства (контролируемые словари, правила разбора значения). Например, строку УДК можно «квалифицировать», указав вместо строчки объект типа «УДК» (dcterms:UDC), что поможет другим системам узнать систему классификации. Кроме того, можно указать название этого термина классификатора УДК помимо номера и пр. Рассмотренный нами механизм наращивания структуризации – некоторая аналогия «квалификаторов значений» Dublin Core. Механизм же введения подсвойств для уточнения семантики свойства обозначается в Dublin Core термином Element Refinement [5]. Вместе эти два механизма, плюс механизм импорта схем в OWL, возможность расширения набора свойств классов, уточнения характеристик классов и свойств в подсхемах,

расширения и специализации в подсхемах контролируемых словарей - служат основой для чёткой методики развития и специализации схем.

Как пример интерпретации менее специализированными системами (например, агрегирующими каталогами) информации, предоставляемой им более специализированными, можно рассмотреть алгоритм упрощения (DumbDown [5]), предлагаемый Dublin Core. Этот алгоритм заменяет все подсвойства (Element Refinements) свойств DC на их суперсвойства, входящие в 15 базовых элементов Dublin Core, а всякий «квалификатор значения» заменяет на «собственно значение», указанное в rdf:value. В результате детализированные структурированные данные сводятся к абстрактным и планарным. Аналогичный алгоритм можно использовать и для упрощения информации от более специализированной схемы ЕНИП некоторой предметной области к более общей («базовой», «минимальной»).

Как уже упоминалось, схемы метаданных ЕНИП играют двойную роль – во первых, они служат схемами обмена данными, а с другой стороны – служат для построения конкретных информационных систем с помощью технологий ИСИР [1,2]. В соответствии с рассмотренным нами делением на подсхемы, построение информационной системы подразумевает, в частности:

- выбор используемых предметных областей – «модулей» информационной системы (либо из набора стандартных, предлагаемых схемами ЕНИП, либо собственных, введённых в рамках данного специализированного приложения);
- выбор уровней специализации для каждой предметной области – минимальный, базовый, расширенный, специализированный....

6 Заключение

Схемы метаданных являются ключевой компонентой в организации Единого Научного Информационного Пространства РАН. Ввиду большого объёма, мы не приводим описания и рекомендации по применению схем в данной статье, полное описание будет опубликовано отдельно.

Литература

- [1] Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В. А., Сысоев Т.М. "Возможности технологий ИСИР в поддержке Единого Научного Информационного Пространства РАН" // Статья в данном сборнике.
- [2] А.А. Бездушный, А.Н. Бездушный, А.К. Нестеренко, В.А. Серебряков, Т.М. Сысоев, "Архитектура RDFS-системы. Практика использования открытых стандартов и технологий Semantic Web в системе ИСИР", Пятая Всероссийская научная конференция: "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Санкт-Петербург, 2003.
- [3] А.А. Бездушный, А.Н. Бездушный, А.Б. Жижченко, М.В. Кулагин, В.А. Серебряков. RDF схема метаданных ИСИР. // Сборник научных трудов X научно-практического семинара "Новые технологии в информационном обеспечении науки". Москва: 2003, с.141-159.
- [4] RDF Primer. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/rdf-primer/>
- [5] Expressing Qualified Dublin Core in RDF / XML. <http://dublincore.org/documents/dcq-rdf-xml/>
- [6] DCMI Metadata Terms. <http://dublincore.org/documents/dcmi-terms/>
- [7] Library Application Profile. <http://www.dublincore.org/documents/library-application-profile/>
- [8] Education Working Group: Draft Proposal. <http://dublincore.org/documents/education-namespace/>
- [9] Government Application Profile. <http://dublincore.org/documents/gov-application-profile/>
- [10] DCMI Agents Working Group. <http://dublincore.org/groups/agents/>
- [11] PRISM: Publishing Requirements for Industry Standard Metadata. <http://www.primstandard.org/>
- [12] National Archives of Australia – AGLS. http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html
- [13] Representing vCard Objects in RDF/XML. W3C Note 22 February 2001 <http://www.w3.org/TR/vcard-rdf>
- [14] FOAF Vocabulary Specification. Namespace Document 1 May 2004. <http://xmlns.com/foaf/0.1/>
- [15] BIBLINK Project. <http://hosted.ukoln.ac.uk/biblink/>
- [16] bibTeX Definition in Web Ontology Language (OWL) Version 0.1. Working Draft, 2004. <http://visus.mit.edu/bibtex/0.1/>
- [17] Math-Net Schemes. <http://www.iwi-iuk.org/material/RDF/1.1/>
- [18] UKOLN Research Support Libraries Programme Collection Description. <http://www.ukoln.ac.uk/metadata/rslp/schema/>
- [19] AKT Reference Ontology. <http://www.aktors.org/publications/ontology/>
- [20] KA2 - Knowledge Acquisition Community Ontology. <http://ontobroker.semanticweb.org/ontos/ka2.html>
- [21] SWRC - Semantic Web Research Community Ontology. <http://ontobroker.semanticweb.org/ontos/swrc.html>
- [22] CERIF: Common European Research Information Format. <http://www.cordis.lu/cerif/src/about.htm>

- [23] The CIDOC Conceptual Reference Model.
<http://cidoc.ics.forth.gr/>
- [24] Вежневцев А.А., Бездушный А.Н., Серебряков В.А., Цыганов С.А. "О реализации систем поддержки применения результатов фундаментальных исследований" // Статья в данном сборнике.
- [25] Лаврёнова О.А., Вежневцев А.А. Структура и реализация электронной библиотеки диссертаций в РГБ // Сборник докладов Второй Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Санкт-Петербург, 2003
- [26] П.М. Курив, Д.В. Котеров, Н.Е. Калёнов. Архитектура и функциональность Библиотечной Подсистемы ИСИР РАН // Электронный журнал, посвященный созданию и использованию электронных библиотек, том 7, выпуск 1. Москва: Институт развития информационного общества - 2004.
- [27] А. С. Аджиев, А. Н. Бездушный, С. П. Коновалов, В. А. Серебряков. Общероссийский WEB-портал математических ресурсов. // Сборник докладов Второй Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Санкт-Петербург, 2003
- [28] А.Н.Алексеев, А.В.Созыкин, Г.Ф.Масич, А.Н. Бездушный. Подсистема проведения конференций и ее метаданные // Электронный журнал, посвященный созданию и использованию электронных библиотек, том 7, выпуск 2. Москва: Институт развития информационного общества - 2004.
- [29] Р.А. Коротченко, И.О. Ярощук, А.Н. Бездушный. Версия схемы метаданных экспериментальных исследований с приложением в гидроакустике // Электронный журнал, посвященный созданию и использованию электронных библиотек, том 7, выпуск 1. Москва: Институт развития информационного общества - 2004.
- [30] А.В. Котов, В.А. Серебряков, А.А. Столяров. Информационная система "Востоковедение" // Сборник докладов Третьей Всероссийской научной конференции "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Петрозаводск, 09.2001
- [31] М.Х. Нгуен, А.С. Аджиев. Описание и использование тезаурусов в информационных системах, подходы и реализация // Электронный журнал, посвященный созданию и использованию электронных библиотек, том 7, выпуск 1. Москва: Институт развития информационного общества - 2004.

Proposals for metadata element sets for scientific information resources in USIS RAS

Bezdushny A.N. Zhizhchenko A.B. Kalenov N.E.
Kulagin M.V. Serebriakov V.A. Bezdushny A.A.

The paper is devoted to the proposals for metadata element sets and ontologies for scientific information resources of the Russian Academy of Sciences in the framework of the Unified Scientific Information Space project (USIS RAS). We consider the requirements, problems and tasks of establishing USIS RAS, a space of interconnected distributed heterogeneous information systems. We briefly outline the application domains and information resource types, which are planned to be presented in USIS. Further the article describes the principles used to describe metadata, and related standards and proposals for metadata schemas that were analyzed and used as a basis for USIS RAS schemas.