

А. АНТОПОЛЬСКИЙ, А. КАЛЕНКОВА, Н. КАЛЕНОВ, В. СЕРЕБРЯКОВ, А. СОТНИКОВ

## ПРИНЦИПЫ РАЗРАБОТКИ ИНТЕГРИРОВАННОЙ СИСТЕМЫ ДЛЯ НАУЧНЫХ БИБЛИОТЕК, АРХИВОВ И МУЗЕЕВ\*

### Введение

В настоящее время основным направлением разработки информационных систем публичного назначения является их интеграция, что обусловлено, прежде всего, быстрым развитием общего для всех пространства интернета. В сфере научного и культурного наследия актуальной становится идея интеграции в едином информационном пространстве информационных систем различных институтов памяти (библиотек, архивов и музеев). Цель этого направления - **обеспечить навигацию пользователя и сквозной поиск объектов научного и, шире, культурного наследия в фондах различных институтов памяти**. Причина - существенное пересечение хранения и обработки объектов культурного наследия в институтах памяти. Сведения, или артефакты (в том числе документы), относящиеся к персоне, событию, учреждению, территории или другой сущности, можно найти в ресурсах разных институтов памяти.

Общая постановка задачи интеграции библиотек, архивов и музеев, а также обзор некоторых методов интеграции, применяемых в России и за рубежом, представлены в работе [1].

Эту идею взял на вооружение коллектив разработчиков электронной библиотеки «Научное наследие России». Интегрированная система рассматривается нами как новый этап развития этой электронной библиотеки, что, конечно, не препятствует использованию создаваемой Интегрированной системы в других приложениях.

Описанное в настоящей статье исследование, проводимое при поддержке РФФИ (грант № 11-07-00331-а), предполагает разработку экспериментальной модели Интегрированной системы, объединяющей АИС нескольких участников ЭБ «Научное наследие России», а именно: БЕН РАН, ГПНТБ России, Архива РАН (АРАН), Государственного Дарвиновского музея (ГДМ). Естественно, в перспективе Интегрированная система должна иметь возможность подключения новых участников.

Главным препятствием к интеграции является различие в принципах, целях и задачах сбора и хранения объектов научного наследия в разных институтах памяти, что выражается в различных способах организации фондов и, как следствие, в принципах учета и каталогизации этих объектов. В свою очередь эти принципы реализуются в виде моделей метаданных, существенно различающихся во внутренних АИС библиотек, архивов и музеев. Ниже приводятся исходные модели метаданных, используемые участниками проекта.

### Состав метаданных исходных АИС

#### Состав метаданных библиотечной АИС БЕН РАН:

- тип объекта - «издание»;
- вид издания (монография, статья, диссертация и т.п.);
- название издания;
- ссылки на персон, интеллектуально связанных с изданием, с указанием вида связи (автор, художник, редактор, составитель, член редколлегии, переводчик);
- ссылки на коллективы (если имеются коллективные авторы);
- год издания;
- библиографическое описание;
- ссылка на персону как предмет публикации;

- ссылка на коллектив как предмет публикации;
- ссылка на событие как предмет публикации;
- ссылка на территорию как предмет публикации;
- тематика (УДК, ГРНТИ);
- URL;
- шифр хранения оригинала;
- ключевые слова (аннотация);
- ссылка на вышестоящий объект (для статей - на выпуск журнала - на описание журнала).

#### Состав метаданных ГПНТБ России

В АБИС ГПНТБ имеется действующая система экспорта метаданных из внутреннего формата АИС в коммуникативные форматы на базе Дублинского ядра, RUSMARC, MODS с использованием синтаксиса xml.

#### Состав метаданных ГДМ

В АИС ГДМ используются метаданные для трех типов объектов: естественно-научные коллекции, художественные коллекции, редкая книга.

#### Естественно-научные коллекции:

Название русское  
Название латинское  
Иные названия

Систематика (царство, тип (у растений отдел), класс, отряд (у растений порядок), семейство, род, вид)

Пол

Стадия развития

Абберация

Дата сбора

Место сбора (GPS, Страна, Регион, Область, ...)

**АНТОПОЛЬСКИЙ Александр Борисович** - доктор технических наук, зав. лабораторией Института научной информации и мониторинга (ИНИМ) РАН.  
Адрес: 142432, Московская обл., Ногинский р-он, пос. Черноголовка, Школьный б-р, 1  
e-mail: ale5695@yandex.ru

**КАЛЕНКОВА Анна Алексеевна** - младший научный сотрудник учреждения Российской академии наук Вычислительный центр им. А.А. Дородницына РАН (ВЦ РАН).  
Адрес: 119333, г. Москва, ул. Вавилова, 40  
e-mail: akalenkova@ultimeta.ru

**КАЛЕНОВ Николай Евгеньевич** - доктор технических наук, профессор, директор Библиотеки по естественным наукам РАН (БЕН РАН).  
Адрес: 119991, г. Москва, ул. Знаменка, 11/11  
e-mail: head@benran.ru

**СЕРЕБРЯКОВ Владимир Алексеевич** - доктор физико-математических наук, профессор, зав. отделом систем математического обеспечения ВЦ РАН.  
Адрес: 119333, г. Москва, ул. Вавилова, 40  
e-mail: serebr@ccas.ru

**СОТНИКОВ Александр Николаевич** - доктор физико-математических наук, профессор, зам. директора Межведомственного суперкомпьютерного центра РАН (МСЦ РАН).  
Адрес: 119991, г. Москва, Ленинский проспект, 32 а  
e-mail: ASotnikov@jssc.ru

\* Работа поддержана грантом РФФИ № 11-07-00331-а.

Персоналия(и) (*Коллектор, Препаратор, Владелец, Организация...* в разных комбинациях)

Коллекция (принадлежность к группе)

Описание

Изображение(я)

*Художественные коллекции:*

Название

Комментарий

Материал

Техника

Дата

Персоналия(и)

Коллекция (принадлежность к группе)

Описание

Изображение(я).

*Редкая книга:*

Заглавие

Перевод заглавия

Год издания

Место издания

Издательство

ББК

Номер тома, части

Серия

Персоналия(и)

Описание

Изображение(я).

## Состав метаданных Архива РАН

В проектируемой системе Архива РАН используется система метаданных, базирующаяся на международном стандарте ISAD(G):

1. Код(ы) идентификации

2. Заглавие

3. Дата(ы)

4. Уровень описания:

1. Фонд (Fonds)

2. Опись (Series)

3. Дело (File)

4. Единица описания (Item)

5. Объем и носитель хранения единицы описания (количество, объем, размер)

6. Имя создателя(ей)

7. Административная / Биографическая история

8. Архивная история

9. Непосредственный источник комплектования или перевод

10. Рамки и содержание

11. Оценка, уничтожение и плановая информация

12. Дальнейшие поступления

13. Система расположения и организации материала

14. Условия, регламентирующие доступ

15. Условия, регламентирующие воспроизводство

16. Язык / графика материала

17. Физическая характеристика и технические требования

18. Научно-справочный аппарат

19. Наличие и местоположение оригиналов

20. Наличие и местонахождение копий

21. Связанные единицы описания

22. Примечания о публикациях

23. Примечание

24. Примечание архивиста

25. Правила, по которым составлено описание

26. Дата(ы) описания.

Нужно иметь в виду, что все 26 элементов ISAD(G) доступны для использования, но для отдельного описания необходимы лишь некоторые из них. И совсем небольшое количество элементов считаются основными при международном обмене описательной информацией:

а) код идентификации;

б) заглавие;

в) создатель;

г) дата(ы);

д) объем единицы описания;

е) уровень описания.

## Принципы интеграции

Проект направлен на создание модели Интегрированной системы, являющейся надстройкой над существующими технологическими системами, обеспечивающей поиск объектов в распределенной среде. При этом Интегрированная система работает с метаданными, формируемыми в результате конвертирования информации из внутренних систем. АИС библиотек, музеев, архивов. Такой подход позволяет не заниматься переработкой внутренних АИС участников в процессе интеграции. С другой стороны, поисковые возможности АИС определяются исходными метаданными. Иной подход требует интеллектуальной переработки имеющихся описаний информационных объектов в исходных АИС, что обычно неприемлемо по ресурсным соображениям.

Поэтому предлагается построить онтологию Интегрированной системы, объединяющую модели метаданных участников. Онтология Интегрированной системы включает согласованный перечень сущностей:

- публикация;
- единица архивного описания;
- персона;
- организация (коллектив);
- проект;
- физический объект (музейный экспонат);
- коллекция;
- событие.

Каждая сущность имеет установленное множество метаданных, в том числе подмножество обязательных элементов. Любой информационный объект, доступный в Интегрированной системе, отнесен к одной из сущностей, а также имеет связи с другими информационными объектами.

Среди метаданных имеются элементы, значения которых нормируются словарями. Совокупность словарей Интегрированной системы будем называть ее тезаурусом.

Для представления онтологии используется язык OWL.

Состав метаданных Интегрированной системы (как в целом, так и по отдельным сущностям) определяется, исходя из анализа запросов, допустимых в системе.

Предлагается на первоначальном этапе ограничить виды запросов, на обработку которых ориентирована Интегрированная система, следующими видами:

1. Кто работал [по заданной тематике] [в заданное время] [на заданной территории]. Выдается список персон с указанием дат и мест рождения и смерти, от каждой из которых имеются ссылки на связанные организации, события, библиотечные, архивные и музейные объекты. При переходе по ссылкам предоставляется подробная информация об объекте из соответствующей системы по ее правилам.

2. Что имеется в Интегрированной системе [по данной тематике - выбор из тезауруса] [данному событию - выбор из тезауруса] [в заданное время - эра, век, интервал годов] [на заданной территории - выбор из тезауруса]. Выдается список объектов с возможностью просмотра (в системе) связанных персон, коллективов (организаций), событий и подробной информации об объектах из соответствующих систем.

3. Что имеется в Интегрированной системе, относящееся к данной персоне. Выдается список объектов со ссылками на системы.

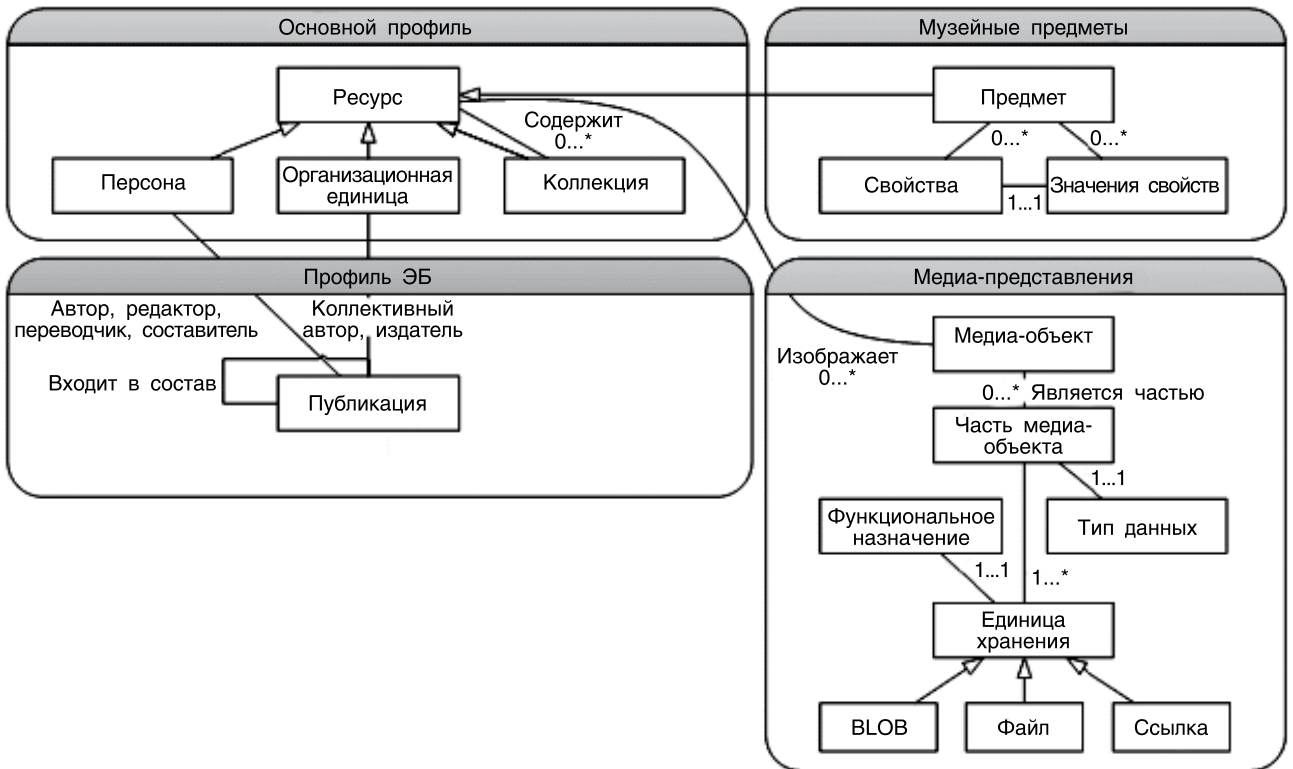


Рис. 1. Схема профилей метаданных СУЭБ LibMeta

4. Что имеется в Интегрированной системе, относящееся к данному коллективу. Выдается список объектов со ссылками на системы.

5. Лексический поиск информационных объектов и их атрибутов, представленных в Интегрированной системе на естественном языке (прежде всего - наименований объектов, описаний и аннотаций).

Метаданные формируются из информации, имеющейся в исходных АИС. Основой формирования метаданных являются описания информационных объектов, имеющиеся в библиотеках, музеях, архивах. При этом основным информационным объектом для библиотеки будет описание единицы хранения для книг и публикации - для периодики, а для музея - описание экспоната (единицы хранения). В архиве ситуация сложнее. Там более или менее подробное описание составляется на укруп-

ненную коллекцию (архивный фонд), сокращенное описание на единицу хранения - архивное дело. На архивный документ описание составляется в редких случаях, но это также возможно. Основным вариантом для поиска в архивных описаниях будет все-таки поиск по фоновому каталогу. Каждый информационный объект имеет обязательные поля: название, тип объекта (публикация, музейный экспонат, архивный фонд), временные характеристики (время создания [и время обнаружения - для музейных объектов]), тематический(е) индекс(ы) и описание (свободный текст и (или) ключевые слова, по которому реализуется поиск), может иметь связи с персонами, коллективами (организациями), территорией, временным отрезком, событием. Соответственно в Интегрированной системе отражаются эти же сущности.

Как уже указывалось, основу информационного языка Интегрированной системы составляет онтология. В ее состав входит тезаурус, т.е. совокупность словарей и авторитетных файлов, отражающих синонимию и иерархическую подчиненность лексических единиц.

Интерфейс Интегрированной системы предоставляет возможность просмотра тезауруса и отдельных словарей, входящих в его состав, и включения выбранных терминов в запрос. Теоретически сущности, принятые в Интегрированной системе, должны отражаться в исходных системах, поисковые элементы из которых перегружаются в Интегрированную систему. Если данных в исходных системах нет, по согласованию участников они создаются и вводятся в Интегрированную систему. Аналогичный принцип принят в действующей информационной системе «Научного наследия России».

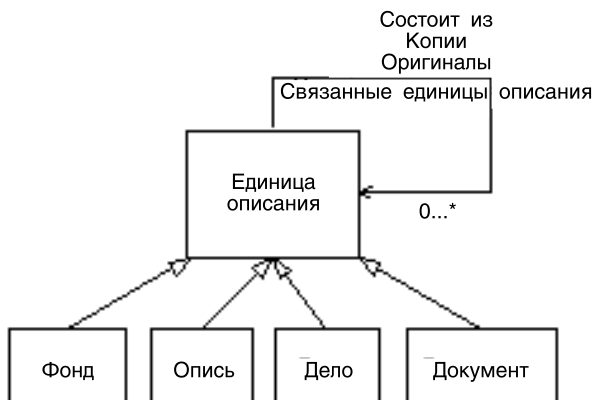


Рис. 2. Основные сущности профиля представления архивной информации

**Исходная онтология СУЭБ LibMeta**

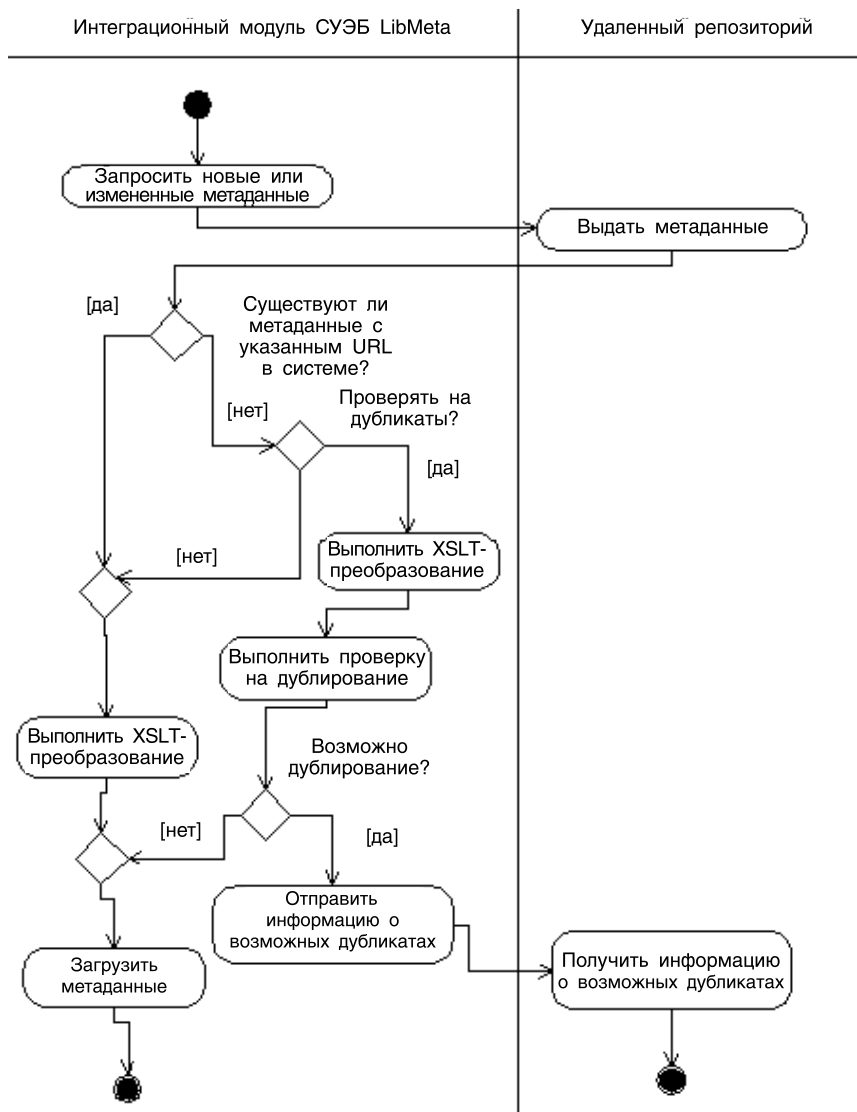
В основу онтологии Интегрированной системы предлагается положить OWL-онтологию Системы управления электронной библиотекой (СУЭБ) LibMeta, разработанную в ВЦ РАН на базе библиотечного профиля ЕНИП (Единое научное информационное пространство). В профиле метаданных ЕНИП для электронных библиотек используются ресурсы, такие как Организации, Персоны, Публикации. Общая схема профилей метаданных, применяемых в СУЭБ LibMeta, а также основных сущностей в данных профилях приведена на **рис. 1**. К основным типам данных, представленных в СУЭБ LibMeta, относятся Публикации, Персоны (авторы), Предметы. Кроме того, в СУЭБ LibMeta на основе стандарта ISAD разработан дополнительный профиль метаданных для представления архивной информации, основные сущности этого профиля показаны на **рис. 2**. Более подробное описание СУЭБ LibMeta приводится в работе [2].

Доработка СУЭБ LibMeta должна проводиться по следующим направлениям:

- Включение в метаописание музейного предмета, реквизитов, отсутствующих в описании СУЭБ LibMeta
- Реализация функций редактирования, просмотра и загрузки архивной информации.

**Интеграция СУЭБ LibMeta с другими информационными системами**

В СУЭБ LibMeta создан универсальный модуль загрузки метаданных в произвольном XML-формате в соответствии с протоколом OAI-PMH. Алгоритм получения метаданных некоторого ресурса, реализованный в этом модуле, представлен на **рис. 3**. С определенной периодичностью интеграционный модуль запрашивает вновь созданные или измененные метаданные из удаленного хранилища по протоколу OAI-PMH. В первую очередь проверяется URI (Unified Resource Identifier) получаемых метаданных. Если метаданные с указанным URI уже представлены в системе, то выполняется XSLT-преобразование (метаданные приводятся к внутреннему RDF/XML формату СУЭБ LibMeta) и производится загрузка в режиме «дозапись». При загрузке в режиме «Дозапись новых данных поверх существующих» для каждого свойства, загружаемого из RDF/XML, все прежние значения этого свойства стираются и заменяются на значения из RDF/XML. При этом значения тех свойств, которые были указаны в базе, но отсутствуют в RDF/XML, оставляются неизменными. Такой режим загрузки обеспечивает корректную инкрементную «дозапись» данных поверх существующих. Если метаданных с указанным URI в системе нет, то они являют-



**Рис. 3.** Алгоритм работы интеграционного модуля СУЭБ LibMeta

ся новыми и также должны быть загружены. Однако в силу того, что СУЭБ LibMeta представляет собой единый интеграционный узел, метаданные, соответствующие некоторому информационному ресурсу, могут быть получены ранее из другого источника. Для того чтобы в СУЭБ LibMeta не возникало дубликатов, используется вспомогательный модуль автоматической проверки на дубликаты. Этот модуль предоставляет возможность указания параметров проверки. Так, например, может быть указано, что два ресурса представляют одну и ту же публикацию, если у них совпадают названия и списки авторов. Если есть предположение о том, что загружаемые метаданные уже хранятся в системе, источнику метаданных отправляется информация о схожих метаданных (и их уникальных идентификаторах), находящихся в СУЭБ LibMeta. Источнику сообщается: какие метаданные были загружены, какие метаданные не были загружены в силу ошибок (в этом случае приводится описание ошибок загрузки), какие метаданные имеют сходство с некоторыми метаданными, уже находящимися в СУЭБ LibMeta, они не загружаются, и приводится список дубликатов и их URI. При возникновении подозрений на наличие дубликатов на стороне источника



определяется, соответствуют ли метаданные одному и тому же информационному ресурсу. Если принимается решение о том, что эти метаданные уже есть в системе, для них устанавливается URI уже загруженных метаданных (тогда при следующей загрузке метаданные в репозитории могут быть дополнены новыми значениями полей), во всех своих метаданных, ссылающихся на этот ресурс, также устанавливается в качестве значимых ссылок выбранный идентификатор. Иначе для загружаемых метаданных выставляется признак, что они должны быть загружены, несмотря на наличие схожих метаданных в СУЭБ LibMeta, и они попадают в систему при следующей загрузке без проверки на дублирование. Таким образом, интеграционный модуль СУЭБ LibMeta реализует некоторый общий подход к загрузке метаданных из удаленных репозиторий.

## Литература:

1. Вопросы интеграции библиотек, архивов и музеев по научному наследию // А.Б. Антопольский / Труды XV научно-практического семинара «Информационное обеспечение науки: новые технологии». - Таруса, 2011.

## Заключение

Итак, основными предметами разработки Интегрированной системы должны быть:

- общая онтология системы, включающая перечень сущностей, состав метаданных для каждой сущности и в основном описанная в настоящей статье;
- тезаурус, нормирующий лексику метаданных, а также задающий иерархические и синонимические связи между лексикой метаданных;
- средства ведения тезаурусов (онтологий) - ввод термина, установление связей, редактирование;
- интерфейс системы, задающий состав поисковых метаданных, средства их просмотра и выборки; программные средства поиска в тезаурусе, формирования запроса на поиск, навигации по ресурсам.

2. Интеграционные возможности Системы управления электронными библиотеками LibMeta // А.А. Каленкова, В.А. Серебряков / Труды XIV Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2011), Санкт-Петербург, Россия, 2011. - С. 81-85.

## НАША ИНФОРМАЦИЯ

### Светодиодные светильники челябинцев освещают город и магистрали

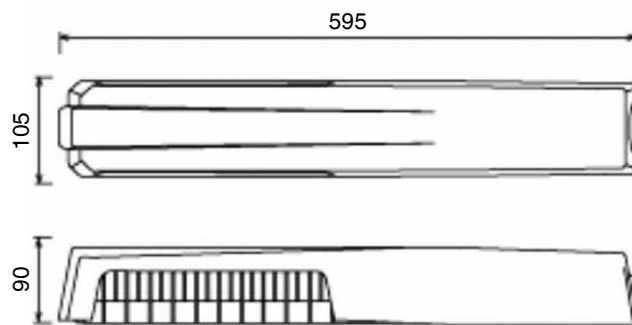
350 модульных светодиодных светильников серии LAD LED R320, созданных челябинским ООО «Лайт Аудио Дизайн», хорошо себя зарекомендовали в процессе эксплуатации на улицах и площадях г. Магнитогорска в осенне-зимний период. Недавно с успехом завершен проект оснащения этими светильниками проходящей через г. Братск новой автодороги «Виллю».

#### Характеристики светильника LAD LED R320-1-MG-50 консоль:

- замена ДНаТ 150, ДРЛ 250;
- экономия электроэнергии в 2-5 раз;
- успешно эксплуатируется в сложных климатических условиях;



- обладает высокой механической прочностью, виброустойчивостью, влаго- и пылезащищенностью, устойчивостью к воздействию температур и коррозии;
- не требует обслуживания;
- количество диодов: 21 шт. серии XP-G (CREE, США);
- световой поток: 6 875 Lm;
- угол раскрытия луча: 60°x120°;
- цветовая температура: 4000 К~4500 К;
- коэффициент цветопередачи: Ra ≥ 70;
- коэффициент мощности: ≥ 0,95;
- напряжение питания: AC 90-265 V, DC 127-370 V;
- энергопотребление: не более 55 W;
- влагозащита: IP 65;
- температура окружающей среды: -50°... +45°;
- алюминиевый корпус с кожухом из ударопрочного полистирола;
- крепление: на консоли (труба от Ø 48 мм до Ø 54,5 мм);
- габаритные размеры (мм): 595\*105\*90;
- вес: 2,95 кг.



Подробная информация на сайте:  
[www.ladesign.ru](http://www.ladesign.ru).