

ПРЕДЛОЖЕНИЯ ПО НАБОРАМ МЕТАДААННЫХ ДЛЯ НАУЧНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ

А. Н. БЕЗДУШНЫЙ, М. В. КУЛАГИН, В. А. СЕРЕБРЯКОВ,
А. А. БЕЗДУШНЫЙ, А. К. НЕСТЕРЕНКО, Т. М. СЫСОВЕВ
Вычислительный центр РАН, Москва, Россия
e-mail: bezdushn@ccas.ru, ql@ccas.ru, serebr@ccas.ru,
alix@7ka.mipt.ru, alexn@ccas.ru, tim@ccas.ru

This paper addresses proposals on the formation of metadata element sets 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, aims and specific tasks arising in USIS RAS, as a space of interconnected distributed heterogeneous information systems. We briefly outline the application domains and information resource types, which are intended to be represented in USIS. Further, the report 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. The capability of the integrated system of information resources technology (ISIR) is discussed.

1. Цели и задачи ЕНИП РАН

В мире имеется достаточно большое количество информационных систем для работы с научными данными, наукоемкой информацией. Каждое научное учреждение имеет данные о публикациях сотрудников, научных исследованиях и проектах, располагает теми или иными наукоемкими данными. Многие из учреждений имеют собственные информационные системы для наукоемкой информации, которые в каком-то виде ее хранят. Однако интересы пользователей не исчерпываются одной информационной системой, пусть даже замечательной. Как правило, интересы ученых шире поставленной задачи, находятся на стыке научных областей. Практически невозможно в рамках одной организации собрать информацию, которая удовлетворила бы запросы всех сотрудников. Даже если бы в какой-то момент это удалось, то в силу огромного динамизма научных исследований невозможно обеспечить какую-то приемлемую полноту, актуальность данных, для которых системы служат. Попытки объединить данные научных учреждений в одной централизованной системе на достаточно высоком уровне не могут привести к положительному результату. Это можно увидеть на примере системы ERGO [1] и финского проекта создания национальной университетской системы [2]. Препятствиями на этом пути являются как объемы информации, так и сложность обеспечения полноты, актуальности данных, невозможность сведения данных всего разнообразия научных областей к единой структуре.

Российская академия наук объединяет большое число научно-исследовательских учреждений и коллективов, расположенных на всей территории России и вовлеченных во все 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 членов мирового научного сообщества, работников 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 мало внимания уделяется вопросам интероперабельности — слабо применяются соглашения по стандартизации электронного представления информационных ресурсов и соответствующие средства, призванные поддержать интеграцию информационных ресурсов, повысить точность поиска и т. п. В результате пользователь не может получить полную и достоверную информацию об интересующих его ресурсах. Очевидно, что каждая область науки, оперируя со своими специфичными 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.

Инициатива по организации единого научного информационного пространства (ЕНИП) РАН призвана помочь научным коллективам сделать ряд шагов в направлении интеграции разнородных научных информационных и программных ресурсов отдельных научных учреждений, предоставления пользователям более эффективных средств интеграции и поиска информации, научной коммуникации, сотрудничества и совместной работы. Под единым пространством понимается не формирование централизованной системы, не навязывание всем одних и тех же решений, а стремление последовательностью практических шагов, совместными усилиями научных коллективов РАН:

- сформулировать взаимосогласованный набор соглашений, правил и открытых стандартов;

- приготовить совокупность макетов и типовых решений для реализации адаптеров прикладных систем, инфраструктурных служб, поддерживающих разные уровни интероперабельности распределенных гетерогенных данных и приложений;

- создать ряд информационных систем общего назначения, следующих этим соглашениям, использующих эти реализации, допускающих модульную организацию, наращивание функциональных возможностей;

- применить эти результаты для решения соответствующих задач учреждений РАН.

Все нацелено на то, чтобы помочь учреждениям РАН в решении общих информационных задач, достижении требуемой интеграции с другими учреждениями РАН.

В общем случае можно сказать, что информационные системы учреждений РАН отличаются огромные объемы и низкая структурированность данных, распределенный характер, неоднородность, независимость и разные условия сопровождения, управления и политики доступа к информационным источникам и сервисам. В таких случаях выделяют и стара-

ются решить проблемы общего вида.

Техническая интероперабельность. Для обеспечения взаимодействия разнородных информационных источников необходимо поддерживать согласованные интерфейсы, протоколы и механизмы доступа к информационным ресурсам.

Синтаксическая интероперабельность. Данные, доступные из информационных источников, как правило, отличаются синтаксической неоднородностью, разнообразием моделей данных и форм представления. Следовательно, необходимо выработать и согласовать унифицирующий подход приведения данных к наиболее распространенным моделям данных и форматам.

Сбор метаданных. В сложившейся ситуации, когда сведения о ресурсах в большом объеме 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 трудно выбрать действительно полезные знания. Учитывая это обстоятельство, для получения 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 разнотипной информации. Центральным в обслуживании слабоструктурированных и унаследованных коллекций информации является процесс “сбора” метаданных, в ходе которого из коллекций в соответствии с требованиями синтаксической интероперабельности извлекаются и структурируются метаданные, формируется индексная информация для обеспечения локального поиска, маршрутизации распределенных запросов, ранжирования результатов запросов.

Семантическая интероперабельность. Метаданные могут относиться к различным предметным областям, в рамках одной иметь разное выражение и интерпретацию. Создание и согласование стандартных прикладных профилей метаданных и онтологий упростят интеграцию разнообразных систем, позволят автоматизировать обмен метаданными, их обработку и преобразование, повысить точность и эффективность поиска. Необходимо разработать набор элементов метаданных для общей научной информации и некоторые профили метаданных конкретных научных областей, согласуя их с научным сообществом и международными стандартами, обеспечить выделение и согласование стандартных классификаторов ресурсов и тезаурусов.

Поддержка глобальной идентификации ресурсов. Использование глобально уникальных идентификаторов дает возможность установления взаимосвязей между ресурсами разных репозиторий (под репозиторием мы понимаем интероперабельный информационный источник в указанном выше смысле) распределенной среды, объединять связанные данные отдельных репозиторий в виртуально-единые ресурсы. Это предоставит пользователям возможность производить навигацию среди ресурсов всей информационной системы, выполнять косвенный поиск, в том числе и по связям между ресурсами в разных репозиториях, упростит задачу объединения результатов поисковых запросов разных репозиторий.

Совместный “поиск” — маршрутизация запросов и объединение ответов. Для понижения нагрузки на сеть и повышения эффективности распределенные запросы должны выполняться не во всем множестве репозиторий, а только в соответствующем запросу подмножестве. Этот процесс называют маршрутизацией запросов, при принятии решения он использует “предварительные знания” — информацию, распространяемую в среде

именно с целью обоснованной рассылки поисковых запросов, формируемую на основе локальных индексов. Процесс объединения ответов репозиторийев, к которым был направлен запрос, в единый ответ системы должен обеспечивать как устранение вторичных вхождений описаний одного и того же ресурса (дублирование описаний), которые с большой вероятностью могут появиться из разных частей распределенной среды, так и совместное ранжирование результатов, поступающих от этих частей.

Балансировка нагрузки. Для снижения нагрузки на телекоммуникационные и вычислительные ресурсы при обработке запросов в случае доступа к часто используемой информации применяются механизмы балансировки нагрузки. Балансировка нагрузки предполагает репликацию метаданных с маломощных серверов на более мощные. В этом случае происходит концентрация поисковой информации на ограниченном числе мощных серверов, участвующих в ответе на поисковые запросы. В рамках обмена и репликации данных встают проблемы обеспечения связывания и интеграции ресурсов независимо сопровождаемых источников информации, выявления дубликатов.

Распределенная авторизация доступа и принцип единой аутентификации. Различные информационные источники, составляющие распределенную среду, имеют различные механизмы контроля доступа к информации. Средства контроля доступа должны быть также предоставлены и интегрированы средой, должен быть указан общий подход к безопасности систем. Для того чтобы избавить пользователя от необходимости регистрироваться в каждом информационном источнике, должен быть поддержан принцип единого входа.

Примером эффективной интеграции данных в контексте потребностей Российской академии наук может быть следующий. Организации, обладающие крупными цифровыми коллекциями, например библиотеки, смогут с помощью решений ЕНИП РАН предоставить непосредственный доступ к своим информационным ресурсам. Научные учреждения, не обладающие подобными ресурсами, могут обеспечить возможность индексирования своих данных, экспорт данных и агрегацию их в более крупные хранилища. Такие хранилища могут предоставляться крупными учреждениями, научными центрами и отделениями РАН. Таким образом, можно поддержать агрегацию научных данных в соответствии с организационной структурой РАН (рис. 1). Находящиеся на верхнем уровне информационные хранилища отделений, в основном поддерживая профильные коллекции информации и услуг, могут участвовать в маршрутизации запросов объединенной среды, доступ к которой открывается через “единое окно”, портал научного пространства РАН.

Организации могут предоставлять свои данные в слабоструктурированном виде, например как текстовые или XML-документы. Для интеграции таких данных в систему необходимо выделять в них структуру, приводить к агрегатным моделям и схемам данных и связывать между собой и с имеющимися в среде данными. К примеру, организация может предоставлять информацию о научных публикациях, отчетах, результатах научной деятельности за некоторый промежуток времени. Желательно, чтобы информация была автоматизированным образом проанализирована и структурирована (рис. 2) с выделением авторов публикации, издательства и других атрибутов — названия, ISBN и пр. При интеграции данных может выясниться, что в хранилище уже имеется некоторая информация об авторах публикации, например другие статьи, или информация об участии в конференциях и пр., поэтому желательно, чтобы обеспечивались автоматизированное сопоставление и связывание соответствующих ресурсов.

Для решения вышеуказанных задач необходимо сформировать и стандартизировать тематические профили метаданных [3], их модели данных, определить средства записи схем

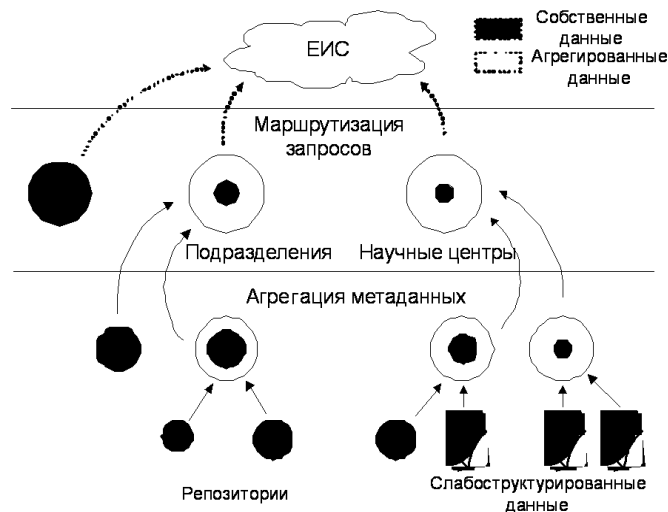


Рис. 1. Агрегация данных в соответствии с организационной структурой РАН.

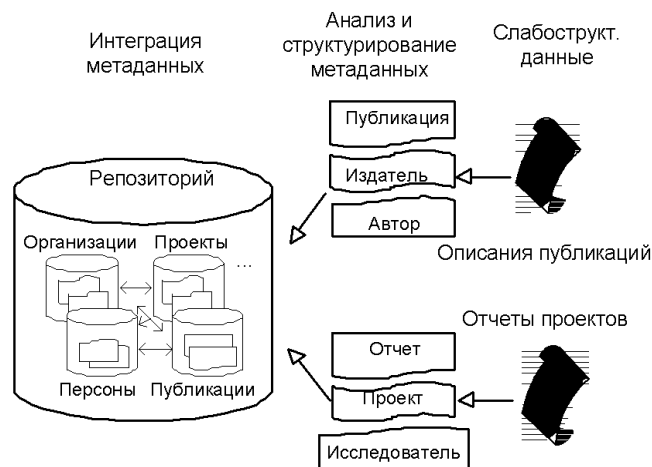


Рис. 2. Анализ и структуризация слабоструктурированных данных.

данных, правила их специализации и согласования, выработать типовые интерфейсы взаимодействия [4]. Необходимо предоставить учреждениям РАН типовые программные и инструментальные средства [5], которые упростят или обеспечат участие учреждений РАН в едином информационном пространстве, структуризацию данных и их интеграцию в единую среду. Помимо прочего стоят задачи создания сети взаимосвязанных инфраструктурных или агрегатных узлов, скорее всего, соответствующих структурным и тематическим объединениям РАН (отделениям, научным центрам РАН), предоставления доступа к среде в целом, ее соответствующим (объединениям) фрагментам через “единое окно”.

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

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 ЕНИИ РАН 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, производятся 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.

2. Информационное наполнение ЕНИИ РАН

Рассмотрим информационное наполнение единого научного информационного пространства на начальном этапе поддержки ЕНИИ. 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 основной “информационный ресурс” науки — люди, которые эту науку делают (рис. 3), 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 (рис. 4). Это, в первую очередь, поддержка следующих видов ресурсов

- “Участники научной деятельности” — центральное звено, вся информация в РАН связана с научной деятельностью ее 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, отчеты сотрудников (научные “Публикации” и “Диссертации” сотрудников РАН, их “Коллекции”), возможно, административные “Постановления” и “Распоряжения”. Примерами специализации публикаций могут служить, например, “Тезисы конференций” и т. п.

- “Научная деятельность” — документы, в частности “Проекты”, отражающие процесс научной деятельности, информация о результатах проектов, патентах и т. п.

- “Научные события” — ресурсы, 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.

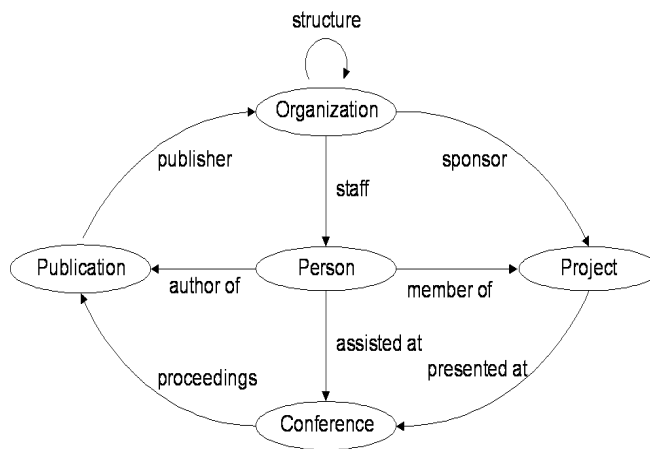


Рис. 3. Пример взаимосвязи ресурсов ЕНИИП.

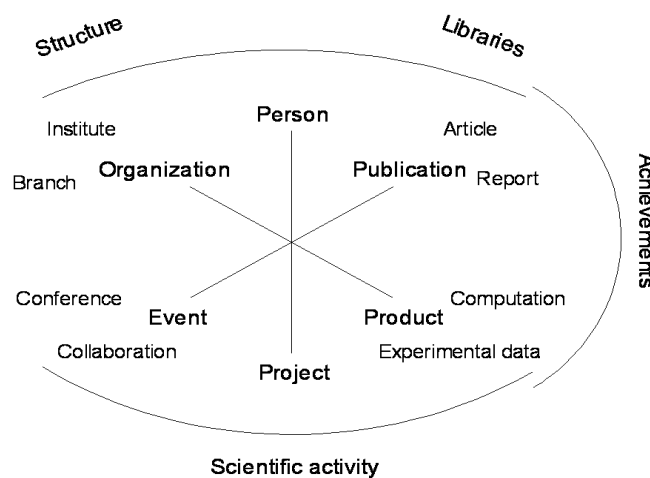


Рис. 4. Базовые предметные области ЕНИИП.

- “Базы данных”, предоставляющие автономные коллекции информации с той или иной степенью интеграции с ЕНИИП РАН, 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 несущественна.

3. Semantic Web как основа методики

На основе 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 распределенных 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 имен — следствие “распределенности” 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 еще 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, это позволяет разделить их по “профилям”, соответствующим разным “пространствам имен”. 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 ЕНИИ РАН

В настоящее время заметна тенденция 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 не известных 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 инфор-

мации (BIBLINK [13], bibTeX [14]...): европейская инициатива по разработке схем для Math-Net [15], UKOLN CLD [16], профиль метаданных для описания цифровых коллекций и пр. Широкое применение нашли предложения по представлению информации стандарта VCard (“визитная карточка”) в RDF [17], который 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) [18].

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 [19]: это КА2 — 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 (АКТ) — “АКТive Portal” [22].

При разработке предложений по наборам метаданных ЕНИИП РАН 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 [23], CIDOC [24], 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 ЕИС РАН (<http://www.ras.ru>), а также в разработке информационных систем, в частности Инновационного портала РФФИ [25], портала библиотеки диссертаций РГБ [26], информационного портала ГСНТИ, портала ВМиК МГУ (<http://cnc.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 интерактивных вычислительных сервисов (web-сервисов) и т. п.

В плане взаимодействия с отечественными информационными системами были рассмотрены варианты интеграции с Библиотекой естественных наук, системой Socionet Отделения общественных наук, системой Informika ГСНТИ и др.

5. Методика разработки и развития схем

Схемы метаданных играют в ЕНИИП РАН 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 ЕНИИП РАН 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 ЕНИИП РАН.

Различные информационные системы могут ориентироваться на различные предметные области. Например, одни имеют дело с научными публикациями, другие — с про-

ектами, 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 РАН типовых информационных систем, порталов по отдельным компонентам.

Наряду с выделением профилей метаданных как таковых делается попытка определиться со стратегиями, методиками 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, попытаться выделить их ортогональные “измерения”, например (рис. 5):

— *минимальный* — необходимый разумный минимум, минимально достаточный для обмена метаданными, поддержки взаимосвязей ресурсов;

— *базовый* — объем, достаточный для эффективной работы “дилетантов” в конкретной предметной области;

— *расширенный* — объем, достаточный для эффективной работы “специалистов” предметной подобласти;

— *специализированный* — 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 в дополнение к утверждениям базовой схемы. Web Ontology Language (OWL) позволяет указывать метаданные о схемах, в частности их функциональную зависимость — “импорт” схем. При импорте все утверждения импортируемой схемы становятся

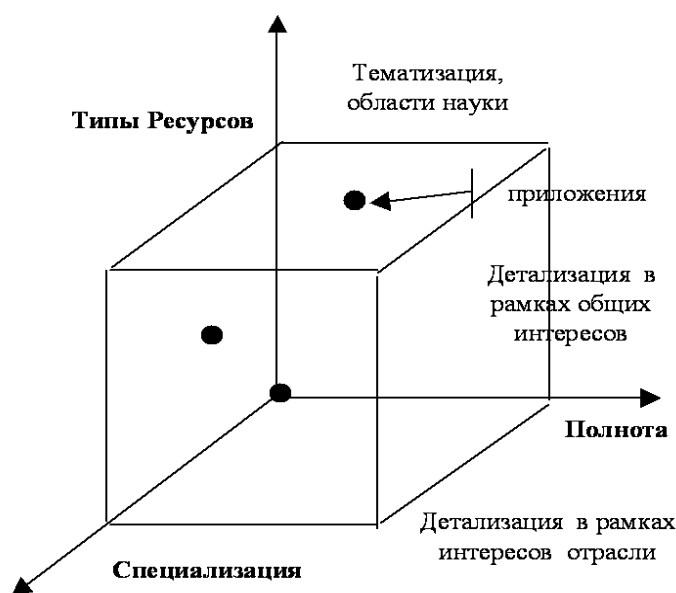


Рис. 5. “Размерности” описания схем.

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 отображение на стандартизованные и уже применяющиеся предложения по профилям метаданных. Рассмотрим методические приемы, предоставляемые 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 как использоваться 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, уточнять семантику термина и набор 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, уточнить их смысл. Приведем 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 “официальное название” (my:legal) для именования 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 идет 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, т. е. некоторая специализация стандартизованного в Dublin Core термина “название”, и используется для именования 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 ее предметной области. Помимо уточнения смысла подсвойство может уточнять характеристики суперсвойства. В частности, подсвойства могут иметь более специализированный тип значений (см. ниже пример с “отчетами 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 встает вопрос о более четкой спецификации структуры данных, в частности четкой *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 уточнен впоследствии, тогда более специализированная схема сможет указать специализацию этого типа. Если же тип значений с большой вероятностью подойдет 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 четко 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 еще 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 введем понятие “отчета по проекту”: заведем соответствующий класс “Проектный отчет” (имеющий дополнительные метаданные, такие как номер отчета) и свойство “отчет по проекту”, позволяющее сопоставлять проектам “отчеты”. 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. Еще пример — степень (кандидата, доктора наук по некоторой дисциплине) можно минимально представить контролируемым словарем. 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 в этом, — предопределенное свойство *rdf:value* [32]. 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 не известного 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`.

Вместе механизмы подсвойств и `rdf:value` плюс импорт 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 четкой методики развития и специализации схем.

6. Технологии ИСИР — архитектура и средства разработки

Рассмотренные выше видение и требования к формированию ЕНИП РАН в определенной степени могут быть поддержаны решениями на основе кросс-платформенной технологии интегрированной системы информационных ресурсов (ИСИР) [33–39]. Технологии ИСИР позволяют поддержать многие элементы жизненного цикла распределенного гетерогенного информационного пространства. Помимо средств интеграции систем программными компонентами ИСИР предоставляется также ряд средств для построения сложных информационных корпоративных порталов, шлюзов к информационным системам. Интегрированная система информационных ресурсов базируется на открытых технологиях и стандартах [40, 41].

Архитектура интегрированной распределенной среды, формируемой на базе решений ИСИР, следует традиционному многоуровневому строению, принятому в федеративных базах данных и системах на основе механизма посредников. В подходе цифровых библиотек такому строению следует, например, архитектура MIA [42]. В нашем случае специализация архитектуры связана с применением технологий Semantic Web как базиса архитектуры (рис. 6). Нижний уровень соответствует разнородным провайдерам данных, информацию которых необходимо интегрировать в среду.

Уровень адаптеров обеспечивает единообразный доступ к данным репозитория, отображение модели данных провайдера в каноническую модель данных репозитория, в нашем случае — в модель данных Semantic Web. Каждый источник погружается в стандартную оболочку, предоставляемую ИСИР, либо должен самостоятельно поддерживать согласованные интерфейсы, протоколы, модель данных и язык запросов. Провайдер данных и оболочка могут быть распределены по сети. На данный момент имеются адаптеры для RDBMS, ODBMS и RDF-источников информации, ведется разработка поддержки прозрачного взаимодействия с LDAP-каталогами. Для спецификации канонической (интегрирующей) схемы данных репозитория, объединяющей “схемы экспорта” отдельных источников информации, мы используем язык RDFS-схем, а точнее, подмножество Web Ontology Language (OWL).

Специализированные компоненты-посредники, надстраиваемые над адаптерами, позволяют репозиториям участвовать в обмене и репликации данных, маршрутизации запросов между узлами. Эти компоненты позволяют прозрачно объединить данные информационных источников в интегрированную информационную среду, которая соответствует “интегрирующей” схеме данных. Интегрированные данные могут предоставляться различным клиентам, в частности, имеется возможность построения web-порталов к информационной системе, а также поддержки протоколов SOAP, Z39.50 и SDLIP.

Решения ИСИР содержат стандартные средства, обеспечивающие функционирование

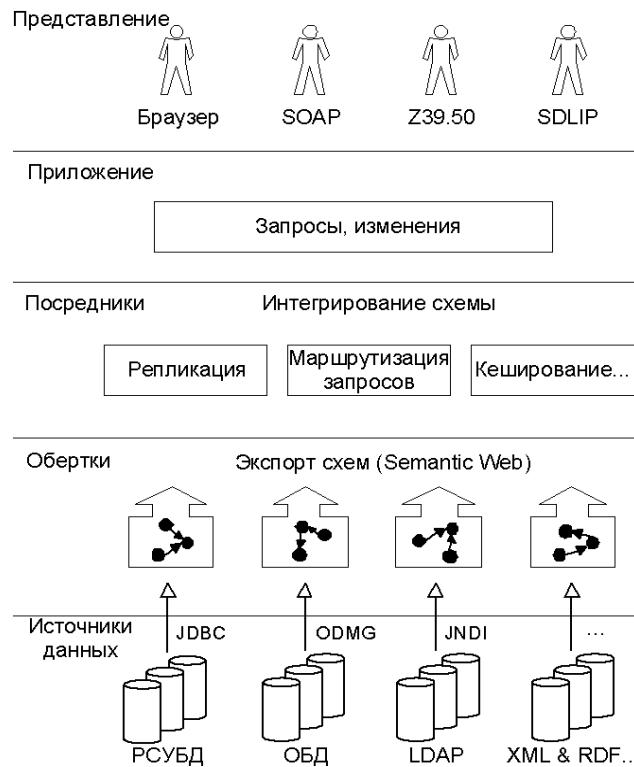


Рис. 6. Многоуровневая архитектура.

всех уровней архитектуры. Стандартная реализация уровня адаптеров базируется на так называемом ядре ИСИР, обеспечивающем отображение модели данных провайдера в модель данных языка Java. Ядро следует стандартизируемому подходу к прозрачному хранению Java-объектов в различного рода СУБД.

Средства разработки ИСИР в определенной степени автоматизируют настройку ядра на источник. Рассмотрим случай провайдера данных в виде реляционной базы данных как наиболее распространенный на практике. При разработке нового репозитория с новой БД необходимо лишь представить схему данных, которые планируется поставлять в интегрированную среду, — это OWL-онтология предметной области.

Как было сказано, основной задачей ядра ИСИР является формирование “объектной базы данных” из исходного источника данных, для поиска и управления объектами в ней используются подмножества объектного языка запросов OQL и интерфейса объектных баз данных ODMG. Помимо поддержки хранения объектов (хранимых persistent объектов) ядро ИСИР включает дополнительные базовые сервисы, необходимые в реальных системах: службы безопасности, учета изменений, средств категоризации объектов и пр.

Прикладные и системные программные компоненты ИСИР (рис. 7) работают с данными либо в их объектном представлении, используя ODMG и OQL, либо в Semantic Web + XML-представлении, используя RDF/XML и XPath. XML-представление объектных данных позволяет применить к ним всю мощь XML-технологий и интегрировать в архитектуру различные разработки, связанные с XML [41]: XSLT и XForms, OWL/XSD Validation, SOAP и web-сервисы.

ИСИР-технологии специально оптимизированы для создания информационных порталов. Средства построения порталов плотно интегрированы с инфраструктурой ИСИР и

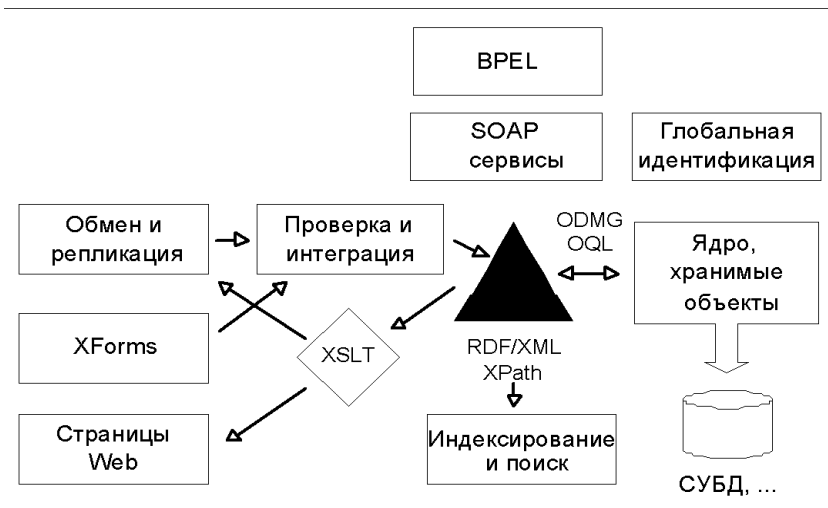


Рис. 7. Взаимосвязь применяемых технологий.

ориентируются на применение новейших W3C XML-технологий. Представление страниц сайта опирается на принципы XML/XSLT-публикации.

ИСИП-решения содержат все стандартные базовые блоки, необходимые для быстрой разработки порталов: регистрация и персонализация, портлеты и агрегация контента, новости, форумы, рассылки, система ведения документации и управления контентом [38], административный интерфейс и пр. [39]. Эти компоненты реализованы на основе упомянутых выше средств. ИСИП-решения апробированы в ряде применений, как научных [36, 42], так и коммерческих.

7. Инфраструктурные службы распределенной среды

7.1. Обмен данными

Как видно, в распределенной среде необходима поддержка обмена как минимум двух типов информации между репозиториями: обмен (репликация) метаинформацией и обмен индексами для осуществления совместного поиска.

Обмен информацией строится на базовых принципах, предложенных в спецификации протокола Common Indexing Protocol (CIP). Согласно данным предложениям, обмен рассматривается только между парой серверов, более сложные случаи сводятся к серии таких обменов. Для пары серверов определяют два различных отношения: push и poll. В первом случае (push) соединение инициируется сервером, который передает информацию другой стороне. Во втором случае (poll) инициирующая сторона запрашивает необходимые ей данные у другого сервера.

У инициирующей обмен стороны имеется ряд правил, которые описывают, когда осуществлять взаимодействие и какие данные передавать или запрашивать. Обмен начинается или в результате наступления заданного события (например, изменение определенного количества ресурсов со времени последней связи), или по заданному расписанию. Состав передаваемых объектов определяется запросом, который может накладывать ограничения на их тип, значение свойств, принадлежность заданному репозиторию, а также выделять ресурсы, изменившиеся после указанного момента времени.

Наиболее простым по затратам методом подключения источника данных к распределенной среде является реализация поддержки отношения типа push. Для этого достаточно уметь выгружать данные в формате RDF/XML и передавать их по протоколу SOAP на заданный адрес, например, один раз в день. Данный подход может быть использован для наиболее удаленных или маломощных узлов сети, а также в том случае, когда нет возможности установить сервисы ИСИР или совместимые с ними.

Если репозиторий, передающий данные, должен поддерживать отношение poll, требования к нему возрастают: необходимо иметь web-сервис обмена, принимающий запросы на репликацию, а также уметь выбирать ресурсы согласно приведенным выше критериям. Существующие технологии ИСИР удовлетворяют этим требованиям.

В качестве транспортного протокола используется SOAP, собственно данные передаются как бинарные объекты в соответствии с соглашением “SOAP with Attachments”. В спецификации SIP для этих целей предложен простой текстовый протокол, но разработчиками было принято решение использовать SOAP благодаря его популярности, а также существующей поддержке ядром ИСИР.

7.2. Репликация ресурсов

Сервис глобальной идентификации помимо того, что гарантирует уникальность генерируемых идентификаторов, позволяет централизованно хранить дополнительную информацию, ассоциированную с идентификатором. В частности, вместе с идентификатором хранится информация о репозитории, которому принадлежит данный ресурс.

Метаданные ресурса могут быть изменены только в рамках репозитория, к которому относится ресурс. Если информация была реплицирована на другие серверы, то там она рассматривается как информация “только для чтения”. В частности, это ограничение не позволяет редактировать одни и те же метаданные в двух разных местах.

Таким образом, из всех доступных копий метаданных в распределенной среде только одна считается актуальной, а остальные потенциально могут содержать устаревшую информацию. Для поддержания копий в актуальном состоянии используется процесс репликации. Репликация использует общий сервис обмена данными, в качестве передаваемой информации выступает RDF/XML-представление ресурсов.

7.3. Совместный поиск

Совместный поиск позволяет пользователю искать нужную ему информацию среди всех распределенных по среде ресурсов. При этом неизбежны некоторые компромиссы, из-за которых поиск выполняется не настолько точно, как если бы запрос осуществлялся в рамках одного репозитория. Это связано с тем, что ресурсы хранятся децентрализованно и могут изменяться в произвольный момент времени, а также с большими размерами сети и различными пропускными способностями каналов. В то же время поиск должен выполняться достаточно эффективно по времени, что не дает возможности опрашивать все репозитории. Для решения этой задачи используется методика маршрутизации запросов, основанная на обмене описателей коллекций и индексов.

Индексирование. Для хранимых ресурсов строится атрибутно-полнотекстовый индекс, с помощью которого можно искать как полнотекстовую информацию (например, в текстовых полях описаний), так и выбирать ресурсы по значению их атрибутов. Индекс

хранится в реляционной СУБД и обновляется по заданному расписанию. Кроме непосредственно поиска индекс используется для формирования описателей коллекций, которые основаны на частотных характеристиках встречающихся в индексе термов.

Описатели коллекций предназначены для определения степени релевантности репозитория какому-либо поисковому запросу и применяются при маршрутизации запросов.

Протокол поиска. Стандартизованный поисковый протокол необходим как для совместного поиска, так и для предоставления возможности поиска внешним системам. Для ИСИР был выбран протокол Simple Digital Library Interoperability Protocol (SDLIP) благодаря его ориентированности на XML, относительной простоте и большим возможностям (в частности, “поиск в найденном”). Протокол SDLIP может работать на основе различных транспортных протоколов. В стандарте есть примеры использования SDLIP поверх HTTP и CORBA. В ИСИР SDLIP используется поверх SOAP, он оформлен в виде web-сервиса.

Заключение

Рассмотрены вопросы интеграции информационных ресурсов РАН, информационной поддержки научных исследований в рамках ЕНИП РАН. Обсуждены вопросы формирования наборов элементов метаданных и онтологий для научных информационных ресурсов. Схемы метаданных являются ключевой компонентой в организации ЕНИП РАН. Ввиду большого объема здесь мы не приводим описания и рекомендации по применению схем, полное описание будет опубликовано отдельно. Проанализированы возможности технологии ИСИР, ее применения для обеспечения интеграции информационных ресурсов РАН. Сопоставление требований ЕНИП РАН и возможностей кросс-платформенной технологии ИСИР позволяет сделать вывод, что проблемы ЕНИП РАН в существенной степени могут быть решены с помощью ИСИР.

Список литературы

- [1] EUROPEAN Research Gateways Online. <http://www.cordis.lu/ergo>
- [2] LAITINEN, Sauli; Sutela Pirjo & Tirronen, Kerttu, Development of Current Research Information Systems in Finland. Proc. of CRIS-2000.
- [3] DUBLIN Core Metadata Initiative. <http://dublincore.org/>
- [4] Бездушный А.Н., Жижченко А.Б., Каленов Н.Е. и др. Предложения по наборам метаданных для научных информационных ресурсов ЕНИП РАН // Шестая Всерос. науч. конф. “Электронные библиотеки: перспективные методы и технологии, электронные коллекции” — RCDL’2004. Пушцино, 2004.
- [5] НЕСТЕРЕНКО А.К., СЫСОЕВ Т.М., БЕЗДУШНЫЙ А.А. и др. Интеграция распределенных данных на основе технологий Semantic Web и рабочих процессов // Там же.
- [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] DCMII Agents Working Group. <http://dublincore.org/groups/agents/>
- [11] PRISM: Publishing Requirements for Industry Standard Metadata. <http://www.prismstandard.org/>
- [12] NATIONAL Archives of Australia — AGLS. http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html
- [13] BIBLINK Project. <http://hosted.ukoln.ac.uk/biblink/>
- [14] BIBTEX Definition in Web Ontology Language (OWL) Version 0.1. Working Draft, 2004. <http://visus.mit.edu/bibtex/0.1/>
- [15] MATH-NET Schemes. <http://www.iwi-iuk.org/material/RDF/1.1/>
- [16] UKOLN Research Support Libraries Programme Collection Description. <http://www.ukoln.ac.uk/metadata/rslp/schema/>
- [17] REPRESENTING vCard Objects in RDF/XML. W3C Note 22 February 2001. <http://www.w3.org/TR/vcard-rdf>
- [18] FOAF Vocabulary Specification. Namespace Document 1 May 2004. <http://xmlns.com/foaf/0.1/>
- [19] БЕЗДУШНЫЙ А.А., БЕЗДУШНЫЙ А.Н., ЖИЖЧЕНКО А.Б. и др. RDF схема метаданных ИСИР // Сб. науч. тр. X научно-практ. семинара “Новые технологии в информационном обеспечении науки”. М., 2003. С. 141–159.
- [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] AKT Reference Ontology. <http://www.aktors.org/publications/ontology/>
- [23] CERIF: Common European Research Information Format. <http://www.cordis.lu/cerif/src/about.htm>
- [24] THE CIDOC Conceptual Reference Model. <http://cidoc.ics.forth.gr/>
- [25] ВЕЖНЕВЕЦ А.А., БЕЗДУШНЫЙ А.Н., СЕРЕБРЯКОВ В.А., ЦЫГАНОВ С.А. О реализации систем поддержки применения результатов фундаментальных исследований // Шестая Всерос. науч. конф. “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”. Пущино, 2004.
- [26] ЛАВРЕНОВА О.А., ВЕЖНЕВЕЦ А.А. Структура и реализация электронной библиотеки диссертаций в РГБ // Сб. докл. Второй Всерос. науч. конф. “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”. СПб., 2003.
- [27] АДЖИЕВ А.С., БЕЗДУШНЫЙ А.Н., КОНОВАЛОВ С.П., СЕРЕБРЯКОВ В.А. Общероссийский WEB-портал математических ресурсов // Там же.

- [28] АЛЕКСЕЕВ А.Н., СОЗЫКИН А.В., МАСИЧ Г.Ф., БЕЗДУШНЫЙ А.Н. Подсистема проведения конференций и ее метаданные // Электронный журнал, посвященный созданию и использованию электронных библиотек. Т. 7, вып. 2. М.: Ин-т развития информационного общества, 2004.
- [29] КОРОТЧЕНКО Р.А., ЯРОЩУК И.О., БЕЗДУШНЫЙ А.Н. Версия схемы метаданных экспериментальных исследований с приложением в гидроакустике // Электронный журнал, посвященный созданию и использованию электронных библиотек, Т. 7, вып. 1. М.: Ин-т развития информационного общества, 2004.
- [30] КОТОВ А.В., СЕРЕБРЯКОВ В.А., СТОЛЯРОВ А.А. Информационная система “Востоковедение” // Сб. докл. Третьей Всерос. науч. конф. “Электронные библиотеки: перспективные методы и технологии, электронные коллекции”. Петрозаводск, 2001.
- [31] НГУЕН М.Х., АДЖИЕВ А.С. Описание и использование тезаурусов в информационных системах, подходы и реализация // Электронный журнал, посвященный созданию и использованию электронных библиотек, Т. 7, вып. 1. М.: Ин-т развития информационного общества, 2004.
- [32] EXPRESSING Qualified Dublin Core in RDF / XML. <http://dublincore.org/documents/dcqrdf-xml/>
- [33] BEZDUSHNYI A.N., ZHIZHCHENKO A.B., KULAGIN M.V., SEREBRYAKOV V.A. Integrated information resource system of the russian academy of sciences and a technology for developing digital libraries // Programming and Computer Software. 2000. Vol. 26, N 4. P. 177–185.
- [34] БЕЗДУШНЫЙ А.А., БЕЗДУШНЫЙ А.Н., НЕСТЕРЕНКО А.К. и др. Архитектура RDFS-системы. Практика использования открытых стандартов и технологий Semantic Web в системе ИСИР // Пятая Всерос. науч. конф. “Электронные библиотеки: перспективные методы и технологии, электронные коллекции” — RCDL’2003. СПб., 2003. <http://rcdl2003.spbu.ru/proceedings/J1.pdf>
- [35] БЕЗДУШНЫЙ А.А., БЕЗДУШНЫЙ А.Н., НЕСТЕРЕНКО А.К. и др. RDFS как основа среды разработки цифровых библиотек и Web-порталов // Рос. науч. электронный журн. “Цифровые библиотеки”. 2003. Т. 6, вып. 3. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part 3/BBNSS>
- [36] БЕЗДУШНЫЙ А.А., НЕСТЕРЕНКО А.К., СЫСОЕВ Т.М. и др. Архитектура и технологии RDFS-среды разработки цифровых библиотек и Web-порталов // Рос. науч. электронный журн. “Цифровые библиотеки”. 2003. Т. 6, вып. 4. <http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2003/part 4/BNSBS>
- [37] БЕЗДУШНЫЙ А.Н., КОВАЛЕВ Д.А., СЕРЕБРЯКОВ В.А. Архитектура сервисов интегрированной системы информационных ресурсов (ИСИР) // Рос. науч. электронный журн. “Цифровые библиотеки”. 2002. Т. 5, вып. 1. <http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2002/pa rt1/BKS>
- [38] СЫСОЕВ Т.М., НЕСТЕРЕНКО А.К., БЕЗДУШНЫЙ А.А. и др. О реализации службы управления содержанием // Рос. науч. электронный журн. “Цифровые библиотеки”. 2003. Т. 6, вып. 6. <http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2003/pa rt6/SNBBS>

- [39] НЕСТЕРЕНКО А.К., БЕЗДУШНЫЙ А.А., СЫСОВЕВ Т.М. и др. Служба управления потоками работ по манипулированию ресурсами репозитория // Рос. науч. электронный журн. "Цифровые библиотеки". 2003. Т. 6, вып. 5.
[http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2003/part 5/NBSBS](http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2003/part%205/NBSBS)
- [40] MIA Development: Architecture and Functional Model, Tracy Gardner, UKOLN, University of Bath.
- [41] АРАСНЕ Cocoon Project. <http://cocoon.apache.org>
- [42] ВЕЖНЕВЕЦ А.А., БЕЗДУШНЫЙ А.Н. Вопросы построения информационного портала поддержки использования результатов фундаментальных исследований // Рос. науч. электронный журн. "Цифровые библиотеки". 2003. Т. 6, вып. 6.
[http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2003/par t6/VB](http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2003/par%20t6/VB)
- [43] БЕЗДУШНЫЙ А.А., БЕЗДУШНЫЙ А.Н., НЕСТЕРЕНКО А.К. и др. Возможности технологий ИСИР в поддержке единого научного информационного пространства РАН // Шестая Всерос. науч. конф. "Электронные библиотеки: перспективные методы и технологии, электронные коллекции". Пущино, 2004.
- [44] БЕЗДУШНЫЙ А.Н., КОВАЛЕВ Д.А., СЕРЕБРЯКОВ В.А. Технологии репликации данных и распределенного поиска в ИСИР РАН // Сб. докл. Третьей Всерос. науч. конф. "Электронные библиотеки: перспективные методы и технологии, электронные коллекции", Петрозаводск, 2001.
- [45] RDF Primer. W3C Recommendation 10 February 2004. <http://www.w3.org/TR/rdf-primer/>
- [46] КУРИВ П.М., КОТЕРОВ Д.В., КАЛЕНОВ Н.Е. Архитектура и функциональность библиотечной подсистемы ИСИР РАН // Электронный журнал, посвященный созданию и использованию электронных библиотек. Т. 7, вып. 1. М.: Ин-т развития информационного общества, 2004.
- [47] W3C Semantic Web Activity. <http://www.w3.org/2001/sw/>

Поступила в редакцию 18 марта 2005 г.