

**ВСЕРОССИЙСКИЙ ИНСТИТУТ АГРАРНЫХ ПРОБЛЕМ И  
ИНФОРМАТИКИ ИМЕНИ А.А. НИКОНОВА –  
ФИЛИАЛ ФГБНУ ФНЦ АГРАРНОЙ ЭКОНОМИКИ И СОЦИАЛЬНОГО  
РАЗВИТИЯ СЕЛЬСКИХ ТЕРРИТОРИЙ – ВНИИЭСХ**

A stylized globe with glowing blue and white lines representing global connectivity or data flow. The globe is set against a light blue background with starburst effects.

**Методы и модели интеграции  
аграрных научных информационных  
ресурсов в Интернете**

Москва - 2018

**ФЕДЕРАЛЬНОЕ АГЕНТСТВО НАУЧНЫХ ОРГАНИЗАЦИЙ  
ВСЕРОССИЙСКИЙ ИНСТИТУТ АГРАРНЫХ ПРОБЛЕМ И  
ИНФОРМАТИКИ ИМЕНИ А.А. НИКОНОВА – ФИЛИАЛ  
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО  
БЮДЖЕТНОГО НАУЧНОГО УЧРЕЖДЕНИЯ «ФЕДЕРАЛЬНЫЙ  
НАУЧНЫЙ ЦЕНТР АГРАРНОЙ ЭКОНОМИКИ И СОЦИАЛЬНОГО  
РАЗВИТИЯ СЕЛЬСКИХ  
ТЕРРИТОРИЙ - ВСЕРОССИЙСКИЙ НАУЧНО-  
ИССЛЕДОВАТЕЛЬСКИЙ  
ИНСТИТУТ ЭКОНОМИКИ СЕЛЬСКОГО ХОЗЯЙСТВА»**

## **"Методы и модели интеграции аграрных научных информационных ресурсов в Интернете"**

Меденников В.И., Луппов В.В., Личман А.А.,  
Муратова Л.Г., Сальников С.Г., Тухина Н.Ю.

Москва 2018

УДК 002:338.43(470)  
ББК 32.81:65.32-55(2Рос)  
М54

**"Методы и модели интеграции аграрных научных информационных ресурсов в Интернете", 2018. – 101 с.**

Рассмотрена общая проблема формирования единого информационного Интернет-пространства аграрных знаний для АПК. Описаны подходы к проблеме интеграции информационных ресурсов в единой базе данных из разрозненных гетерогенных баз данных, размещенных на сайтах различных агентств и научно-исследовательских организаций. Разработана база данных для представления информационных ресурсов и описан разработанный аппарат для сбора данных и размещения их в Интернете. В рамках портала [www.eipaz.ru](http://www.eipaz.ru) сформированы базы данных «Публикации научных работ» и «Прикладные разработки и инновационные проекты» научно-исследовательских учреждений аграрно-экономического направления.

Разделы книги написали:

**В.И. Меденников**, д.т.н. - введение, глава 1, раздел 2.1 заключение;

**В.В. Луппов**, к.т.н. – разделы 2.1, 2.2, 2.3, 2.5, 2.6;

**А.А. Личман**, к.э.н. – глава 1, раздел 2.2;

**Л.Г. Муратова**, к.э.н. – глава 1, раздел 2.3;

**С.Г. Сальников**, к.ф.-м.н. – глава 1, раздел 2.4;

**Н.Ю. Тухина**, н.с. – раздел 2.1, редактирование, оформление.

В работе использованы материалы исследований отдела информатизации АПК ФГБНУ ВИАПИ имени А.А. Никонова.

**Рецензенты:**

доктор экономических наук, **Романенко И.А.**, ФГБНУ ВИАПИ имени А.А. Никонова  
кандидат экономических наук, **Евдокимова Н.Е.**, ФГБНУ ВИАПИ имени А.А. Никонова

ISBN 978-5-6040748-9-3

© Авторы, 2018

© ВИАПИ имени А.А. Никонова, 2018

## **СОДЕРЖАНИЕ**

|                                                                                                                           |            |
|---------------------------------------------------------------------------------------------------------------------------|------------|
| <b>ВВЕДЕНИЕ .....</b>                                                                                                     | <b>4</b>   |
| <b>1 АНАЛИЗ ТЕХНОЛОГИЙ ПРОЕКТИРОВАНИЯ БАЗ ДАННЫХ И ПРИМЕРЫ<br/>ИНТЕГРАЦИИ РАЗРАБОТАННЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ .....</b> | <b>7</b>   |
| 1.1 Анализ технологий и проблем проектирования баз данных в сельском хозяйстве .....                                      | 7          |
| 1.2 Примеры интеграции информационных ресурсов.....                                                                       | 20         |
| <b>2 СОЗДАНИЕ БАЗ ДАННЫХ АГРАРНЫХ ЗНАНИЙ «ПУБЛИКАЦИИ» И<br/>«РАЗРАБОТКИ» НИУ АГРАРНОГО ПРОФИЛЯ .....</b>                  | <b>37</b>  |
| 2.1 Источники информации для баз данных аграрных знаний.....                                                              | 37         |
| 2.2 Организация и логическая структура базы данных в ЕИПАЗ .....                                                          | 41         |
| 2.3 Информационная структура баз данных ЕИПАЗ.....                                                                        | 44         |
| 2.4 Сбор данных в интернете для формирования баз данных публикаций и разработок ..                                        | 67         |
| 2.5 Занесение данных в систему ЕИПАЗ для формирования баз данных публикаций и<br>разработок .....                         | 80         |
| 2.6 Работа с базами данных публикаций и разработок в ЕИПАЗ.....                                                           | 87         |
| <b>ЗАКЛЮЧЕНИЕ.....</b>                                                                                                    | <b>89</b>  |
| <b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....</b>                                                                             | <b>90</b>  |
| <b>ПРИЛОЖЕНИЕ 1 ЛОГИЧЕСКАЯ СТРУКТУРА БД ЕИПАЗ .....</b>                                                                   | <b>92</b>  |
| <b>ПРИЛОЖЕНИЕ 2 ИЗВЛЕЧЕНИЕ ДАННЫХ СО СТРАНИЦ САЙТОВ (ТЕКСТ<br/>ПРОГРАММЫ НА ЯЗЫКЕ R).....</b>                             | <b>95</b>  |
| <b>ПРИЛОЖЕНИЕ 3 ФРАГМЕНТ ФОРМЫ ЗАНЕСЕНИЯ ЗАПИСИ ПУБЛИКАЦИИ... ..</b>                                                      | <b>101</b> |

## **ВВЕДЕНИЕ**

В Российской Федерации еще только определяются направления развития цифровой экономики (ЦЭ) в соответствии с новыми возможностями Интернет-технологий, но отдельные тенденции этого процесса уже достаточно явно прослеживаются.

Ключевую роль в успехе ЦЭ будут определять глобальные организационно-технологические Интернет-платформы на федеральном и отраслевом уровнях. Основной Интернет-платформой должно стать единое информационное Интернет-пространство страны (ЕИИП РФ). На отраслевом уровне должны быть разработаны, в частности, такие платформы, как единое информационное Интернет-пространство научно-образовательных ресурсов (ЕИИП НОР) и единое информационное Интернет-пространство АПК (ЕИИП АПК), интегрированные в ЕИИП РФ.

В Российской Федерации значительная доля ВВП создается государственными корпорациями (или компаниями со значительной долей государственного участия). В связи с этим, создание ряда промышленных цифровых платформ может производиться как под руководством профильных министерств, так и государственных корпораций, при их тесном взаимодействии. При построении платформ ЦЭ предполагается фокусировать усилия на ключевых направлениях: транспорт, телекоммуникации, энергетика, обработка данных.

Считается, что целенаправленное построение ряда промышленных платформ ЦЭ с единой архитектурой и стандартами позволит в будущем построить единое цифровое пространство, объединяющее все индустрии и отрасли. Такой подход представляется наиболее целесообразным для России сегодня, но он не лишен недостатков:

- остальным отраслям, в том числе и АПК придется продолжать стихийный процесс самоорганизации цифрового пространства, после чего переход на разработанную платформу ЦЭ окажется очень дорогим;

- нет уверенности, что выделенные средства не будут использованы нецелевым образом, тем более, что эти средства будут изъяты у государственных корпораций;

- есть большие сомнения, что создание технологических платформ ЦЭ государственными корпорациями без единой концепции, архитектуры ее формирования приведет к их интеграции в дальнейшем.

Например, государство не смогло создать единую систему сбора, хранения и предоставления широкому кругу пользователей научных знаний, произведенных научным сообществом. В настоящее время эти знания размыты в различных базах данных, никак не связанных между собой. При этом тратятся огромные средства на их создание и поддержание, но лишь незначительная часть баз данных доступна для научной

общественности. Это стало возможным вследствие отстранения государства от рычагов регулирования информатизацией страны, а теперь и процессом цифровизации экономики и общества.

Отстранение стало очевидным, исходя из анализа архитектурных решений, применяемых проектировщиками баз данных научных информационных ресурсов. Что касается более сложных концептуальных интегрированных решений по ИС страны, реализация которых сулила значительную эффективность, то можно привести в пример проект ОГАС Глушкова В.М., отказ от которого отрицательно сказался на дальнейшей информатизации России, когда пошли по пути оригинального проектирования ИС, создав к настоящему времени большое количество разнородных, неинтегрированных ИС.

Из множества определений цифровой экономики, в наибольшей степени отвечающее тематике наших исследований следующее: “хозяйственная деятельность, в которой ключевым фактором производства являются данные в цифровом виде, обработка больших объемов и использование результатов анализа которых по сравнению с традиционными формами хозяйствования позволяют существенно повысить эффективность различных видов производства, технологий, оборудования, хранения, продажи, доставки товаров и услуг”, следует, что важнейшей научной проблемой становится исследование оценки эффективности использования информационных ресурсов в цифровой экономике.

Увеличение объемов производства, частоты обновления номенклатуры выпускаемой продукции и технологий, и, как следствие, технологической и научной информации; усложнение управления развивающимися регионами, производственными системами, непромышленной сферой привели к увеличению и усложнению информационных потоков, поэтому без цифровизации этих процессов дальнейшее развитие общества будет замедляться.

Западные эксперты единодушны в том, что информационные технологии не работают без выстраивания отношений между субъектами экономики и управления в целом. В настоящее время на Западе эволюция информационных технологий привела к тому, что эти технологии становятся частью системы управления предприятием, т.е. средством координации действий людей. Именно это срастание информационной технологии и технологии управления заставило пересмотреть идеологию, технологию и организацию управления предприятиями и облечь их в виде стандартов.

Аналогично, интеграция как информационных ресурсов (ИР), так и информационных систем (ИС) достигается путем формирования стандартов на представление ИР, функции управления, подобно международным стандартам управления MRPII, ERP, CSRP, а также

путем комплексного подхода к проектированию, разработке и внедрению ИС, подготовке соответствующих специалистов.

Анализ сайтов информационно-консультационной службы, НИУ РАН, сельскохозяйственных ВУЗов, других предприятий, занимающихся сельскохозяйственной тематикой позволил выделить семь видов представлений аграрных знаний, присутствующих в том или ином виде на этих сайтах: публикации, разработки, консультационная деятельность, нормативно-правовая информация, дистанционное обучение, пакеты прикладных программ, базы данных. Исходя из этого, под информационными научными ресурсами будем понимать эти виды ИР.

Неотъемлемой составляющей частью интеграции информационных ресурсов является создание интерактивных баз данных в интернет-пространстве для обеспечения наиболее полной и актуальной информацией управленческих работников, участников производственных процессов, научных работников, преподавателей учебных заведений и обучающихся. В данной монографии рассматриваются вопросы проектирования и создания баз данных в виде электронных библиотек публикаций научных работ и прикладных разработок и инновационных проектов научно-исследовательских организаций аграрно-экономического направления. Базы данных созданы в рамках разработанного авторами специализированного кластера в Интернете Единого Информационного Пространства Аграрных Знаний (ЕИПАЗ) – портала [www.eipaz.ru](http://www.eipaz.ru).

Рассмотрены: общая проблема поиска и сбора информации в Интернете для наполнения баз данных, программные средства сбора информации и построения таблиц для переноса в базу данных портала, логическая структура базы данных портала с включением в нее реляционных таблиц публикаций и разработок, информационная структура сформированных баз данных портала, порядок работы с информацией портала.

# **1 АНАЛИЗ ТЕХНОЛОГИЙ ПРОЕКТИРОВАНИЯ БАЗ ДАННЫХ И ПРИМЕРЫ ИНТЕГРАЦИИ РАЗРАБОТАННЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ**

## **1.1 Анализ технологий и проблем проектирования баз данных в сельском хозяйстве**

В настоящее время каждая страна при переходе к цифровой экономике (ЦЭ) в зависимости от развития экономики, уровня образования, подготовленности нормативно-правовой базы, состояния и применяемых технологий разработки информационных систем делает упор на ту или иную стратегию цифровой трансформации для достижения положительных эффектов от этого. Наша страна еще только нащупывает направления развития ЦЭ в соответствии с новыми возможностями Интернет-технологий, однако отдельные тенденции этого процесса уже достаточно явно прослеживаются. Можно выделить следующие отличительные черты ЦЭ применительно к информатизации сельского хозяйства.

1 Ключевую роль в успехе ЦЭ при выработке решений, продуктов, сервисов и распределения создаваемой ими добавленной стоимости, будут определять глобальные организационно-технологические Интернет-платформы на федеральном и отраслевом уровнях. Основной Интернет-платформой должно стать единое информационное Интернет-пространство страны (ЕИИП РФ), представляющее собой типовые производственные, региональные, отраслевые и ведомственные информационные системы (ИС), порталы многоцелевой направленности, интегрированные между собой по формату данных, по классификаторам. Мероприятия по его формированию должны сопровождаться разработкой единой системы сбора и анализа статистической и учетной отчетности, разработкой унифицированных производственных типовых информационно-управляющих систем, информационно-вычислительных систем в науке и образовании, типовых информационно-управляющих систем для управления транспортными, логистическими, энергетическими и другими инфраструктурными системами. Это и была бы структура ЕИИП РФ, которая бы учитывала потребности и возможности большинства информационных систем и потребителей информационных ресурсов. На отраслевом уровне должны быть разработаны, в частности, такие платформы, как единое информационное Интернет-пространство научно-образовательных ресурсов (ЕИИПНОР) и единое информационное Интернет-пространство АПК (ЕИИП АПК), интегрированные в ЕИИП РФ. ЕИИПНОР представляет интеграцию в единой БД в некотором облаке у провайдера информации о разработках, публикациях, консультационной деятельности, нормативно-правовой информации, дистанционном обучении, пакетах прикладных программ, базах данных, разработанных НИУ РАН, ВУЗами, предприятиями и другими организациями, имеющими данные семь видов представлений научных знаний.



2 Отличительной чертой этого перехода является формирование открытого информационного общества, когда знания, информация становятся доступны для широкого круга пользователей. Например, Южная Корея считает, что для ускорения темпов роста экономики необходимо создать открытое общество, когда все инновации, почти вся информация должна быть открыта для всего общества. Особенно это стало необходимо с появлением и использованием Интернета, который дал возможность доступа неограниченного числа пользователей к различным информационным системам.

3 Понятно, что такая концепция потребует интеграции, как информационных ресурсов (ИР), так и ИС. Это достигается путем формирования стандартов на представление ИР, функции управления, подобно международным стандартам управления MRPII, ERP, CSRP, а также путем комплексного подхода к проектированию, разработке и внедрению ИС, подготовке соответствующих специалистов в рамках ЕИИП РФ. Одна из основных причин слабого внедрения информационных средств - это отсутствие целостной формализованной теории управления. Если бы такая теория была, то можно было бы, скажем, для вновь создаваемых или реорганизуемых предприятий предложить оптимальную систему управления, а через нее определить и информационно-управляющую систему (ИУС), наиболее ей удовлетворяющую. Указывая на отсутствие целостной организационной науки академик А.И. Берг отмечал, что из-за этого КПД общественного производства равен 2-4 %. Недавно академик РАН Александр Сергеев сказал об этом же следующее: «Физики пришли к выводу, что наука не может предсказать развитие сложных систем на достаточно больших временах эволюции. Они в принципе непредсказуемы! Во всем мире ученые поняли, что не стоит больше искать способов их предсказания — надо изменить свое отношение к проблеме. А именно найти способы ограничения динамического хаоса и компенсации действия непредсказуемых факторов». Относительно информатизации данную проблему можно выразить фразой «нельзя автоматизировать хаос».

4 Совершенствование программно-технических средств, удешевление их привело к тому, что потребность в них возникла у большинства организаций. А это уже потребовало создания программного обеспечения (ПО) в виде программного продукта на принципах типизации. В связи с этим стоит вспомнить один из методологических принципов создания информационных систем, сформулированный академиком В.М. Глушковым, относящийся к данной работе - принцип модульности и типизации. Этот принцип сводится к выделению и разработке максимально независимых частей системы или модулей и максимальному их использованию в различных подсистемах, в частности, в отраслях, где большое количество типовых организаций, например, в сельском хозяйстве, в образовательной деятельности

принцип модульности и типизации должен быть взят за основу при разработке ИС, в т.ч. и в Интернет.

5 Основные информационные ресурсы будут доступны для использования напрямую, без посредников, независимо от их физического расположения. Например, при реализации типовых сайтов сельскохозяйственных предприятий в рамках ЕИИП АПК информационно-консультационная служба (ИКС) должна исчезнуть или существенно видоизмениться, как информационный посредник, ввиду того, что вся информация ИКС и ее функции будут включены составной частью в общую информационную систему.

Рассмотрим проблему интеграции ИР и ИС, поскольку в сельском хозяйстве России данная проблема является довольно острой и в ближайшее время будет сдерживать цифровизацию экономики. Товаропроизводителю в настоящее время трудно найти разработки, публикации, прочую информацию по проблемам АПК, поскольку старая система распространения инноваций на бумажных носителях была разрушена, а новая на электронных не создана. Пока же в сельском хозяйстве доминирует «позадачный» метод разработки и внедрения программного обеспечения, когда приобретаются отдельные задачи у различных производителей, не связанные ни функционально, ни информационно.

Различные ситуации, требующие интеграции БД при информационном взаимодействии пользователей БД и информационного пространства (ИП), следующие.

1. Пользователи и ИП находятся территориально в одном месте и осуществляется новое проектирование ИС в виде автоматизированных рабочих мест (АРМ). В этом случае осуществляются традиционные этапы: анализ потоков данных; концептуальное проектирование; нормализация структур данных; логическое проектирование; физическое проектирование; разработка интерфейсов ввода-вывода; сопровождение БД.

2. Пользователи и ИП территориально распределены и осуществляется новое проектирование ИС. Этапы проектирования: анализ потоков данных, каналов передачи данных, ограничений (время, пропускная способность и т.п.); распределение вычислительной техники по местам размещения; распределение информации по узлам управления и техническим средствам; формирование логических структур распределенных баз данных; далее аналогично п.1.

3. Пользователи и ИП территориально распределены и осуществляется допроектирование новых ИС к существующим. После анализа существующих БД и требований, предъявляемых к новым может быть принято два варианта действий.

3.1. Принимается решение о полном перепроектировании всех существующих ИС в соответствии с п.2. Переносится информация из существующих ИС в объединенную новую БД.

3.2. В другом случае проектируется новая БД, учитывающая требования новых и старых ИС. Определяется проекция старой БД в новую и осуществляется синхронизация идентификаторов, преобразование форм передачи данных, преобразование форматов данных; решается проблема асинхронности передачи данных.

4. Наконец, рассмотрим ситуацию, когда ИП и пользователи территориально распределены, но неизвестны параметры ИС и БД (логические структуры, форматы данных и т.д.). Опять же две ситуации.

4.1. **БД однородные.** Например, это - библиотечные БД. В этом случае разрабатываются стандарты в области создания машиночитаемой каталогизации. Так появился международный формат MARC, который со временем разросся до 50 версий. Базовым форматом для них послужил USMARC (США). Появление такого количества форматов можно объяснить особенностями национальных правил каталогизации в разных странах. Для преодоления несовместимости MARC-форматов было принято решение о разработке формата-посредника, который получил название UNIMARC, то есть «Универсальный MARC». Для реализации процесса обмена каждой национальной службе достаточно было составить два конвертора: один для конвертирования в UNIMARC, другой – из UNIMARC [1]. К числу новых, разработанных на базе UNIMARC, можно отнести RUSMARC (Россия). Развитие Интернета привело к появлению XML – стандарта, метаязыка разметки, который обеспечивает простое структурное представление данных (создание собственно языков разметки), имеет весьма широкую область применения и играет все большую роль в информационном обмене. Начали говорить о переходе к XML от MARC-формата, даже об отказе от него, в силу следующих преимуществ [2]:

- **Гибкость:** в нем нет фиксированного набора полей и определенного синтаксиса, это позволяет пользователям создать свои собственные языки. Эта гибкость обеспечивает простую адаптируемость к стандартам каталогизации.
- Отсутствие жестких **ограничений** на размер записи.
- Возможность **включения** в запись других текстов, изображений, звуковых файлов.
- Широкое **распространение** вне библиотечного сообщества.

Однако, далее развитие пошло по пути разработки различных версий XML-схем для MARC-форматов. Это связано, с одной стороны, с огромными усилиями в создание библиотечных систем, с другой стороны, с незаконченностью процесса информационного обмена в Интернете (мир ждет прорыва в разработке универсального семантического поиска, что приведет к интеграции интеллектуального содержания библиотечных систем). В

частности, версия библиотеки Конгресса США, разработанная изначально для MARC21, послужила основой для разработки стандарта ISO 25577 – MarcXchange. Стандарт определяет требования к общей XML-схеме, для обмена как библиографическими записями, так и другими типами метаданных. Это «транспортная схема», которая может быть применена для любого MARC-формата. В стандарте не определяется содержание отдельных записей, не определяется значение конкретных меток, индикаторов, идентификаторов подполей. Описана общая структура, предназначенная для обеспечения обмена между системами обработки данных. Валидация содержания MARC-записей обеспечивается специальными программными средствами, ориентированными на конкретный MARC-формат. Общая схема ISO 25577 не предусматривает использование встроенных полей, для UNIMARCa (соответственно RUSMARCa) схема требует соответствующих дополнений. «Тонкая» XML-схема (RUSMARXML/Slim) для библиографических записей в формате RUSMARC разработана Центром компьютерных технологий Уральского Государственного университета им. А. М. Горького. Схема, соответствующая ISO 25577 и дополненная для UNIMARC, разрабатывается Постоянным комитетом по UNIMARC при участии Национальной службы развития системы форматов RUSMARC [3]. Хотя MARC-формат определяет и структуру, и семантику библиографической информации, прежде всего это стандарт структуры данных, а не стандарт содержания. Содержание записи регламентируются правилами каталогизации, системой предметизации и т.д. В то же время есть ряд вопросов, трактовка которых в формате и в правилах каталогизации различается. Один из таких спорных вопросов – разделение между содержанием записи и носителем. У участников Сводного каталога библиотек России возникает много проблем при описании электронных ресурсов, в частности, при идентификации описываемого ресурса как электронного / видеоматериала / звукозаписи. В RUSMARC (как, собственно говоря, и в UNIMARC, и в MARC21) принят следующий подход: при описании ресурса, который сочетает характеристики разных видов материала, в записи отражаются характеристики каждого вида материала для описания всех аспектов ресурса, включая и содержание, и физический носитель, при этом основным является именно содержание, а потом уже – носитель. При определении типа записи, который указывается в маркере (по сути – идентификация ресурса как электронного, видео, аудио и т.д.) отдается предпочтение коду вида каталогизируемого материала, а не коду вторичной физической формы (формат). В Российских правилах каталогизации принимается во внимание как форма воплощения, так и содержание ресурса, но первичным оказывается носитель.

Вообще, практика каталогизации нацелена прежде всего на документ как физический объект описания, а не на описание интеллектуального содержания. В результате во многих

случаях создаются десятки записей на документы с идентичным содержанием. Эту ситуацию призвана разрешить модель FRBR (Functional Requirements to Bibliographic Data – Функциональные требования к библиографическим записям), которая ориентирована прежде всего на содержание, на поддержку связей между произведениями, создатели модели предполагают, что ее применение позволит пользователю проще ориентироваться в каталоге. В настоящее время нельзя говорить, что модель FRBR инкорпорирована в MARC-формат или правила каталогизации, но работа в этом направлении ведется. Предпринимаются попытки реализовать модель на практике, для поиска в библиотечных системах. Первая попытка применения модели FRBR в правилах каталогизации предпринята создателями RDA (Resource Description and Access – Описание ресурсов и доступ к ним). RDA – свод правил, который призван прийти на смену AACR и разрабатывается в качестве нового стандарта содержания для описания ресурсов и обеспечения доступа к ним. Ключевые моменты создаваемого свода правил: (1) в основе RDA лежат концептуальные модели FRBR и FRAD (Functional Requirements to Authority Data – Функциональные требования к авторитетным данным); RDA ориентирован на объектно-ориентированную структуру баз данных; (2) проводится четкое различие между содержанием записи и ее представлением. Основным является именно содержание: «Задача – создать набор инструкций для записи данных, которые могут быть применены независимо от какой бы то ни было структуры или синтаксиса для хранения или вывода данных». Вывод (представление) записи будет рассмотрен в отдельном приложении, с указанием соответствий между элементами RDA и ISBD, для формирования привычного вывода библиографического описания (ISBD). Кроме того, отдавая должное широкому распространению MARC-форматов в библиотеках, предполагается включить приложение с инструкциями по представлению описательных элементов данных в MARC-формате. В марте 2008 г. Комитетом по разработке RDA под эгидой Британской библиотеки, Библиотеки Конгресса США, Библиотеки и архива Канады создана рабочая группа RDA/MARC, среди задач которой – разработать предложения для обсуждения с MARC-сообществом; определить, какие изменения необходимы в MARC21 для поддержки совместимости с RDA и эффективного обмена данными в будущем. В ближайших планах – обеспечить совместимость RDA и MARC-формата на уровне словарей. Таковы основные элементы информационной инфраструктуры. Для того, чтобы это была действительно единая инфраструктура, мы должны обеспечить реальное взаимодействие этих компонентов друг с другом. Кроме того, каждый компонент должен использоваться максимально эффективно.

Многие проблемы, которые, казалось бы, возникают при работе с MARC- записями, - это на самом деле проблемы не записей, а библиотечных систем. MARC- записи,

накопленные за время использования формата, представляют собой огромный массив стандартизированных данных. Задача библиотечных систем – максимально использовать это стандартизированное, структурированное содержание. Например, большинство элементов FRBR на самом деле включаются в записи уже на протяжении десятилетий, но только сейчас выяснено, что они были бы полезны для пользователя – помогли бы ему ориентироваться в каталоге. Как еще элементы MARC-записи могут облегчить работу с каталогом, сделать ее более удобной и эффективной? (а) Фасетная навигация на основе определителей содержания MARC – многомерный просмотр (browsing), отражающий связи между записями в массиве. Например, тематический поиск может дать тысячу записей. Получить более точный результат может уточнение по индексу, дате публикации, месту хранения и т.д. Если данные для формирования фасетов берутся из реальных записей – такое сужение никогда не приведет к нулевому результату. (б) Ранжирование результатов поиска по релевантности (приписывание различных весов терминам запроса в зависимости от того, в каком поле найден термин) (в) Различная сортировка результатов поиска по релевантности, дате издания, заглавию, автору, востребованности и популярности документа (насколько часто его ищут пользователи) и т.д. Очевидно, что реализация этих функций возможна только при условии достаточной информации в записи. В той или иной мере эти функции реализуются в таких проектах, как: FictionFinder (OCLC), RedLightGreen (проект Research Libraries Group, позднее вошел в каталог OCLC WorldCat, Endeca – web-интерфейс Библиотеки Университета Северной Каролины).

Надо сказать, что значение информации, накопленной в MARC-записях, признается не только библиотечным сообществом. Со стороны поисковых служб Интернета предпринимаются усилия к интеграции традиционного библиотечного содержания и служб в своих поисковых машинах. В качестве примеров можно назвать два проекта Google:

- Национальная служба развития системы форматов RUSMARC 8 1. Google LibraryThing – веб-приложение для создания каталогов библиотек любого уровня, хранения и обмена библиографическими записями. Основная функция LibraryThing – импорт данных от книготорговых организаций и библиотек, при этом библиотеки поставляют записи в MARC-формате и DC (Дублинское ядро);
- Google BookSearch – проект, направленный на создание всеобъемлющего виртуального каталога для поиска всех книг на всех языках при участии издательств и библиотек.

В мае 2008 г. подписано соглашение между Google и OCLC, в соответствии с которым, в частности, библиотеки-партнеры OCLC передают в BookSearch свои MARC-записи. Тем самым расширяются возможности доступа к библиотечным коллекциям через Интернет. Итак, современная библиографическая инфраструктура должна обеспечить быстрый и беспроблемный обмен различными типами данных. Требования, которым должна удовлетворять эта инфраструктура, многочисленны и разнообразны. Первичным является пользователь, его потребности и способность эти потребности удовлетворить:

- универсальность (многосторонность) – возможность работы с метаданными из различных источников, описывающими различные виды объектов, должна быть возможность принять запись от издателя, и использовать ее в качестве прототипа каталогизационной записи, или добавить информацию в запись, которая уже есть в каталоге;
- расширяемость – возможность включения новых стандартов метаданных и использования любого стандарта по выбору;
- открытость и прозрачность;
- поддержка иерархических структур;
- достаточный уровень детализации информации MARC-формат остается одним из компонентов этой библиографической инфраструктуры, наряду с другими схемами метаданных. При этом каждый компонент библиографической инфраструктуры должен выполнять свои функции, обслуживать интересы и потребности «своей» категории пользователей.

**4.2. БД имеют гетерогенные структуры.** В этом случае используются как накопленные в программировании методы, так и разрабатываются новые, например, семантические сети. Гетерогенность данных разделяется на физическую и семантическую [3]. Физическая гетерогенность подразумевает различия в представлении данных; может выражаться в различии типов данных, реализаций моделей данных (например, различные реляционные системы). Могут различаться также языки описания процедур, триггеров, языки запросов, манипулирования и определения данных. Кроме этого, могут различаться и сами модели данных. Данные могут быть представлены не только в БД, но и в электронных таблицах или почтовых файлах и т.д. Проблема физической гетерогенности решается посредством введения стандартов взаимодействия (ODBC, DAO, OLE DB, ADO, ADO.NET). Семантическая гетерогенность проявляется в различиях в наименованиях данных, значениях и логических структурах. В данной области проводится большое количество исследований, но до сих пор не появилось какого-либо стандарта. Хотя проблема гетерогенности данных находится в центре внимания крупнейших компаний-разработчиков программного

обеспечения и ведущих научных коллективов, она еще далека от удовлетворительного решения.

Рассмотрим более подробно современное состояние решения этой проблемы ([4]).

#### **4.2.1 Уровни интеграции данных.**

Системы интеграции данных могут обеспечивать интеграцию данных на физическом, логическом и семантическом уровне. Интеграция данных на физическом уровне с теоретической точки зрения является наиболее простой задачей и сводится к конверсии данных из различных источников в требуемый единый формат их физического представления. Интеграция данных на логическом уровне предусматривает возможность доступа к данным, содержащимся в различных источниках, в терминах единой глобальной схемы, которая описывает их совместное представление с учетом структурных и, возможно, поведенческих (при использовании объектных моделей) свойств данных. Семантические свойства данных при этом не учитываются. Поддержку единого представления данных с учетом их семантических свойств в контексте единой онтологии предметной области обеспечивает интеграция данных на семантическом уровне ([5]).

Процессу интеграции препятствует неоднородность источников данных, в соответствии с уровнем интеграции. Так, при интеграции на физическом уровне в источниках данных могут использоваться различные форматы файлов. На логическом уровне интеграции может иметь место неоднородность используемых моделей данных для различных источников или различные схемы данных при использовании одной и той же модели данных. Одни источники могут быть веб-сайтами, а другие — объектными базами данных и т. д. При интеграции на семантическом уровне различным источникам данных могут соответствовать различные онтологии. Например, возможен случай, когда каждый из источников представляет информационные ресурсы, моделирующие некоторый фрагмент предметной области, которому соответствует своя понятийная система, и эти фрагменты пересекаются.

#### **4.2.2 Возникающие задачи.**

При создании системы интеграции возникает ряд задач, состав которых зависит от требований к ней и используемого подхода. К ним, в частности, относятся:

- Разработка архитектуры системы интеграции данных;
- Создание интегрирующей модели данных, являющейся основой единого пользовательского интерфейса в системе интеграции;
- Разработка методов отображения моделей данных и построение отображений в интегрирующую модель для конкретных моделей, поддерживаемых отдельными источниками данных;



- Интеграция метаданных, используемых в системе источников данных;
- Преодоление неоднородности источников данных;
- Разработка механизмов семантической интеграции источников данных.

### 4.2.3 Архитектуры систем интеграции.

#### 4.2.3.1 Консолидация.

В случае консолидации данные извлекаются из источников, и помещаются в Хранилище данных. Процесс заполнения Хранилища состоит из трех фаз — извлечение, преобразование, загрузка (Extract, Transformation, Loading — ETL). Во многих случаях именно ETL понимают под термином «интеграция данных». Еще одна распространенная технология консолидации данных — управление содержанием корпорации (enterprise content management, сокр. ECM). Большинство решений ECM направлены на консолидацию и управление неструктурированными данными, такими как документы, отчеты и web-страницы.

Консолидация — однонаправленный процесс, то есть данные из нескольких источников сливаются в Хранилище, но не распространяются из него обратно в распределенную систему. Часто консолидированные данные служат основой для приложений бизнес-аналитики (Business Intelligence, BI), OLAP-приложений.

При использовании этого метода обычно существует некоторая задержка между моментом обновления информации в первичных системах и временем, когда данные изменения появляются в конечном месте хранения. Конечные места хранения данных, содержащие данные с большими временами отставания (например, более одного дня), создаются с помощью пакетных приложений интеграции данных, которые извлекают данные из первичных систем с определенными, заранее заданными интервалами. Конечные места хранения данных с небольшим отставанием обновляются с помощью оперативных приложений интеграции данных, которые постоянно отслеживают и передают изменения данных из первичных систем в конечные места хранения.

#### 4.2.3.2 Федерализация.

В федеративных БД физического перемещения данных не происходит: данные остаются у владельцев, доступ к ним осуществляется при необходимости (при выполнении запроса). Изначально федеративные БД предполагали создание в каждом из  $n$  узлов  $n-1$  фрагментов кода, позволяющего обращаться к любому другому узлу. При этом федеративные БД отделяли от медиаторов [6].

При использовании медиатора создается общее представление (модель) данных. Медиатор — посредник, поддерживающий единый пользовательский интерфейс на основе

глобального представления данных, содержащихся в источниках, а также поддержку отображения между глобальным и локальным представлениями данных. Пользовательский запрос, сформулированный в терминах единого интерфейса, декомпозируется на множество подзапросов, адресованных к нужным локальным источникам данных. На основе результатов их обработки синтезируется полный ответ на запрос. Используются две разновидности архитектуры с посредником — Global as View и Local as View [7].

Отображение данных из источника в общую модель выполняется при каждом запросе специальной оболочкой (wrapper). Для этого необходима интерпретация запроса к отдельным источникам и последующее отображение полученных данных в единую модель. Сейчас этот способ также относят к федеративным БД.

Интеграция корпоративной информации (Enterprise information integration, сокр. ЕИ) — это пример технологии, которая поддерживает федеративный подход к интеграции данных.

Изучение и профилирование первичных данных, необходимых для федерализации, незначительно отличаются от аналогичных процедур, требуемых для консолидации.

#### **4.2.3.3 Распространение данных.**

Приложения распространения данных осуществляют копирование данных из одного места в другое. Эти приложения обычно работают в оперативном режиме и производят перемещение данных к местам назначения, то есть зависят от определенных событий. Обновления в первичной системе могут передаваться в конечную систему синхронно или асинхронно. Синхронная передача требует, чтобы обновления в обеих системах происходили во время одной и той же физической транзакции. Независимо от используемого типа синхронизации, метод распространения гарантирует доставку данных в систему назначения. Такая гарантия — это ключевой отличительный признак распространения данных. Большинство технологий синхронного распространения данных поддерживают двусторонний обмен данными между первичными и конечными системами. Примерами технологий, поддерживающих распространение данных, являются интеграция корпоративных приложений (Enterprise application integration, сокр. ЕАИ) и тиражирование корпоративных данных (Enterprise data replication, сокр. EDR). От федеративных БД этот способ отличает двустороннее распространение данных.[4]

#### **4.2.3.4 Сервисный подход.**

Сервисно-ориентированная архитектура SOA (Service Oriented Architecture), успешно применяемая при интеграции приложений, применима и при интеграции данных. Данные

также остаются у владельцев. При запросе происходит обращение к определённым сервисам, которые связаны с источниками, где находится информация и ее конкретный адрес.

Интеграция данных объединяет информацию из нескольких источников таким образом, чтобы ее можно было показать клиенту в виде сервиса. Сервис — это не запрос в традиционном смысле обращения к данным, скорее, это извлечение некоторой бизнес-сущности (или сущностей), которое может быть выполнено сервисом интеграции через серию запросов и других сервисов. Подход SOA концентрируется, в первую очередь, на определении и совместном использовании в форме сервисов относительно ограниченного количества самых важных бизнес-функций в корпорации. Следовательно, сервис-ориентированные интерфейсы строятся на ограниченном количестве запросов на необходимую информацию, которую нужно представить потребителю.

Имея соответствующие учетные данные системы безопасности, потребитель может осуществить выборку любых данных из источника через почти неограниченное количество различных запросов SQL. Но для этого потребитель должен иметь представление о модели источника данных и способе создания результата с использованием этой базовой модели. Чем сложнее модель источника данных, тем более сложной может оказаться эта задача.[6]

#### **4.2.4 Проблемы интеграции информации.**

Вне зависимости от выбранных технологии и метода интеграции данных, остаются вопросы, связанные с их смысловой интерпретацией и различиями в представлении одних и тех же вещей. Именно, приходится преодолевать несоответствие схем данных [8] и несоответствие самих данных.

##### **4.2.4.1 Типы несоответствия схем данных:**

- Конфликты неоднородности (используются различные модели данных для различных источников);
- Конфликты именования (в различных схемах используется различная терминология, что приводит к омонимии и синонимии в именованиях);
- Семантические конфликты (выбраны различные уровни абстракции для моделирования подобных сущностей реального мира);
- Структурные конфликты (одни и те же сущности представляются в разных источниках разными структурами данных).

Структурные и семантические конфликты приводят к следующим проблемам:

- Различие в типах данных. Некоторый домен в одном источнике может представляться числом, в другом — строкой фиксированной длины, в третьем — строкой переменной длины.
- Различие в единицах измерения. В одной БД указана величина в сантиметрах, в другой — в дюймах. В этом случае существует отображение 1:1.
- Различие в множестве допустимых значений. Один и тот же признак может определяться разными наборами констант. Например, выполнение задания одним источником может оцениваться по четырехбальной шкале (неудовлетворительно, удовлетворительно, хорошо, отлично), другим — по трехбальной, третьим — по стобальной. Отображение не является 1:1, оно может быть многозначным, может не иметь обратного, может зависеть от сторонних данных (например, 30 по математике соответствовать «удовлетворительно», а по русскому языку — «неудовлетворительно»).
- Различие «домен-отношение». Домен в одной БД (например, строковое значение) соответствует таблице в другой БД (записи из таблицы-справочника).
- Различие «домен — группа доменов». В одном источнике адрес записывается одной строкой, в другом могут быть выделены отдельные поля для улицы, дома, строения, квартиры.
- Различие «данные-схема». Данные одной БД соответствуют схеме (метаданным) другой. В одной БД «инженер» — значение атрибута «должность» отношения «работник», в другой «инженеры» — отношение, содержащее некоторых работников, в то время как «бухгалтеры» содержит других.
- Отсутствующие значения. В каком-то из источников может отсутствовать информация, имеющаяся в большинстве других.

Разрешение этих несоответствий часто выполняется вручную. Обзор автоматических методов разрешения несоответствия схем можно найти в [9,10].

#### **4.2.4.2 Типы несоответствия собственно данных.**

##### **4.2.4.2.1 Различие формата данных:**

«ул. Бахрушина, 18-1» или «Бахрушина, д.18, стр.1»; «8(910)234-45-32» или «8-910-234-45-32».

##### **4.2.4.2.2 Различие в представлении значений.**

Например, некая организация может быть записана в отдельных источниках как «Новолипецкий металлургический комбинат», «НЛМК», «ОАО НЛМК».

#### 4.2.4.2.3 Потеря актуальности данных одним из источников.

Например, смена фамилии при замужестве: в одной БД записана новая фамилия, в другой старая, и они не совпадают.

4.2.4.2.4 Наличие ошибок операторского ввода (или ошибок распознавания бланков) в отдельных источниках данных. Сюда относятся механические опечатки, ошибки восприятия на слух сложнопроизносимых имен или названий, отсутствие единых стандартов транскрипции с иностранных языков.

4.2.4.2.5 Намеренное внесение искажений с целью затруднить идентификацию сущностей.

Перечисленные различия приводят к дублированию записей при интеграции данных в одну БД. Разрешение перечисленных проблем и устранение дублирования записей вручную практически невозможно. Имеется множество методов для ее автоматического и полуавтоматического решения. По-русски задача не имеет устоявшегося термина (применяются «сопоставление записей», «вероятностное соединение», «нестрогое соединение», «нестрогое соответствие»). В зарубежных работах эта задача носит название Identity resolution, или Record linkage (есть и другие синонимы).

## 1.2 Примеры интеграции информационных ресурсов

### 1.2.1 Современная система электронного взаимодействия в Эстонии

Разработанная и введенная в эксплуатацию система интеграция информационных ресурсов X-Road [11] обеспечивает гражданам право свободного доступа к своим данным. В системе x-Road используется ролевая модель доступа, представляющая архитектуру систем интеграции в виде федерализация (п. 4.2.3.2). Благодаря x-Road теперь каждый гражданин не только может ознакомиться со всеми своими данными, которые там имеются, но и узнать, кто еще интересовался этими данными, так как каждый запрос там фиксируется и оставляет свой след.

X-Road – это децентрализованная система, в которой определены единые интерфейсы и протоколы взаимодействия и обмена данными. Любой госорган может по определенным правилам подключить свою информационную систему к X-Road и настроить права доступа к нему для других ведомств. Коммерческие ИТ-системы тоже могут взаимодействовать с X-Road и могут брать оттуда данные. В результате — регистрация юрлица занимает 15 минут, более 25% эстонцев голосует через интернет на последних парламентских выборах, более 99% налоговых деклараций сдается в электронном виде и т.д.

Таков X-Road — достижение эстонцев, сделавшее государство, бизнес и граждан ближайшими союзниками. По заказу государства частные компании разработали X-Road — протокол для обмена данных между разными базами данных. Это краеугольный камень всей системы госуслуг Эстонии.

X-Road соединяет разные реестры. Через него общаются, например, реестр коммерческих компаний и реестр автомобилей. Протокол оказался настолько удобным, что его стал использовать бизнес. Например, через X-Road ведут общие дела две компании телеком-оператора TeliaSonera. В последние годы система становится главным экспортным продуктом эстонского электронного государства. Финляндия уже переводит на неё все свои е-госуслуги (сам код передали «финским братьям» бесплатно, но на инсталляции зарабатывают частные эстонские компании), другие страны ЕС используют её для некоторых из своих е-сервисов. В Эстонии X-Road «гоняет» информацию между 170 базами данных.

Каждое министерство и ведомство имеет свою базу данных, которые чаще не совместимы и работают на различных платформах. X-Road-центр, расположенный в Министерстве экономики и коммуникаций, отвечает за то, чтобы формат, даты и предоставляемая информация совпадали.

В системе X-Road каждое учреждение остается хозяином своей информации и само определяет кому предоставить доступ к какой категории данных. Учреждение само обновляет свои данные. Таким образом, после внедрения X-Road, отпадает надобность дублировать базы данных и, вместе с тем, все пользователи имеют доступ к самой свежей информации.

При разработке X-Road соблюдены следующие принципиально важные условия, связанные с передачей данных: безопасность, стандартизация, наблюдаемость, отслеживаемость, достоверность.

В настоящее время в Эстонии есть 3 способа идентификации личности в Интернете. Это ID-карта, digi-ID и Mobil-ID. Основное и наиболее распространенное, то есть базовое средство идентификации человека - это ID-карта, у которой имеются две основные функции. С одной стороны - это обычный идентификационный документ, который снабжен полутора десятками элементов безопасности (микрочасть, голограмма и т.д.), делающими ID-карту полноценным рейсовым документом, средством идентификации человека в реальной жизни. С другой стороны, в нее включен чип, на котором установлены два приложения - идентификационное с PIN-кодом, а также приложение Электронной цифровой подписи (ЭЦП) со своим PIN-кодом. Это два разных PIN-кода различной длины. Первые ID-карты начали выдавать в Эстонии в 2000 году, а по состоянию на 2011 год их было выдано уже 1,1

млн. штук (при населении страны немногим более 1.3 миллиона человек). Т.е. практически 90% населения страны в настоящее время имеет эти карточки.

Со второй половины прошлого года в Эстонии была введена карта Digi-ID, которая по сути своей является чипом без средств визуальной идентификации и элементов визуальной защиты. На ней нет фотографии, указаны только имя и фамилия человека. Чип же на Digi-ID имеет ту же структуру, что и на обычной ID-карте. Спрашивается, зачем она была введена? Затем, что три четверти государственных услуг в Эстонии можно получить посредством интернета, и идентификация человека в интернете становится критическим моментом. Дело в том, что потеря или приход в негодность обычной ID-карты лишает человека на 1-2 недели возможности потребления услуг, поскольку сама по себе эта карта - достаточно сложный и защищенный документ, производство и оформление которого занимает определенное время. В этой ситуации должна существовать возможность быстрого получения некоего альтернативного варианта карты. Карту Digi-ID можно получить в любом департаменте гражданства и миграции буквально за 20 минут (они печатаются и персонифицируются на месте), и после ее получения продолжать потреблять госуслуги в интернете и спокойно ждать того момента, когда будет изготовлена основная ID-карта.

Третий способ идентификации - мобильная идентификация, Mobil-ID. Впервые он был внедрен эстонским оператором EMT в 2007 году. Идея Mobil-ID состоит в следующем. На SIM-карту мобильного оператора сверху записываются вышеуказанные два приложения - идентификационное и ЭЦП. Таким образом, пользователь Mobil-ID может идентифицировать себя с помощью мобильного телефона, с той лишь разницей, что ему не нужно для этого считывающее устройство для ID-карты. Это очень удобно тогда, когда человек оказывается за рубежом. Mobil-ID работает и в международном роуминге, когда человека обслуживает другой сотовый оператор. Мобильная идентификация Mobil-ID очень популярна и доступна любому жителю Эстонии от школьника до пенсионера, поскольку около 80% жителей Эстонии пользуются мобильным телефоном, а мобильной сетью покрыта почти вся территория страны. К тому же на законодательном уровне разрешено совершение микроплатежей с мобильного счета абонента, стимулирующих участие мобильных операторов.

Новейшие технологии в сфере электронного государства разработали и координируют частные компании и государственные ведомства, среди них e-Governance laboratory, Actors, Cybernetica. В результате их совместных 10-летних разработок в области эффективного и безопасного обмена данными создана современная система электронного взаимодействия - X-Road. Разработка и запуск среды обмена данными заняла в Эстонии 2 года. Проект начался в 2000 году, внедрение системы стартовало в начале 2002 года.

### **1.2.2 Модель интеграции оптимальных информационных систем в сельском хозяйстве России**

Рассматривается ситуация [12], когда пользователи и ИП территориально распределены и осуществляется допроектирование новых ИС к существующим. После анализа существующих БД и требований, предъявляемых к новым принимается решение о полном перепроектировании всех ИС, и новых и старых. Переносится информация из существующих ИС в объединенную новую.

Под структурой информационно-управляющей системы (ИУС) АПК будем понимать организационную совокупность ее взаимосвязанных элементов, определяющих их место как в чисто физическом, так и технологическом смысле (уровень и конкретное место размещения элемента в пространстве и технологической схеме принятия решений и обработки информации).

В общем случае под элементом структуры ИУС АПК понимается некоторый физический элемент (узел, техническое устройство или средство, исполнитель или коллектив исполнителей и т.п.), реализующий вполне определенным образом некоторую функцию системы (задачу, группу взаимосвязанных функций или задач).

Таким образом, понятие структуры ИУС АПК предполагает частичную упорядоченность ее элементов относительно друг друга как в смысле их размещения по физическим узлам и уровням, так и в смысле решаемых ими функциональных задач процесса управления. Под синтезом структуры ИУС АПК понимается процесс построения взаимосвязей элементов структуры ИУС АПК и самих элементов в соответствии с заданными критериями эффективности ИУС АПК в целом.

В общем случае задача синтеза структуры ИУС АПК с учетом его особенностей включает в себя: 1) выбор принципов построения системы управления; 2) распределение функций управления и обработки информации по узлам и уровням иерархии ИУС АПК; 3) рациональное размещение задач принятия решений и обработки информации по конкретным элементам с одновременным выбором методов их решения и согласованием целей отдельных элементов, уровней и подсистем с общим для всей системы; 4) распределение функций и задач между техническими средствами и коллективами специалистов по управлению; 5) определение наиболее эффективных взаимосвязей между всеми составляющими структуры системы; 6) набор и назначение технических средств для решения задач управления и передачи данных в ИУС АПК [13].

Постановка и решение общей задачи синтеза структуры ИУС АПК из-за особенностей агропромышленного производства, приведенных выше, представляют значительные трудности, поэтому формулируются и решаются частные задачи, позволяющие оценить правильность принимаемых решений о структуре создаваемой ИУС АПК.



Как правило, частные задачи отражают конкретные аспекты процесса синтеза структуры ИУС АПК, а результаты их решения позволяют специалистам по синтезу структур наметить наиболее перспективные пути решения общей задачи и выделить область наиболее целесообразных вариантов построения всей ИУС АПК.

Рассмотрим процесс синтеза структуры ИУС АПК с формальных позиций, для чего введем следующие обозначения.

Обозначим множество возможных принципов построения системы и ее элементов через  $P$  (к ним можно отнести, например, принципы централизованного или децентрализованного управления). Каждому набору принципов  $\pi$  построения системы соответствует некоторое множество возможных функций  $f \in F(\pi)$ , из которого, при проектировании системы необходимо выбрать подмножество, удовлетворяющее некоторым критериям. Через  $m$  обозначим множество взаимосвязанных физических элементов структуры ИУС АПК. Подобными элементами могут быть узлы системы, технические средства, пункты обслуживания, отдельные исполнители, коллективы и т.д. Через  $A$  обозначим операцию отображения  $F(\pi)$  на  $m$ . Оптимальное отображение должно обеспечивать экстремум некоторой целевой функции при выполнении заданных ограничений.

С учетом введенных обозначений можно выделить следующие задачи анализа и синтеза структуры ИУС АПК:

Задача 1. В общем случае задача синтеза оптимальной структуры состоит в

$$\Phi^{\wedge} | \{ [f \in F(\pi)] A[\bar{m} \in m] \} \rightarrow extr; \quad (1)$$

$$\pi \in P; \quad (2)$$

$$f \in F(\pi); \quad (3)$$

$$\bar{m} \in m. \quad (4)$$

определении.

Задача 2. Если принципы построения системы заданы, то задача синтеза оптимальной структуры состоит в определении (1), (3) и (4).

Задача 3. Если заданы принципы построения системы и выполняемые ею функции, то задача синтеза оптимальной структуры состоит в определении (1) и (4).

Задача 4. Если заданы принципы построения системы, выполняемые ею функции и элементы системы, то задача синтеза оптимальной структуры состоит в определении (1), т.е. рационального отображения множества взаимосвязанных функций на множество взаимосвязанных элементов системы.

Задача 5. Задача анализа состоит в определении характеристик системы при заданных (1) - (4).

В зависимости от цели исследования в понятие "разработка структуры системы" включаются различные вопросы из числа рассмотренных выше.

Так, при разработке, например, структуры ИУС АПК необходимо определить множество узлов системы (множество центров на каждом уровне и число ступеней иерархии или уровней управления) и связей между ними. Необходимо распределить множество задач управления и обработки информации по различным ступеням и узлам системы и выбрать комплекс технических средств, наилучшим образом удовлетворяющих заданным требованиям [14].

По нашей классификации информационно-управляющие системы по степени влияния на объект управления условно делятся на 4 класса.

1. Системы, которые на каждом уровне и в каждом звене управления автоматизируют существенные функции управления.
2. Системы, которые оптимизируют систему управления в части затрат на информационную технику и передачу информации, дублирование функций и данных.
3. Системы, которые изменяют структуру системы управления объектом,
4. Системы, которые способствуют изменению самого объекта, например, структуры производства.

Модель интеграции оптимальных информационных систем, состоит из следующих блоков:

- блок размещения запросов пользователей (задач) по узлам управления и техническим средствам;
- блок расчета потребностей в средствах вычислительной техники и распределения их по узлам управления и закрепления за пользователями;
- блок расчета инвестиций в средства связи;
- блок расчета распределения информации по узлам управления и техническим средствам;
- блок формирования логических структур распределенных баз данных (РБД).

Модель позволяет создать архитектуру ИУС второго, третьего и, частично, четвертого класса.

Подавляющее же количество работ по проектированию ИУС, в частности РБД, относится к первому классу, так проектирование РБД основывается на постулате, что

структура системы и узлы управления заданы, информационная техника с заданными характеристиками установлена, потребности в информации определены.

На основании материалов предыдущих разделов можно сделать вывод, что было бы наиболее рационально (в общем виде) проектировать ИУС одновременно, взаимосвязанно с реорганизацией (реинжинирингом) как системы управления, так и самой производственной системы.

Для описания структуры системы управления рассмотрим систему, состоящую из множества узлов управления  $j$  (например, животноводческих ферм, растениеводческих отделений, механических мастерских и т.д.), множества задач  $K$ , кластеров данных  $L$ , типов связи  $R$ .

Считаем, что информация делится на 2 типа: первичную, возникающую в узлах управления, и вторичную, возникающую после решения некоторой задачи. С другой стороны, разбиение на кластеры осуществляется процедурой разбиения данных на предметные базы данных.

Процесс управления предполагается периодически с периодом  $T$ , и все операции расчетов, передачи данных и т.д. усреднены по времени.

Будем считать, что любая задача может решаться в любом узле, в том числе разбиваться по этим узлам. Например, начисление заработной платы может решаться в каждом узле на своем объеме информации.

Рассмотрим математическую модель гипотетической ИУС.

Введем обозначения:

$K$  - номер задачи (запрос на информацию, на корректировку БД);

$l$  - номер группового информационного элемента;

$l \in L^1$  - первичная информация;

$l \in L^2$  - вторичная;

$K^1$  - множество задач, использующих лишь первичную информацию;

$K^2$  - множество задач, использующих хотя бы один групповой информационный элемент;

$j$  - номер узла управления,  $j \in J$ ;

$i$  - номер технического средства,  $i \in I$ ;

$L_k$  - множество групп  $l$ , используемых в заданных  $k$ ;

$\varphi_{pk} = \begin{cases} 1, & \text{если } p\text{-ый пользователь использует } k\text{-ю задачу,} \\ 0, & \text{иначе,} \end{cases}$

$$\omega_{pj} = \begin{cases} 1, & \text{если } p\text{-ый пользователь прикреплен к } j\text{-му узлу,} \\ 0, & \text{иначе,} \end{cases}$$

$g_l$  - вектор длины групп;

$f_l$  - вектор количества экземпляров в группах;

$T = \|\tau_{pl}\|$ , где  $\tau_{pl}$  - максимально достижимое время получения информации l-й группы

по запросу p-го пользователя;

$$A = \|a_{ij}\| = \begin{cases} a_{ij} = 1, & \text{если существует семантическая связь между } i\text{-й и } j\text{-й группами,} \\ 0, & \text{иначе,} \end{cases}$$

$f_{klj}^e$  - средние характеристики (объем информации, временные, пропускные способности и т.д.) на информацию l-ой группы, необходимый для задачи k, возникающий в узле j,  $e \in E$ ;

$p_{kln}^e$  - средние характеристики на информацию l-й группы, необходимой для задачи k, после решения задачи n;

$$x_{ijk} = \begin{cases} 1, & \text{если } k\text{-я задача решается в} \\ & \text{узле } j\text{-м на } i\text{-м типе оборудования} \\ 0, & \text{иначе,} \end{cases}$$

$$x_{lt} = \begin{cases} 1, & \text{если } l\text{-я информационная группа входит в} \\ & t\text{-й тип записи} \\ 0, & \text{иначе,} \end{cases}$$

$$y_{klj} = \begin{cases} 1, & \text{если } t\text{-й тип записи размещается на } i\text{-м} \\ & \text{типе оборудования } j\text{-го узла,} \\ 0, & \text{иначе,} \end{cases}$$

$$a_{klj} = \begin{cases} 1, & \text{если } l\text{-я группа возникает в узле } j \\ & \text{для } k\text{-й задачи,} \\ 0, & \text{иначе,} \end{cases}$$

$$y_{l_1 l_2 j_1 j_2 r} = \begin{cases} 1, & \text{если информация из } l_1\text{-й группы не-} \\ & \text{редается из } j_1\text{-го типа оборудования} \\ & j_1\text{-го узла в } j_2\text{-й тип оборудования} \\ & j_2\text{-й узел посредством связи } r; \\ 0, & \text{иначе,} \end{cases}$$

$$b_{kln} = \begin{cases} 1, & \text{если } l\text{-я группа используется в } k\text{-й} \\ & \text{задачепослерешения задачи } n; \\ 0, & \text{иначе} \end{cases}$$

$Z_{ij}$  - количество оборудования  $i$ -го типа в  $j$ -м узле;

$d_{mijk}$  - необходимые ресурсы  $m$ -го типа для решения  $k$ -й задачи на  $i$ -м типе оборудования в  $j$ -м узле;

$M_{mi}$  -  $m$ -е ресурсы  $i$ -го типа оборудования;

$$s_{i_1 i_2 j_1 j_2 r} = \begin{cases} 1, & \text{если } r\text{-й тип связи используется для} \\ & \text{передачи информации из } i_1\text{-го типа} \\ & \text{оборудования } j_1\text{-го узла в } i_2\text{-й тип} \\ & \text{оборудования } j_2\text{-го узла,} \\ 0, & \text{иначе} \end{cases}$$

$G_r^e$  - характеристики средств связи;

$c_{ij}^1$  - стоимость  $i$ -го оборудования в  $j$ -м узле;

$c_{i_1 i_2 j_1 j_2 r}^2$  - стоимость  $r$ -го средства связи при передаче информации из  $(i_1, j_1)$  в  $(i_2, j_2)$ ;

$c_{i_1 j_1 i_2 j_2 r}^3$  - затраты на передачу единицы информации из  $(i_1, j_1)$  в  $(i_2, j_2)$ ;

$c_{ijk}^4$  - стоимость  $m$ -го ресурса для решения  $k$ -й задачи на  $i$ -м типе оборудования в  $j$ -м узле;

$c_k^5$  - обобщенная стоимость  $k$ -й задачи;

$c^0$  - средства, выделенные на внедрение ИУС;

$$v_k = \begin{cases} 1, & \text{если } \sum_{i,j} x_{ijk} \geq 1, \\ 0, & \text{иначе} \end{cases} \quad U_{kn} = \max_{l \in L_k} b_{kln};$$

Ограничения на размещение задач по узлам и техническим средствам:

$$\sum_{i,j} x_{ijk} \geq 1, \quad k \in K^3 \in K, \quad (1), \quad \text{либо}$$

$$\sum_{i,j,p} x_{ijk} y_{pk} \omega_{pj} \geq 1, \quad \text{то есть } k\text{-я задача должна быть решена хотя бы в одном узле;}$$

$$\sum_i x_{ijk} \geq 1, \quad j \in J_1 \quad k \in K^4 \in K, \quad (2), \quad \text{либо}$$

$\sum_{i,p} x_{ijk} y_{pk} \omega_{pj} \geq 1$ , т.е. некоторые задачи из множества  $k$  должны быть обязательно

решены в узлах  $j \in J_1$ , например, в центральной бухгалтерии предприятия, в том числе вторичные задачи;

$$\sum_i x_{ijn} \geq U_{kn} V_k, \quad k, n \in K, \quad i \in I, \quad j \in J, \quad (3),$$

то есть должны быть решены все задачи  $n$ , информация которых используется для решения задачи  $k$ .

Условия передачи информации из узла  $j_2$  в узел  $j_1$ :

$$\begin{aligned} \sum_z y_{l_{i_1 j_1 i_2 j_2}^r} &= \sum_n \sum_k b_k l_n x_{i_1 j_1^n} x_{i_2 j_2^k}, \quad l \in L^2 \\ \sum_{r, l_2} y_{l_{i_1 j_1 i_2 j_2}^r} &= \sum_k a_{kl j_2} x_{i_1 j_1^k} \quad l \in L^1 \end{aligned} \quad (4)$$

Информация передается из узла  $j_2$  в узел  $j_1$ , когда она либо возникает в узле  $j_2$  и используется в узле  $j_1$  для задачи  $k$ , либо когда вторичная информация задана  $n$ , решаемой в узле  $j_2$ , необходима для задачи  $k$ , решаемой в узле  $j_1$ ;

$$\sum_r y_{l_{i_1 j_1 i_2 j_2}^r} \leq 1, \quad (5)$$

информация передается одним средством связи;

$$\sum_{i_2, j_2, r} y_{l_{i_1 j_1 i_2 j_2}^r} = \sum_{t, k} \omega_{kl} x_{lt} y_{t i_1 j_1}, \quad (6)$$

условие, что все группы в узле  $j_1$  на  $i_1$ -м типе оборудования формируются в записи  $t$ .

Ограничение на число групп в составе логической записи

$$\sum_t x_{lt} \leq N_t, \quad t \in [1, t_0] \quad (7)$$

Ограничения на длину формируемой логической записи

$$\sum_t x_{lt} y_{tij} y_l \leq x_{tij}, \quad (8)$$

Ограничение на однократность включения групп в логические записи:

$$\sum_t x_{lt} = 1, \quad \forall l \in L \quad (9)$$

Ограничение на число типов записей, размещаемых в БД j-го узла на i-м типе оборудования

$$\sum_t y_{tij} \leq h_{ij}, \quad (10)$$

Ограничение на загрузку оборудования:

$$\sum d_{mijk} x_{ijk} \leq M_{m_i}, \quad (11)$$

Ограничения на канал связи:

$$\sum_{l,k} y_{l i_1 j_1 i_2 j_2^r} f_{kl j_2}^e + \sum_{l,k,n} p_{kln}^e y_{l i_1 j_1 i_2 j_2^r} \leq G_r^e s_{i_1 j_1 i_2 j_2^r} \quad (12)$$

Финансовые ограничения:

$$\sum_{i,j,k} c_{ij}^1 x_{ijk} + \sum c_{i_1 j_1 i_2 j_2^r}^2 s_{i_1 j_1 i_2 j_2^r} + \sum c_k^5 x_{ijk} \leq c^0 \quad (13)$$

Критерий эффективности:

$$\begin{aligned} & \sum c_{ij}^1 x_{ijk} + \sum c_{i_1 j_1 i_2 j_2^r}^2 s_{i_1 j_1 i_2 j_2^r} + \sum c_{i_1 j_1 i_2 j_2^r}^3 f_{kl j_2}^e y_{l i_1 j_1 i_2 j_2^r} + \\ & + \sum c_{i_1 j_1 i_2 j_2^r}^3 p_{kln}^t y_{l i_1 j_1 i_2 j_2^r} + \sum c_{mijk}^4 d_{mijk} x_{ijk} + \\ & + \sum c_k^5 x_{ijk} \rightarrow \min \end{aligned} \quad (14)$$

Очевидно, что решить данную задачу в общем виде - сложная проблема. Для ее решения существует несколько подходов. Первый - воспользоваться алгоритмами, предложенными в [15] с небольшими модификациями, связанными с появлением блока синтеза структур. Второй - исходя из конкретики каждого хозяйства, на основе требований и пожеланий руководства, финансовых возможностей, сложившихся структур управления, ее

консерватизма, попытаться свести задачу к многоэтапной итеративной процедуре, аналогично изложенной в [16].

Для этого введем предположение, что информация для задач из множества  $K^2$  всегда имеется, например, задача из множества  $K^1$  решается либо вручную, либо на ЭВМ, и информация эта возникает в месте ее решения. Тогда система уравнений примет следующий вид:

Уравнения (1), (2) остаются без изменений, в уравнении (3) нет необходимости.

Уравнение (4) принимает вид:

$$\sum_{r, l_2} y_{l_1 j_1 i_2 j_2^r} = \sum_k a_{kl j_2} x_{i_1 j_1^k} \quad (15)$$

Уравнения (5)-(11) остаются без изменений. Уравнение (12) принимает вид:

$$\sum_{l, k} y_{l_1 j_1 i_2 j_2^r} f_{kl j_2}^e \leq G_r s_{i_1 j_1 i_2 j_2^r} \quad (16)$$

Уравнение (13) остается без изменений. Критерий эффективности приобретает вид:

$$\begin{aligned} & \sum c_{ij}^1 x_{ijk} + \sum c_{i_1 j_1 i_2 j_2^r}^2 s_{i_1 j_1 i_2 j_2^r} + \sum c_{i_1 j_1 i_2 j_2^r}^3 f_{kl j_2}^e y_{l_1 j_1 i_2 j_2^r} + \\ & + \sum c_{mijk}^4 d_{mijk} x_{ijk} + \\ & + \sum c_k^5 x_{ijk} \rightarrow \min \end{aligned} \quad (17)$$

Данная система уравнений все еще имеет нелинейный вид. Для приведения ее к линейному виду произведем замену переменных:

$$z_{ltij} = x_{lt} y_{tij} \quad (18)$$

с исключением из системы уравнения (10).

После решения данной системы уравнений получим, в частности, конкретные значения  $x_{lt}^0$ ,  $y_{tij}^0$ . Если при этом выполняется условие (10), то есть



$$\sum_t y_{ij}^0 \leq h_{ij}, \quad (19)$$

то на этом процедура расчета по модели завершается. Если же условие (10) не выполняется при некоторых  $(i, j) \in (I^0, J^0)$ , тогда для данных  $(i, j) \in (I^0, J^0)$  либо подбирается такая СУБД, которая удовлетворяет условию (19), либо, если такую замену совершить невозможно, вместо выражения (10) в систему включается следующее:

$$\sum_t z_{litj} \leq h_{ij}, \quad \forall l \in L \quad (20)$$

Очевидно, что при  $\chi_{lt} = 1$ , выражение (20) совпадает с (10), при  $\chi_{lt} = 0$  условие (10) выполняется автоматически.

Труднее обстоит дело в том случае, когда приобретаемое программное обеспечение жестко завязано по логической структуре собственной БД. В этом случае придется написать интерфейс по доступу к этим БД из других задач и при расчете по рассматриваемой в работе модели соответствующие переменные заменить константами, определяющими заданные логические структуры БД.

Модель также упрощается, когда, например, руководство предприятия жестко требует создания локальной сети в центральной конторе на основе уже имеющейся вычислительной техники, расставленной по рабочим местам, причем не только в центральной конторе. В такой ситуации модель распадается на ряд блоков, решаемых последовательно. Другой пример - руководство предприятия выделяет на компьютеризацию столь малые средства, например, на один или два ПЭВМ. В этом случае модель также существенно упрощается. И таких примеров можно привести много. Модель заработает в полном объеме в том случае, когда руководство решится на комплексную компьютеризацию предприятия с возможной в последующем реорганизацией системы управления.

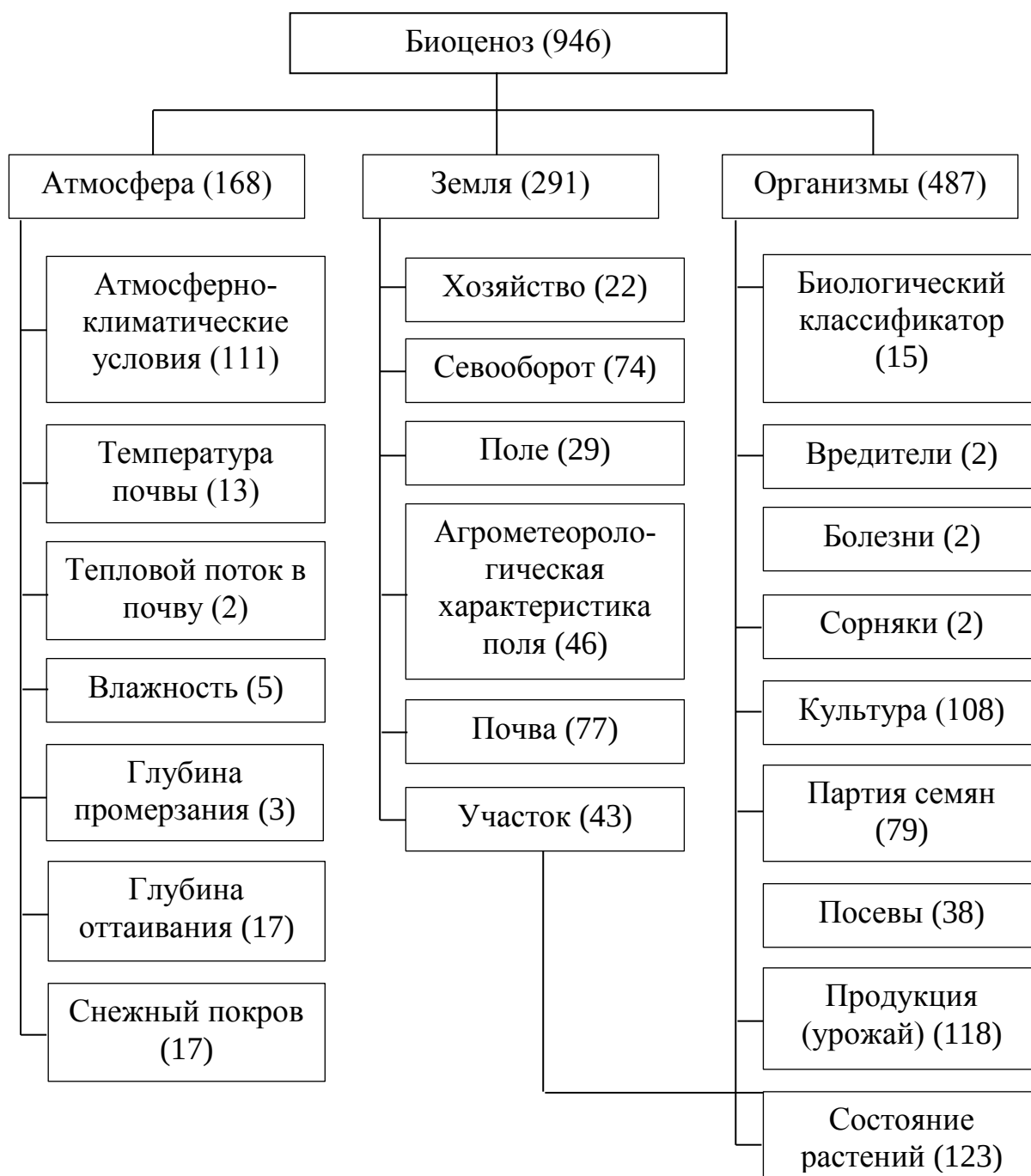
Представленная в работе модель, по нашему мнению, отражает наиболее рациональный уровень влияния ИУС на систему управления, исходя из сложившихся условий в стране. Модель позволяет в пределах выделенных финансовых ресурсов определить наиболее рациональную структуру ИУС, распределяет средства ВТ и решаемые задачи по узлам управления, определяет при необходимости инвестиции в телекоммуникационные средства с оптимизацией информационных потоков и логических структур распределенных баз данных. По мере развития теории управления, теории систем

на основе дальнейших исследований в области ИУС модель может дополняться соответствующими блоками.

### **1.2.3 Интеграция информационных ресурсов при проектировании цифровой платформы АПК России**

Почти все ИС и сайты в сельском хозяйстве имеют гетерогенные структуры БД. Так, прошедшая 5-6 апреля 2017 г. конференция «Информационные технологии на службе агропромышленного комплекса России» показала, что в АПК продолжается «островная информатизация» отдельных предприятий, в частности, у дилеров и разработчиков программного обеспечения точного земледелия, космического мониторинга земель, ИУС базы данных растениеводства гетерогенны, неполны, несовместимы. В то же время, еще на эталонном объекте – агрокомбинате «Кубань» в рамках задания «Электронизация сельского хозяйства» комплексной программы стран-членов СЭВ при реализации подсистемы растениеводства силами творческого коллектива из различных отраслевых растениеводческих НИУ и НИИ кибернетики АПК на единой методической основе была осуществлена интеграция знаний различных агропромышленных технологий в растениеводстве.

Требование интегрированности решения задач по информации, по режимам ее обработки, а также требование функциональной их полноты послужило основой для формирования перечня задач и их классификации (выделено для автоматизации 340 задач), и для проектирования логической структуры базы данных (БД) (151 вид записей), общей для всех растениеводческих предприятий России. На рисунке 1.2.3.1 приведена укрупненная концептуальная модель растениеводства (в скобках указано количество атрибутов в соответствующем информационном блоке).



**Рис.1.2.3.1 Укрупненная информационная модель растениеводства**

В современных условиях при реализации типового сайта сельскохозяйственного предприятия [17], данная логическая структура БД явится интегратором всех систем точного земледелия, космического мониторинга земель и других ИУС с единых позиций. Это позволит реализовать типовые решения, практически бесплатные для товаропроизводителя. Данная логическая структура и классификация указанных задач для автоматизации может стать основой для стандартизации цифровой платформы для отрасли растениеводства. Подобный подход для других отраслей, в т.ч. животноводства, при размещении соответствующей БД в некотором “облаке”, например, у провайдера, имеющего мощную

систему управления базами данных (СУБД), расширит цифровую платформу на все отрасли сельского хозяйства. Очевидно, что эти БД будут интегрированы друг с другом.

Рассмотрим, как интеграция Интернет-технологий должны повлиять на производственный и бухгалтерский учет. Еще 30 лет назад при информатизации агрокомбината «Кубань» в одном из хозяйств были поставлены ПК, связанные сетью, в пункты сбора первичной учетной информации. Вся первичная учетная информация поступала в единую БД, которая была сформирована на универсальной структуре в виде кортежа: вид операции, место проведения, кто проводил, дата, интервал времени, средства производства, объем, вид потребленного ресурса, объем потребленного ресурса. На этой основе можно рассчитать как заработную плату, так и вычислить материальные затраты, осуществить технологический учет и т.д. без создания промежуточных баз данных.

Из единой БД могут получать информацию для расчетов непосредственно, как бухгалтера, так и остальные специалисты. Бухгалтерский же учет в существующем виде до сих пор основан на многократном агрегировании информации из первичного учета в различных срезах. Возможности же современных вычислительных средств позволяют хранить и производить расчеты на основе первичной информации без создания промежуточных баз данных. При этом единая БД первичного учета, которая хранится в «облаке» типового сайта, может заполняться учетчиком с любого мобильного устройства. Часть полей единой БД заполняется автоматически информацией с различных датчиков и приборов.

Наиболее наглядный пример влияния интеграции Интернет-технологии на структуру производства [18] - изменение структуры ИКС в развитых странах. С развитием Интернет-технологий данная структура начинает претерпевать изменения. Исторически так сложилось, что ИКС являлся посредником между товаропроизводителем и носителями аграрных знаний. Типовая схема работы: товаропроизводитель обращается в ИКС с каким-либо вопросом; консультант на основе собственного опыта либо сам готовит ответ, имея в распоряжении соответствующую литературу, базы данных, пакеты прикладных программ и прочие источники знаний, либо переадресует вопрос более знающему консультанту.

С развитием Интернет-технологий данная схема начинает претерпевать изменения. Например, в Австралии почти отказались от посредника-консультанта. В стране нет отдельной федеральной или региональной информационно-консультационной службы. Ресурсы в Интернете, исполняющие функции предоставления аграрной информации и отчасти консультирования фермеров берут на себя сайты Министерств сельского хозяйства провинций. Но у них на сайтах даже нет раздела «Консультирование». Пользователь ищет нужную информацию на этих сайтах самостоятельно, ориентируясь по отраслевому

признаку. Иногда в представленной информации можно найти ответ на интересующий фермера вопрос, но связи «консультант-фермер» не существует.

Этот процесс происходит даже в условиях отсутствия сформированного на Западе единого информационного Интернет-пространства научных знаний, которое только начинает формироваться, в основном, за счет интеграции образовательных ресурсов на базе дистанционного образования (ДО).

Аналогично, при реализации типового сайта сельскохозяйственного предприятия и обязанности отражения в общем «облаке» статистической информации должна существенно видоизмениться структура Росстата.

Таким образом, интеграция единого информационного Интернет-пространства аграрных знаний [19, 20, 21] с типовыми сайтами сельскохозяйственных предприятий, сельскохозяйственных НИИ и ВУЗов; типовыми ИУС и системами первичного учета, реализованными с помощью Интернет-технологий, представляет собой основу цифровой платформы, логическая структура которой готова к интеграции с различными цифровыми устройствами и приборами в дальнейшем.

## **2 СОЗДАНИЕ БАЗ ДАННЫХ АГРАРНЫХ ЗНАНИЙ «ПУБЛИКАЦИИ» И «РАЗРАБОТКИ» НИУ АГРАРНОГО ПРОФИЛЯ**

### **2.1 Источники информации для баз данных аграрных знаний**

В настоящее время можно найти информацию о публикациях и разработках НИУ сельского хозяйства из следующих источников: elibrary, БД ФИПС, БД “ЕГИСУ НИОКТР”, сайты НИУ. Все эти источники, как и следовало ожидать, имеют гетерогенные структуры (табл. 2.1.1). Хотя государство затратило 2 млрд. рублей на разработку нового варианта БД “ЕГИСУ НИОКТР” (с 2013 года). При этом БД до 2013 года не вошла в структуру нового варианта БД “ЕГИСУ НИОКТР” в силу их несовместимости. В новом варианте БД “ЕГИСУ НИОКТР” уже используется ГРНТИ.

Таблица 2.1.1

## Поисковые поля информации в системах ELIBRARY, ЕГИСУ, ФИПС, ВНИИЭСХ, ЦНСХБ

| ELIBRARY                                                  | ЕГИСУ (после 2013 г)                      | ЕГИСУ (до 2013 г. вкл. ВНИИЭСХ) | ФИПС (после 2013 г)                | ЦНСХБ                                                                      | Общие поля                  | Публикации в ЕИПАЗ    | Разработки в ЕИПАЗ    |
|-----------------------------------------------------------|-------------------------------------------|---------------------------------|------------------------------------|----------------------------------------------------------------------------|-----------------------------|-----------------------|-----------------------|
| Контекстный поиск                                         | Виды разработок <sup>^</sup>              | Название, контекстный поиск     | Тематика (слова)                   | Виды информресурсов (БД Агрос, каталог ФАО, каталог книг, журналов и т.д.) | Название, контекстный поиск | Название              | Название              |
| Где искать <sup>1</sup>                                   | Организация                               | Вид разработки                  | Название                           | Вид документа                                                              | Вид разработки, публикации  | Ведомство             | Ведомство             |
| Вид публикации                                            | Приоритетное направление                  | Организация                     | Номер документа                    | Название                                                                   |                             | Департамент ведомства | Департамент ведомства |
| ГРНТИ (множественный выбор)                               | ГРНТИ <sup>^^</sup>                       | Авторы                          | Дата опубликования                 | Авторы (множественный выбор)                                               |                             | Организация-автор     | Организация-автор     |
| Авторы (множественный выбор)                              | Критические технологии <sup>^^^</sup>     |                                 | Индекс МПК*                        | ГРНТИ (до 4-го уровня)                                                     |                             | Подразделение         | Подразделение         |
| Журналы (возможен множественный выбор из списка журналов) | Качество результата                       |                                 | Заявитель(ли)                      | Ключевые слова                                                             |                             | Авторы                | Авторы                |
| Поиск в личной подборке публикаций                        | Ожидаемый эффект                          |                                 | Автор(ы)                           | Страна                                                                     |                             | Вид публикации        | Вид разработки        |
| Параметры <sup>2</sup>                                    | Дата (от)                                 |                                 | Патентообладатель(и) (Организация) | Язык                                                                       |                             | Авторы-коллектив      | Авторы-коллектив      |
| Годы публикации                                           | Дата (до)                                 |                                 | Дата публикации заявки             | Год                                                                        |                             | ГРНТИ                 | ГРНТИ                 |
| За время поступления <sup>3</sup>                         | Ключевые слова                            |                                 | Патентный поверенный               | ISBN/ ISSN                                                                 |                             | Год                   | Год                   |
| Сортировка (по релевантности, по дате, и т.д.)            | Статус лицензии (открытая, безвозмездная) |                                 | Дата заявки РСТ**                  |                                                                            |                             | Организация-место     | Организация-место     |
| Порядок (по возрастанию, по убыванию)                     |                                           |                                 | Номер документа РСТ                |                                                                            |                             | Ключевые слова        | Ключевые слова        |
| Организация                                               |                                           |                                 | Адрес для переписки                |                                                                            |                             |                       |                       |

|  |  |  |                                  |  |  |  |  |
|--|--|--|----------------------------------|--|--|--|--|
|  |  |  | Аналоги слов документа           |  |  |  |  |
|  |  |  | Код вида документа               |  |  |  |  |
|  |  |  | Дата публикации извещения        |  |  |  |  |
|  |  |  | Рег. номер заявки                |  |  |  |  |
|  |  |  | Дата подачи заявки               |  |  |  |  |
|  |  |  | Конвенционный приоритет          |  |  |  |  |
|  |  |  | Дата подачи конвенционной заявки |  |  |  |  |
|  |  |  | Страна приоритета                |  |  |  |  |
|  |  |  | Слова в реферате                 |  |  |  |  |
|  |  |  | Слова в формуле                  |  |  |  |  |
|  |  |  | Слова в описании                 |  |  |  |  |
|  |  |  | Статус документа***              |  |  |  |  |
|  |  |  | База данных**** (Вид разработки) |  |  |  |  |

<sup>1</sup> Где искать – в названии, в аннотации, в ключевых словах, в названии организации, в списках цитирования, в тексте публикации.

<sup>2</sup> Параметры – с учетом морфологии, похожий текст, полный текст, доступные, в результате предыдущего поиска.

<sup>3</sup> За время поступления – за все время, за последний месяц, за последние три месяца, за последние полгода, за последний год

<sup>^</sup> Виды разработок – 4 вида: научный отчет, диссертации, результаты интеллектуально деятельности, научно-исследовательские, опытно-конструкторские, технологические работы.

<sup>^^</sup> 3-х уровневый рубрикатор, выбор однозначного пункта только из нижнего уровня.

<sup>^^^</sup> Критические технологии – выбор из предлагаемого списка видов технологий.

\* МПК - Международная Патентная Классификация

\*\* РСТ - Patent Cooperation Treaty – Договор патентной кооперации

\*\*\* Статус документа - Действует – 1, Может прекратить свое действие – 2, Прекратил действие, но может быть восстановлен – 3, Не действует -4.

\*\*\*\* База данных – выбор данных по типу документов с 2013 года:

- патентные документы;
- товарные знаки;
- промышленные образцы;
- программы для ЭВМ, БД и ТИМС (Топологии интегральных микросхем)



В некоторых западных странах развитие электронных информационных ресурсов проходит также недостаточно интенсивно. Информатизация университетов в США развивается эволюционно. Там нет единого Интернет-пространства научных информационных ресурсов, например, электронные библиотеки университетов находятся в стадии становления, и библиотекам приходится экспериментировать, искать, отлаживать новые технологии. Библиотеки США расходуют на электронные книги не намного больше средств, чем их российские коллеги, и обладают в среднем сопоставимыми электронными ресурсами. Однако в сравнении американской и российской ситуаций есть принципиальная разница. Российские НИУ и ВУЗы нормативно принуждены развивать свои сайты, например, в РФ Министерством образования и науки, Рособнадзором выдвинуты требования о составе информационного наполнения сайтов ВУЗов. Правда, это пока не касается электронных библиотек, да и в целом НИУ. Те в нашей стране, кого к этому не обязывают, как правило, спокойно игнорируют эти ресурсы. В США библиотеки сами беспокоятся о том, чтобы не отставать от времени, и видят в этом необходимое условие своего выживания в меняющейся цифровой среде.

В едином Интернет-пространстве аграрных знаний (ЕИПАЗ) представлены три возможных варианта интеграции научных информационных ресурсов. В первом варианте предполагается перенос в единую базу данных только каталогов информационных массивов. При удачном поиске нужной информации в каталоге пользователь отсылается на сайт и другую базу, где хранится полнотекстовая, либо более подробная информация. При втором варианте вся информация перенесена в единую базу данных. В третьем варианте в единую базу данных перенесены каталоги, а полнотекстовая информация может быть или перенесена в единую базу данных, или же может оставаться в других базах данных, где она была первоначально сформирована. Третий вариант представляет собой наиболее гибкий вариант построения информационного пространства, так как позволяет не переносить лишний раз ту информацию, которая уже представлена в других базах данных, и может быть доступна через ссылки на определенные адреса.

В связи с вышеизложенным, при формировании баз данных научно-исследовательских учреждений аграрно-экономического направления использован, в основном, первый вариант интеграции информационных научных ресурсов, при котором полнотекстовая информация размещенная ранее на сайтах источников не переносится, а в единой базе данных в каталожной карточке имеются ссылки на эти сайты.

Основными подходами по реализации ЕИПАЗ являются создание единого словаря атрибутов (параметров). Создание и поддержка единого словаря параметров системы позволяет привести к единой системе наименований все используемые в информационных

ресурсах (ИР) атрибуты. Особенно это касается атрибутов метаданных, наиболее широко применяемых в различных ИР. При этом появляются возможности объединения и слияния различных ресурсов [22]. Таким образом, единый словарь параметров обеспечивает унификацию разнообразных имен атрибутов во всех программных компонентах ЕИПАЗ.

## **2.2 Организация и логическая структура базы данных в ЕИПАЗ**

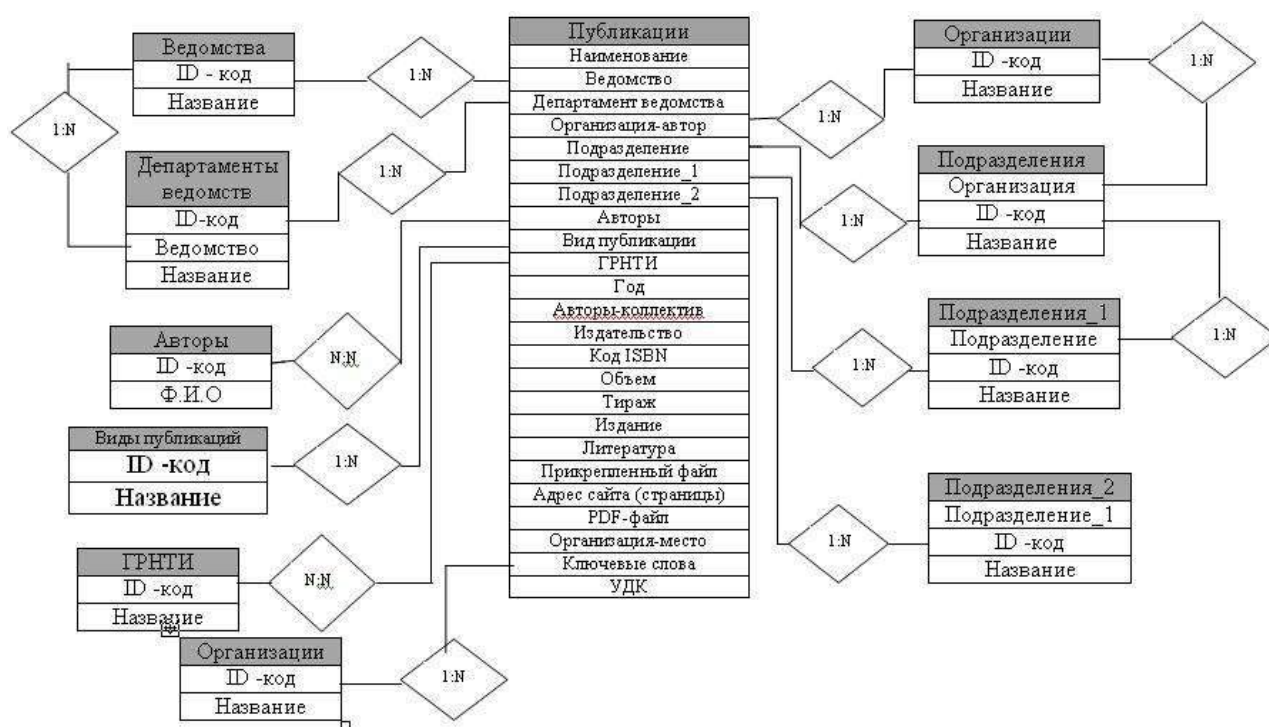
Формирование баз данных информационных научных ресурсов для электронных библиотек «Публикации научных работ» и «Прикладные разработки и инновационные проекты» научно-исследовательских учреждений аграрно-экономического направления выполнено в рамках разработанной ранее единой БД ЕИПАЗ которая реализована на основе информационно-программного комплекса «1С-Битрикс». В дальнейшем для краткости указанные два вида информационных ресурсов называются «Публикации» и «Разработки». Логическая структура «сущность-связь» БД ЕИПАЗ была представлена в [23], в данной работе показана в Приложении 1, как набор реляционных (взаимосвязанных) таблиц данных, каждая из которых описывает определенную сущность (объект) его атрибутами. Данная БД предназначена для размещения динамической информации на серверах провайдера отдельно от страниц сайта в форме взаимосвязанных файлов.

Информация из БД извлекается и выводится на страницы сайта по запросам пользователей программным путем системой управления базы данных (СУБД). Добавление и корректирование информации может выполняться администраторами, имеющими полный допуск, из специальной административной области, или редакторами сайтов, имеющими допуск только к редактированию локальной информации определенной организации из публичной области со страниц сайта.

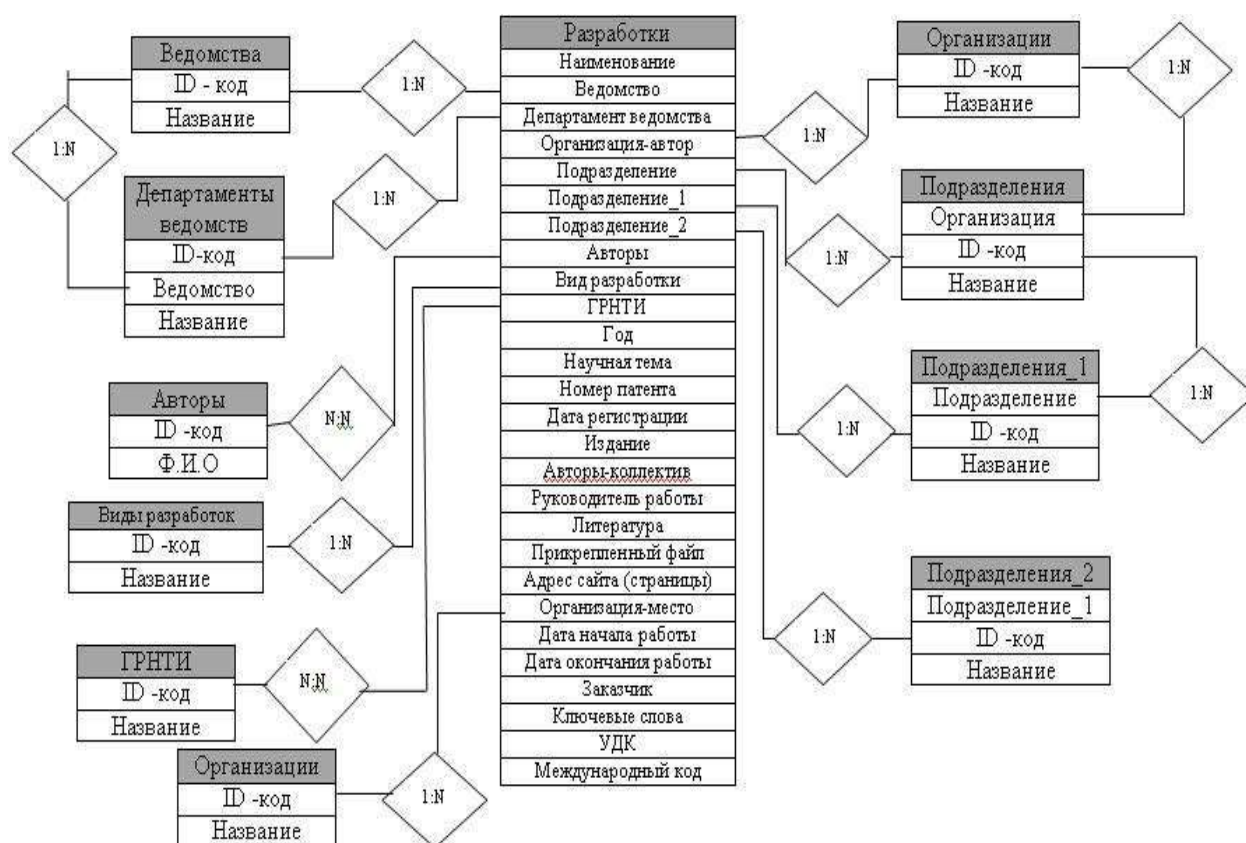
В данном разделе представлена методика формирования баз данных публикаций и разработок научно-исследовательских учреждений аграрно-экономического направления в рамках единой БД ЕИПАЗ. По результатам обследования баз данных источников, представленным ранее в таблице 2.1.1 были сформированы наборы атрибутов объектов информации публикаций и разработок. В этой таблице видно, что все обследованные базы данных имеют мало атрибутов, общих для всех БД. Такими общими атрибутами являются только «Название» и «Вид публикации, разработки». Использование только этих общих атрибутов недостаточно для организации БД с возможностями расширенного поиска по многим атрибутам. В связи со сказанным тезаурус, то есть набор атрибутов, для создаваемых баз данных был составлен путем объединения наиболее существенных атрибутов, имеющихся в представленных в таблице 2.1.1 базах данных, которые выбраны как источники информации.

В созданный тезаурус не включены те атрибуты, которые несущественны для организации поиска, или являются специфическими только для одной из рассматриваемых баз данных. Например, для базы данных ФИПС - это «Адрес для переписки», «Патентный поверенный» и другие; для базы данных ЕГИСУ – это «Качество результата», «Ожидаемый эффект» и другие; для базы данных ELIBRARY – это «Журналы». Эти атрибуты являются дополнительными для конкретных работ. Для просмотра значений полного набора атрибутов конкретной работы в создаваемых базах данных публикаций и разработок введен атрибут «Адрес сайта (страницы)», в котором размещается ссылка на документ в исходной базе данных, где пользователь может просмотреть всю полную информацию о работе.

На рисунках 2.2.1 и 2.2.2 представлены таблицы атрибутов сущностей (объектов), соответственно, публикаций и разработок, где показаны отношения между данными объектами и остальными объектами БД ЕИПАЗ.



**Рис.2.2.1** Логическая структура БД публикаций



**Рис.2.2.2 Логическая структура БД разработок**

Представленная структура связей публикаций и разработок в БД ЕИПАЗ обеспечивает разнообразную и точную «привязку» экземпляров объектов (записей) к их происхождению по местам образования (ведомство, организация, подразделения организации), по авторству, по видам публикаций и разработок, по тематике (ГРНТИ) и т.д. Это обеспечивает возможность быстрого целенаправленного поиска необходимых записей по всей БД.

Особенностью набора атрибутов публикаций и разработок является наличие двух атрибутов (свойств) для привязки к организации: «Организация-автор» и «Организация-место». Свойство «Организация-автор» определяет авторство записи, а в значении свойства «Организация-место» указывается адрес сайта, где размещена информация об этой записи. Значения этих двух полей могут не совпадать, так как допускается возможность опубликования сведения о работах на сайте организации, не являющейся ее автором. Так, например, на сайтах информация о публикации или разработке может быть размещена в каталоге Центральной научной сельскохозяйственной библиотеки (ЦНСХБ), на сайтах научной электронной библиотеки ([www.e-library](http://www.e-library)), единой государственной информационной системы учета научно-исследовательских, опытно-конструкторских и технологических работ гражданского назначения ([www.rosrid.ru](http://www.rosrid.ru)), информационно-

консультационных служб (ИКС), хотя эти организации не являются авторами этих работ. Также могут публиковать «не свои» работы и любые другие организации.

Имеются также два свойства «Авторы» и «Авторы-коллектив». Свойство «Авторы» предназначено для указания авторов, которые авторизованы в ЕИПАЗ, то есть записаны в родительской таблице «Авторы», при «ручном», индивидуальном введении записи в каталог. Это поле заполняется путем выбора персон из таблицы «Авторы». Однако, при организации массового автоматизированного переноса записей в каталог ЕИПАЗ из внешних источников организация выбора фамилий из родительской таблицы «Авторы» не всегда возможна ввиду того, что список авторов работ может быть обширным, и не все авторы могут находиться в списке авторизованных в таблице «Авторы». В этом случае все авторы работ просто переносятся в поле «Авторы-коллектив» как строковые значения. Поэтому поиск работ по авторам в БД организован как в поле «Авторы» при выборе определенной фамилии из родительской таблицы, так и в поле «Авторы-коллектив» путем написания в поисковом поле фамилии с клавиатуры.

Типы полей свойств таблиц «Публикации» и «Разработки» и организация связи с родительскими таблицами в системе ЕИПАЗ описаны в следующем разделе.

## **2.3 Информационная структура баз данных ЕИПАЗ**

### **2.3.1 Реализация БД ЕИПАЗ средствами «1С-Битрикс»**

БД «Публикации» и «Разработки» включены в общую систему ЕИПАЗ и являются частью ее. База данных ЕИПАЗ реализована на основе информационно-программного комплекса «1С-Битрикс». В системе «1С-Битрикс» единая база данных формируется специальными программными конструкциями, называемыми «информационными блоками», являющимися прообразами реляционных таблиц логической структуры БД. Информационные блоки – комплексные модули системы, позволяющие каталогизировать информацию и управлять различными типами однородной информации. С помощью информационных блоков осуществляется размещение, выборка и представление различных типов динамической информации. Сокращенно далее информационный блок может называться – «инфоблок».

В состав модуля инфоблоков входят следующие объекты:

- типы информационных блоков;
- информационные блоки;
- разделы инфоблоков;
- элементы инфоблоков.

На рис. 2.3.1.1 показана иерархическая структура модуля инфоблоков.



**Рис. 2.3.1.1 Иерархия построения модуля инфоблоков**

Верхний уровень представленной иерархии – «Тип инфоблока» включает в себя набор инфоблоков (таблиц), относящихся к одной тематике информации. Информационный блок – это блок записей однородной информации, имеющей единый набор свойств. Это прообраз таблицы, по строкам которой размещены элементы (записи) или входящие в инфоблок наборы записей, объединенные в разделы. Разделы в свою очередь могут иметь конкретные записи или входящие в них разделы следующего уровня. Инфоблоки могут состоять только из разделов, иногда разделы могут образовывать несколько уровней: разделы верхнего, родительского уровня, разделы второго уровня, входящие в разделы родительского уровня и т.д. Обычно по такому принципу создаются многоуровневые справочники.

В таблице 2.3.1.1 представлен список созданных в системе ЕИПАЗ типов инфоблоков, входящих в них инфоблоков, их уникальные идентификаторы (ID), показаны соответствия инфоблоков таблицам логической структуры БД (Приложение 1).

**Таблица 2.3.1.1**

**Типы информационных блоков и входящие инфоблоки**

| Название типа инфоблока | ID типа инфоблока | Инфоблоки                | Символьный код инфоблока | ID инфоблока | Соответствие таблице логической структуры |
|-------------------------|-------------------|--------------------------|--------------------------|--------------|-------------------------------------------|
| Страны и области        | country_region    | Страны                   | country_name             | 84           | Страны                                    |
|                         |                   | Округи                   | okrug                    | 132          | На схеме приложения 1 не показан          |
|                         |                   | Области                  | regions                  | 85           | Регионы                                   |
|                         |                   | Районы                   | district                 | 91           | На схеме приложения 1 не показан          |
| Ведомства               | Vedomctva         | Ведомства высшего уровня | vedomstvo                | 86           | Ведомства высшего уровня                  |

|                        |                      |                                  |                      |     |                                                        |
|------------------------|----------------------|----------------------------------|----------------------|-----|--------------------------------------------------------|
|                        |                      | Департаменты<br>ведомств         | department_vedomstv  | 87  | Подразделения<br>ведомств                              |
| Организации            | organizations        | Организации                      | organization         | 89  | Организации                                            |
|                        |                      | Типы<br>организаций              | tip_organization     | 88  | Типы<br>организаций                                    |
|                        |                      | Подразделения<br>первого уровня  | podrazd_organization | 90  | Подразделения<br>организаций                           |
|                        |                      | Подразделения<br>второго уровня  | podrazd_2            | 118 | На схеме<br>Приложения 1<br>не показаны                |
|                        |                      | Подразделения<br>третьего уровня | podrazd_3            | 119 |                                                        |
| Персоналии             | id_persona           | Персона                          | persona              | 92  | Персоналии                                             |
|                        |                      | Ученые<br>степени                | id_science_stepen    | 94  | На схеме<br>Приложения 1<br>не показаны                |
|                        |                      | Ученые звания                    | id_science_zvanie    | 93  |                                                        |
| Публикации             | id_publication       | Публикации                       | publication          | 95  | Публикации                                             |
|                        |                      | Виды<br>публикаций               | vid_publ             | 97  | Виды<br>публикаций                                     |
|                        |                      | Диссертации                      | dissertation         | 127 | На схеме<br>Приложения 1 не<br>показаны                |
| Разработки             | razrabotki           | Разработки                       | razrabotki           | 98  | Разработки                                             |
|                        |                      | Виды<br>разработок               | vid_razrab           | 131 | Виды<br>разработок                                     |
| Продукция              | Production           | Продажа                          | prodaja              | 101 | Продукция<br>предприятий                               |
|                        |                      | Спрос                            | spros                | 102 | Потребности<br>предприятий                             |
|                        |                      | Виды<br>продукции                | vid_product          | 100 | Виды<br>продукции                                      |
| Работы и<br>услуги     | works                | Предложение<br>работ и услуг     | predlojenie          | 113 | Работы и<br>услуги                                     |
|                        |                      | Спрос на<br>работы и<br>услуги   | works spros          | 112 | Работы и<br>услуги                                     |
|                        |                      | Виды работ и<br>услуг            | works kinds          | 111 | Виды работ и<br>услуг                                  |
| 01 Новости             | news                 | Новости<br>ЕИПАЗ                 | news_eipaz           | 107 | Новости                                                |
| Биржа труда            | labour_exchange      | Вакансии                         | vacancies            | 125 | Вакансии                                               |
|                        |                      | Резюме                           | summary              | 126 | На схеме не<br>показаны                                |
| Эконом.-<br>статистика | Ekon_Stat            | Статистика                       | statistic            | 128 | Приложение 2<br>Таблица<br>«Эконом.-стат.<br>сведения» |
| Мероприятия            | entertainments       | Мероприятия                      | entertainment        | 147 | На схеме не<br>представлена                            |
|                        |                      | Вид<br>мероприятия               | vid_entertainment    | 148 | На схеме не<br>представлена                            |
| ГРНТИ                  | grnti                | ГРНТИ                            | grnti                | 76  | ГРНТИ                                                  |
| Научная<br>подготовка  | science_postgraduate | Научная<br>подготовка            | sci_podgotovka       | 116 | Научная<br>подготовка                                  |
|                        |                      | Формы<br>подготовки              | formi_podgotovki     | 115 | Формы научной<br>подготовки                            |

|                            |                       |                           |                  |     |                                     |
|----------------------------|-----------------------|---------------------------|------------------|-----|-------------------------------------|
|                            |                       | Справочник специальностей | sci_specialnosti | 114 | Номенклатура научных специальностей |
|                            |                       | Диссертационные советы    | dissertation     | 122 | На схеме не представлена            |
| Должности и специальности  | Positions_occupations | Специальности             | occupations      | 110 | Специальности                       |
|                            |                       | Должности                 | positions        | 109 | На схеме не представлена            |
| Нормативы                  | normativ              | Нормативные документы     | norm_dok         | 104 | Нормативно-справочные документы     |
|                            |                       | Виды документов           | vid_dokum        | 103 | Виды НПИ                            |
| Базы данных                | data_bases            | Базы данных               | data bases       | 150 | Разработанные базы данных           |
|                            |                       | Виды БД                   | type data base   | 151 |                                     |
| Пакеты прикладных программ | applied_programs      | Пакеты программ           | applied_programs | 157 | ППП                                 |
|                            |                       | Виды программ             | type program     | 156 |                                     |
| Дистанционное обучение     | distance_education    | Дистанционное обучение    | education        | 155 | Дистанционное обучение              |
|                            |                       | Виды обучения             | type education   | 154 |                                     |

Инфоблоки, входящие в тип «Страны и области», предназначены для отнесения объектов информации (например, организаций) к определенной стране, федеральным округам РФ, областям и районам РФ и других стран при необходимости последующих выборов по этим признакам. Таким образом, имеется возможность расширения географии путем включения в ЕИПАЗ информации других стран. Входящие инфоблоки «Страны», «Округа», «Области» и «Районы» содержат записи, являющиеся дочерними по отношению к записям предыдущего инфоблока.

Тип инфоблока «Ведомства» используется для указания принадлежности носителей информации (например, организаций) к вышестоящим органам: структурам министерств, академическим структурам, если это требуется, или для записи информации, относящейся непосредственно к ведомству или к его департаменту. При этом допускается, что носители информации могут быть самостоятельными, то есть не иметь подчиненности ведомствам высшего уровня. Этот тип инфоблока состоит из двух инфоблоков: «Ведомства высшего уровня» и инфоблока дочерних записей «Департаменты ведомств».

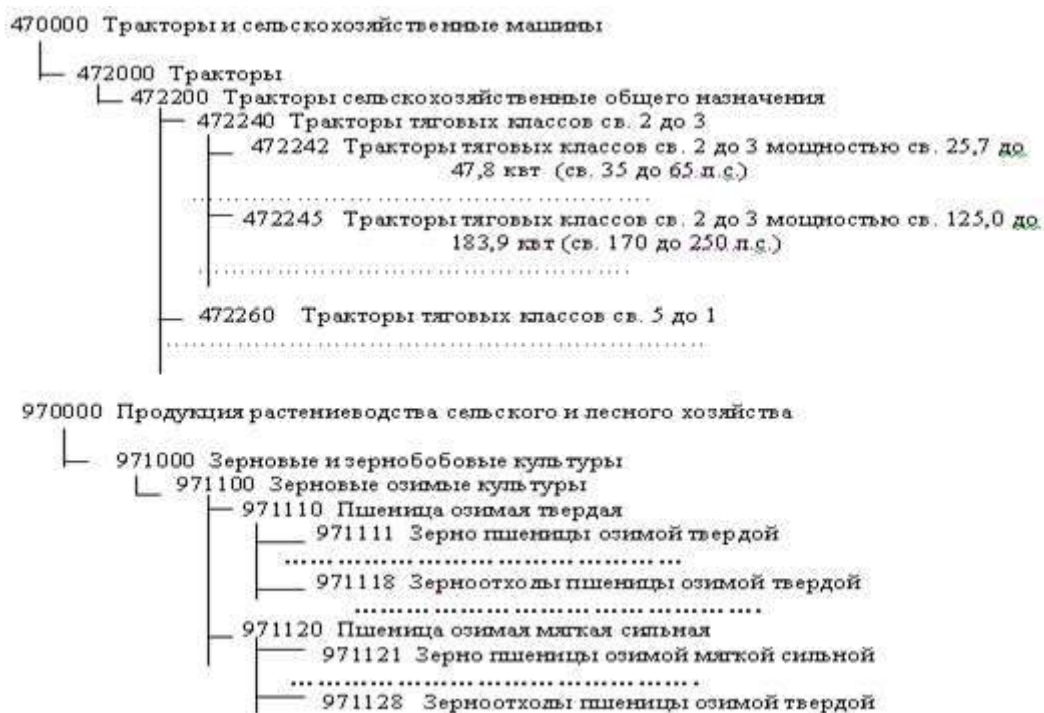
Для типа инфоблока «Организации» предусмотрена возможность многоуровневой структуры. Максимально возможное количество уровней – четыре, которые реализуются инфоблоками в порядке подчиненности: организации – подразделения первого – второго – третьего уровней. Такая схема позволяет реализовать, например, сложную структуру высшего учебного заведения (ВУЗа): университет – институты – факультеты – кафедры (лаборатории). Это применяется, например, при создании единой базы данных



публикаций или разработок ВУЗа, когда можно осуществлять выборку информации по любому подразделению, прописав только один раз в БД информационный объект. Такая структура реализована в инфоблоке родительских записей «Организации» и трех инфоблоков «Подразделения первого уровня», «Подразделения второго уровня» и «Подразделения третьего уровня», записи в которых являются дочерними по отношению к записям предыдущего инфоблока.

В типе инфоблока «Публикации» создан отдельный инфоблок «Диссертации». В этом инфоблоке размещаются записи с атрибутами, которые отражают весь процесс прохождения диссертаций от представления диссертации в ученый совет до защиты. Тем самым атрибуты объекта «диссертации» значительно отличаются от набора атрибутов обычных публикаций. Это можно увидеть непосредственно в настройках свойств того и другого инфоблоков. Защищенные диссертации как публикации затем заносятся в инфоблок «Публикации» с признаком вида публикации «Диссертация». Таким образом, запись о диссертации имеется в обоих инфоблоках.

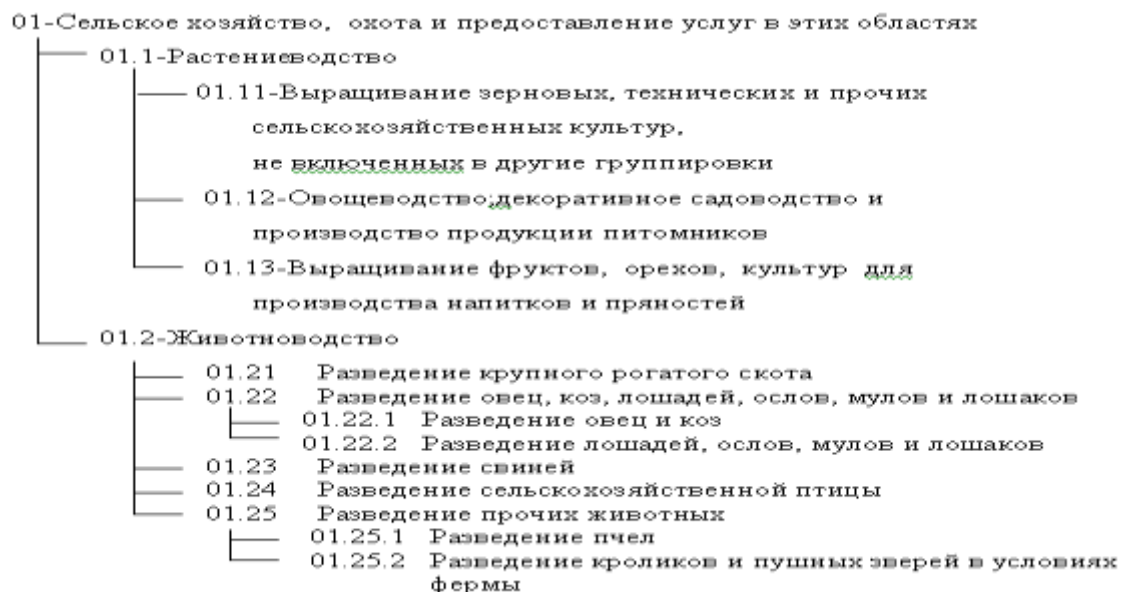
В типе инфоблока «Продукция» представляется информация о продаже и о спросе продукции, для чего созданы отдельные инфоблоки «Продажа» и «Спрос». Это разделение облегчает обработку информации при создании электронной торговой площадки. Для последующей организации целенаправленного поиска записи продукции заносятся с признаком вида продукции в соответствии с «Общероссийским классификатором продукции» (ОКП), реализованном в инфоблоке «Виды продукции» в пятиуровневой структуре. Для примера на рис. 2.3.1.2 показан фрагмент справочника ОКП.



**Рис. 2.3.1.2 Фрагмент иерархической структуры ОКП**

В типе инфоблока «Работы и услуги» информация разделена по инфоблокам «Предложение работ и услуг» и «Спрос на работы и услуги». Записи работ и услуг заносятся с признаком вида деятельности в соответствии с Общероссийским классификатором видов экономической деятельности (ОКВЭД-2016), который реализован в системе в инфоблоке «Виды работ и услуг» в четырехуровневой структуре.

Ниже на рис. 2.3.1.3 показан пример фрагмента справочника видов работ и услуг – Общероссийского классификатора видов экономической деятельности (ОКВЭД).



**Рис. 2.3.1.3 Фрагмент иерархической структуры ОКВЭД**

В типе инфоблока «Биржа труда» имеются два инфоблока: «Вакансии», в котором записываются объявленные организациями вакансии и «Резюме», в котором размещаются записи подаваемых резюме соискателями работы.

Тип инфоблока «Мероприятия» состоит из двух инфоблоков: «Мероприятия» и «Вид мероприятия» предназначен для записи информации о проводимых организациями мероприятиях различных видов, указанных в справочнике «Вид мероприятия».

Тип инфоблока «Научная подготовка» включает в себя четыре инфоблока. В инфоблоке «Научная подготовка» содержатся объявления о научных подготовках, предлагаемых организациями и научными учреждениями, в соответствии с формами научной подготовки, перечисленными в инфоблоке «Формы подготовки». Научная подготовка классифицируется по специальностям, представленным в инфоблоке «Справочник специальностей», созданном на основе «Номенклатуры специальностей научных работников», утвержденной Приказом Министерства образования и науки Российской Федерации от 25 февраля 2009 г. № 59. В инфоблоке «Диссертационные советы» размещается информация о диссертационных советах научных и образовательных учреждениях системы ЕИПАЗ.

Тип инфоблока «Должности и специальности» включает два инфоблока, в которых представлены справочники должностей и специальностей.

Тип инфоблока «Нормативы» имеет в своем составе инфоблок «Нормативные документы», в котором содержатся записи о нормативно-правовых документах различного вида, издаваемых законодательными и руководящими органами РФ, округов, районов, ведомствами и министерствами с указанием авторов этих документов, субъектов, разместивших в системе эти документы, и адресов, где размещены эти документы. Все документы классифицируются в соответствии со справочником, реализованном в инфоблоке «Виды документов».

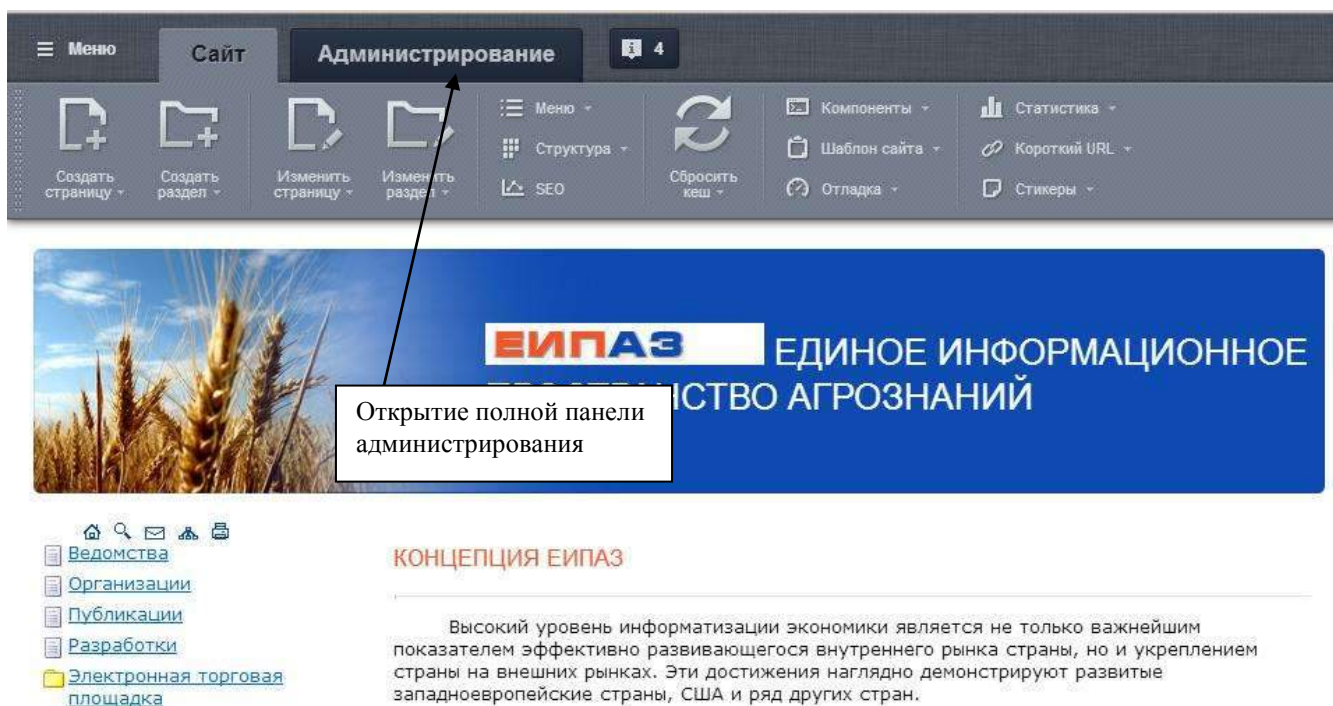
Тип инфоблока «ГРНТИ» в своем единственном инфоблоке «ГРНТИ» содержит записи Государственного рубрикатора научно-технической информации. ГРНТИ используется в системе ЕИПАЗ для классификации публикаций, разработок, области деятельности персонала и области компетенции и консультирования консультантов. Как и сам рубрикатор, инфоблок «ГРНТИ» построен по иерархическому принципу и имеет трехуровневую структуру. Например, 68.00.00 - «Сельское и лесное хозяйство» - 68.01.00 - «Общие вопросы сельского хозяйства» - 68.01.14 – «Коммерческие вопросы, маркетинг, конъюнктура, реклама», и т.д.

Классификаторы ГРНТИ, ОКП и ОКВЭД в системе реализуются в части разделов, которые используются для записи в системе информационных объектов, относящихся к аграрной деятельности, но всегда могут быть дополнены разделами для обозначения и других видов деятельности.

Типы инфоблоков «Базы данных», «Пакеты прикладных программ» и «Дистанционное обучение» предназначены для размещения в соответствующих указанных инфоблоках записей об этих информационных объектах с указанием ссылок доступа к ним, классифицируемых по видам этих объектов. Справочники видов этих объектов размещены в инфоблоках: «Виды БД», «Виды программ» и «Виды обучения».

### 2.3.2 Работа с информационными блоками в «1С-Битрикс»

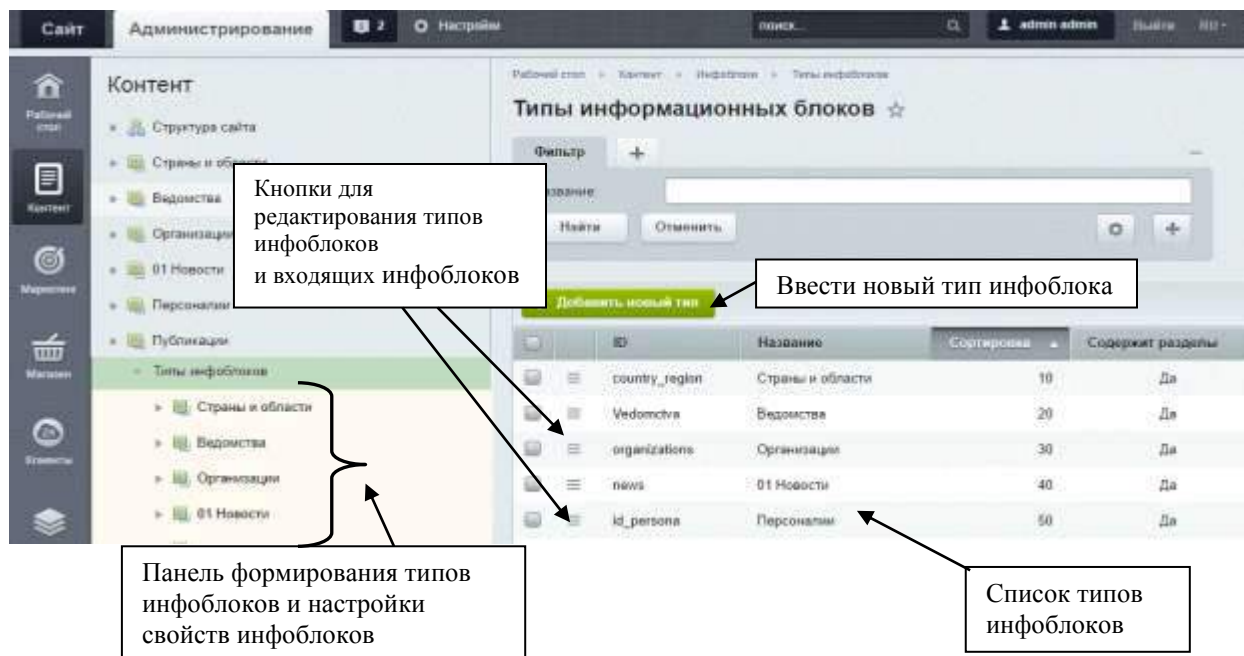
Полный доступ к работе с инфоблоками (создание, настройка свойств и доступа, запись элементов списков и редактирование) имеют авторизованные администраторы. Работа администраторов осуществляется через Административную панель. На панели сайта **www.eipaz.ru** имеются две строки авторизации, в которые вводятся логин и пароль администратора. После авторизации администратора появляется над сайтом Административная панель, верхняя часть которой представлена на рис.2.3.2.1.



**Рис. 2.3.2.1 Административная панель**

На административной панели представлены иконки работы со страницами сайта, с меню, с используемыми компонентами, шаблонами и т.д. При нажатии на вкладку «Администрирование» открывается полная панель с боковой частью. На рис. 2.3.2.2

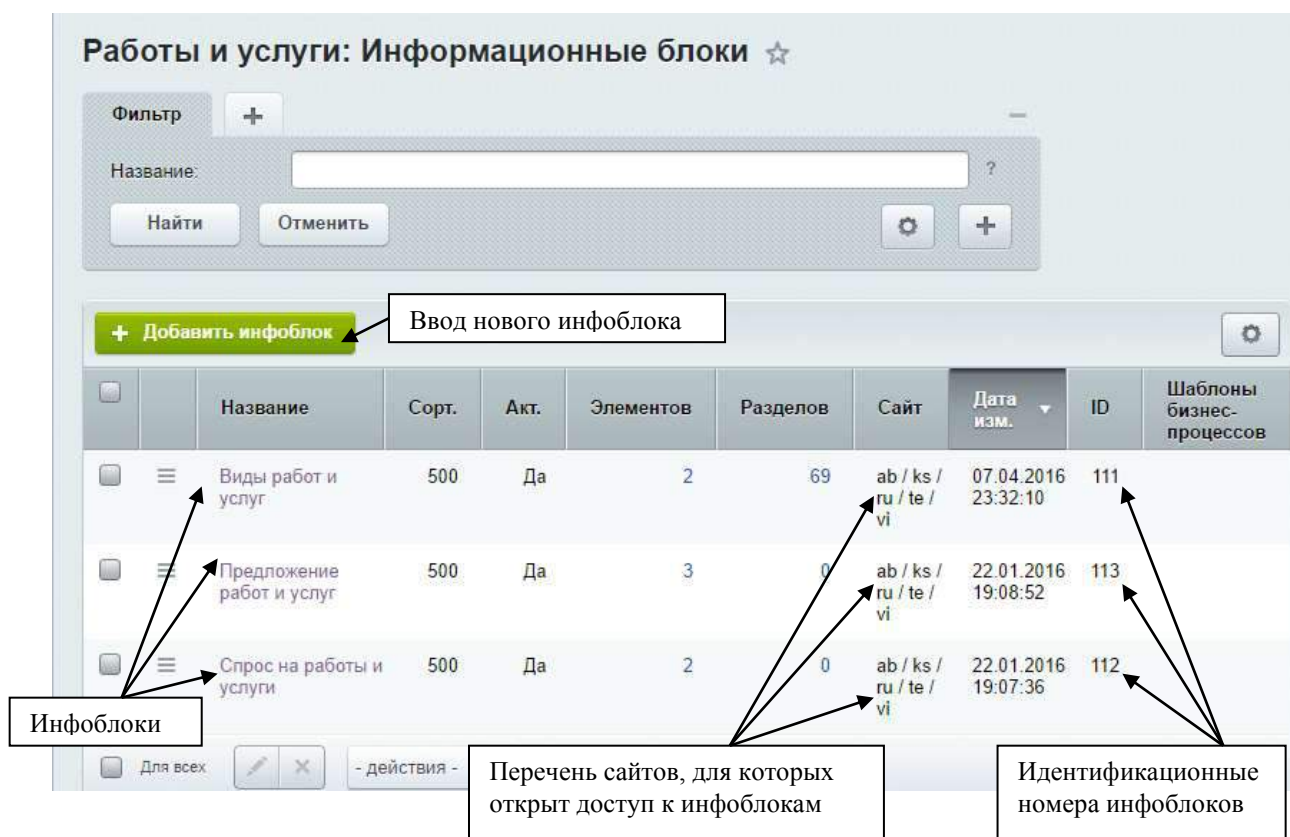
показан вид экрана после выбора на боковой панели иконки «Контент». На средней вертикальной панели «Контент» при нажатии на пункт «Типы инфоблоков» раскрывается список всех типов инфоблоков системы. На панели справа представлены все имеющиеся типы инфоблоков (здесь для сокращения показана только часть инфоблоков от типа инфоблока «Страны и области» до «Публикации»).



**Рис. 2.3.2.2 Список типов инфоблоков**

В этой панели выполняется работа по созданию новых типов инфоблоков, созданию входящих в них инфоблоков и редактированию. При нажатии кнопки для редактирования можно перейти к формированию составляющих инфоблоков и проектированию свойств (полей) инфоблоков, по которым будут вводиться элементы информации в инфоблоки. При раскрытии заголовка «Типы инфоблоков» появляется список имеющихся типов инфоблоков на панели справа. Выбором любого типа инфоблока можно его редактировать. Кнопка «Добавить новый тип» служит для формирования нового типа инфоблока, кнопка редактирования служит для изменения названия типа инфоблока или для его удаления.

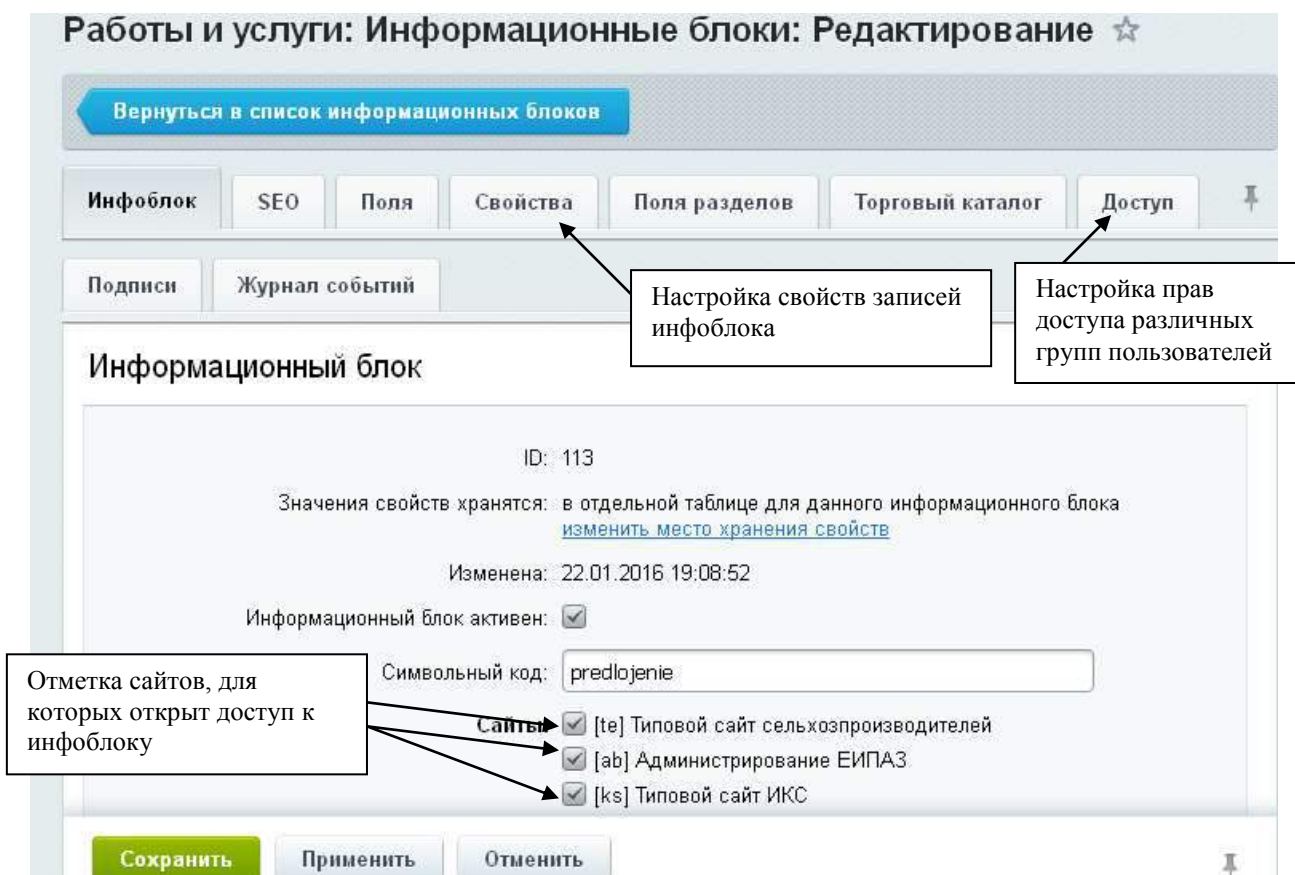
При нажатии на выбранном типе инфоблока на панели «Типы инфоблоков» справа появляется панель ввода, настройки и редактирования инфоблоков, входящих в состав выбранного типа инфоблока. На рис. 2.3.2.3 представлен вид экрана состава типа инфоблока «Работы и услуги»: входящие в него инфоблоки.



**Рис. 2.3.2.3 Состав типа инфоблоков «Работы и услуги»**

В этой таблице можно вводить новые инфоблоки, включаемые в данный тип инфоблоков. Для этого здесь имеется кнопка «Добавить инфоблок». При формировании инфоблока в экранной форме свойств вводится его символьный код, автоматически система присваивает ему идентификационный номер – ID. Показывается также список сайтов, созданных в системе, которым предоставляется доступ к каждому инфоблоку, например, к инфоблоку «Предложение работ и услуг», как показано на рис. 2.3.2.4.

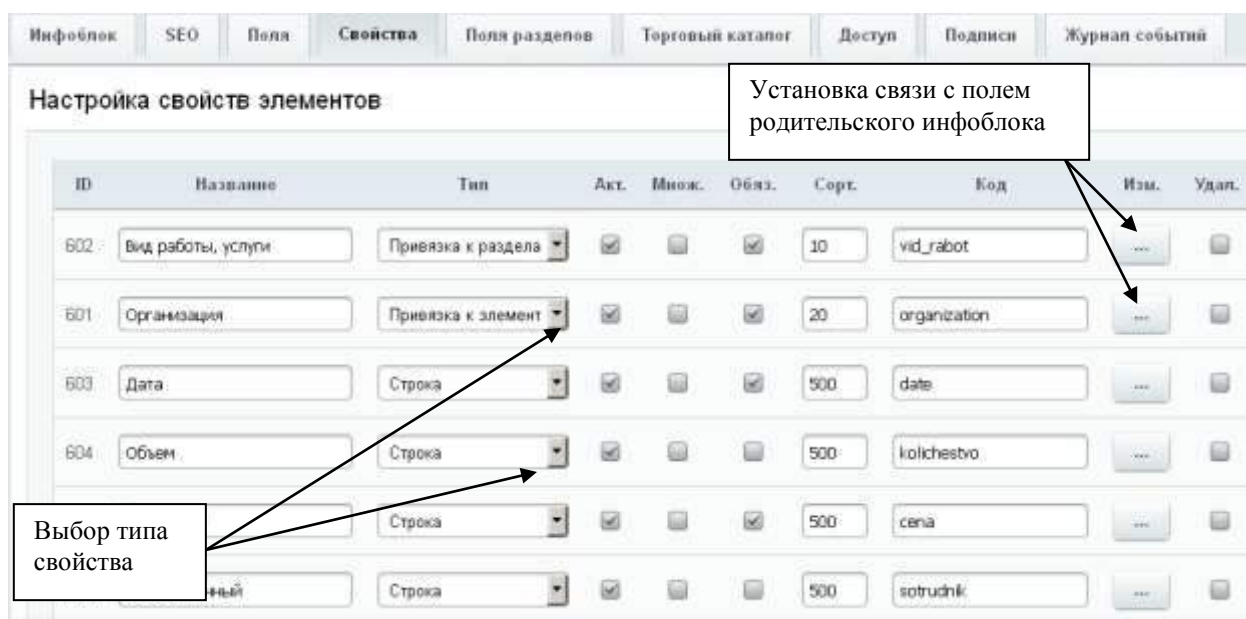




**Рис. 2.3.2.4 Форма представления свойств инфоблока**

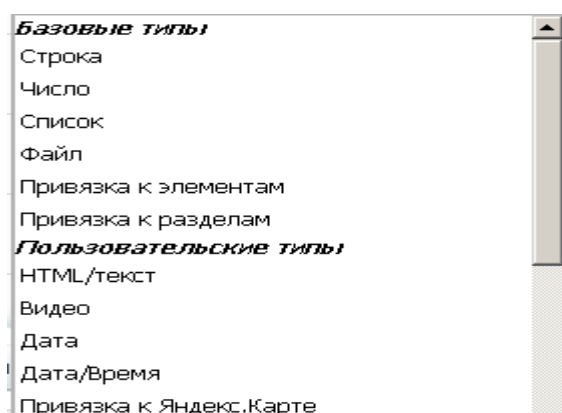
Система «1С-Битрикс» допускает создание многосайтовых приложений, когда с одной базой данных могут работать одновременно несколько сайтов, количество которых, в принципе, не ограничивается. Поэтому в настройках инфоблоков имеется функция регистрации сайтов, которые могут использовать конкретные инфоблоки. На рис. 2.3.2.4 это показано в окне «Отметка сайтов, для которых открыт доступ к инфоблоку».

Затем на закладке «Свойства» выполняется ввод и настройка свойств (полей), по которым будут заноситься в инфоблок элементы информации. Свойства инфоблока соответствуют атрибутам описания предмета информации в таблицах базы данных (рис. 2.2.1, 2.2.2). На рис. 2.3.2.5 показана экранная форма настройки части свойств элементов информации для инфоблока «Предложение работ и услуг». Запись элементов списков инфоблоков начинается с главного, с названия элемента. Название элемента записи – это основной атрибут каждого элемента. Затем уже записываются значения атрибутов элемента.



**Рис. 2.3.2.5 Форма настройки свойств элементов инфоблока**

В колонке «Тип» вводится тип значения свойства, которое описывает вводимый элемент, здесь для выбора раскрывается список возможных типов, показанный на рис. 2.3.2.6.



**Рис. 2.3.2.6 Список выбора возможных типов значений свойств**

Типы «Привязка к элементам» и «Привязка к разделам» устанавливаются в том случае, когда требуется связать значение свойства со значением в родительском инфоблоке. Таким образом, реализуется в системе связь между значениями дочерних и родительских таблиц данных логической структуры БД (Приложение 1). Установка связи с полем родительского инфоблока (таблицы логической структуры БД) выполняется нажатием кнопки в колонке «Изм.» на рис. 2.3.2.5. В результате открывается форма



настройки связи (рис. 2.3.2.7). Здесь показана настройка связи свойства 601 – «Организация» с инфоблоком списка организаций «Организации».

### Настройка свойства инфоблока

ID: 601

Тип: Привязка к элементам

Свойство активно: ☒

Порядок сортировки: 20

Название: Организация

Символьный код: organization

Множественное: ☐

Обязательное: ☒

Выводить на странице списка элементов поле для фильтрации по этому свойству: ☐

Количество полей для ввода новых множественных значений: 5

Подсказка:

Показывать на странице редактирования элемента: ☒

Выбор родительских типа инфоблока и инфоблока

Информационный блок:

Организации [organi]

Организации [89]

Тип инфоблока

Инфоблок

**Рис. 2.3.2.7 Привязка значения поля к родительскому инфоблоку**

Тип «Привязка к элементам» позволяет «связать» определенное свойство записи в некотором инфоблоке со значением свойства в родительском инфоблоке. Например, в записи о сотруднике в инфоблоке «Персона» его свойство «Организация» связать с названием организации из родительского инфоблока «Организации», выбрав организацию, где работает этот сотрудник.

Если родительский классификатор построен по иерархическому принципу, то обычно выбирается тип связи «Привязка к разделам». В этом случае имеется возможность установить привязку одновременно к нескольким разделам в иерархии. Например, Государственный рубрикатор научно-технической информации (ГРНТИ) построен по такому иерархическому принципу, фрагмент его представлен на рис. 2.3.2.8.

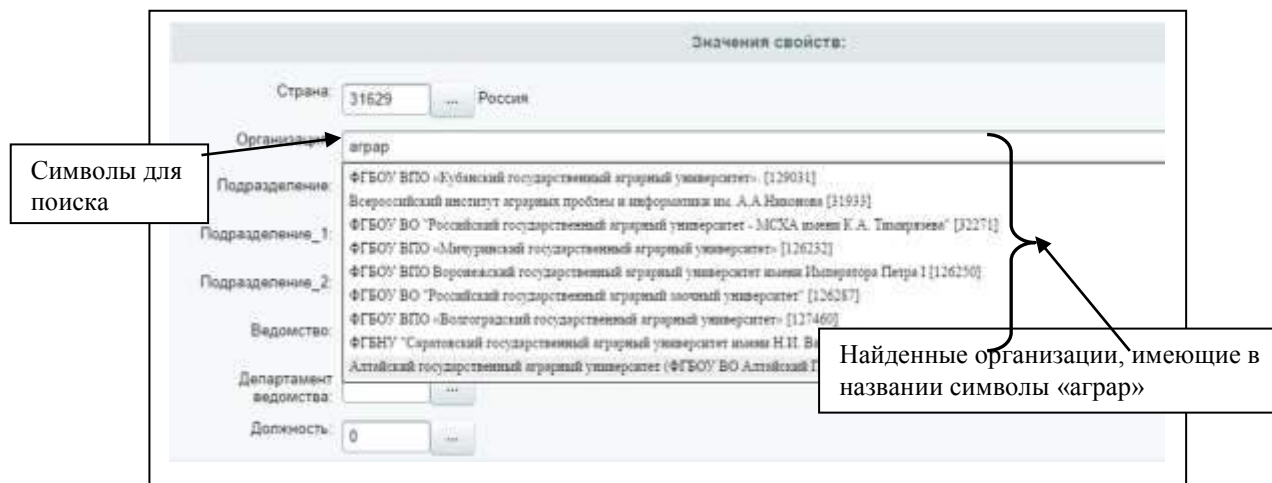
|                                                                                                                                                                          |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 68.00.00 СЕЛЬСКОЕ И ЛЕСНОЕ ХОЗЯЙСТВО                                                                                                                                     |
| 68.75.00 Экономика и организация сельского хозяйства                                                                                                                     |
| 68.75.01 Общие вопросы экономики и организации сельского хозяйства                                                                                                       |
| 68.75.02 Общие экономические проблемы сельского хозяйства                                                                                                                |
| 68.75.13 Аграрная политика. Государственное регулирование сельского хозяйства                                                                                            |
| 68.75.19 Экономика и организация производства в сельскохозяйственных предприятиях различных организационно-правовых форм. Производственная структура сельского хозяйства |
| 68.75.21 Управление, планирование и прогнозирование в сельскохозяйственном производстве                                                                                  |

Рис. 2.3.2.8 Фрагмент ГРНТИ

Для примера, на рис. 2.3.2.8 серым фоном выделены несколько разделов из ГРНТИ, с которыми можно связать свойство персоны «Область деятельности»: общий раздел 68.00.00 и подчиненные, уточняющие разделы 68.75.00 и 68.75.13. В этом случае при задании поиска среди сотрудников эта персона будет выбрана в широкий список сотрудников, работающих в области 68.00.00 «СЕЛЬСКОЕ И ЛЕСНОЕ ХОЗЯЙСТВО», а также в меньший список, работающих в области 68.75.00 «Экономика и организация сельского хозяйства» и в еще меньший, работающих в более узкой области 68.75.13 «Аграрная политика. Государственное регулирование сельского хозяйства».

В некоторых случаях целесообразно использовать тип привязки «Привязка к элементам с автозаполнением». При этом для выбора элемента из родительского инфоблока (списка) необходимо в поле привязки занести некоторое количество символов названия нужного элемента, и тогда в открывающейся таблице будут показаны все элементы, в названии которых имеются внесенные символы. Из этого списка выбирается нужный элемент для занесения в поле дочернего элемента.

Такой тип привязки можно устанавливать в случае, когда родительская таблица состоит из большого количества элементов, и поиск простым перелистыванием будет занимать много времени. На рис.2.3.2.9 показан пример занесения новой персоны в список сотрудников.



**Рис. 2.3.2.9 Выбор места работы (организации) при занесении сотрудника**

Здесь для выбора организации из общего списка организаций в поле «Организация» введены поисковые символы «аграр», и при этом открывается сокращенный список только тех организаций, которые в названии имеют данные символы. Пользователю остается только указать нужную организацию из этого сокращенного списка.

### 2.3.3 Настройка свойств инфоблоков

В данном разделе показаны настройки свойств инфоблоков некоторых типов инфоблоков.

#### 1. Тип инфоблока «Страны и области»

На рис. 2.3.3.1 представлен вид правой панели экрана при выборе типа инфоблока «Страны и области» на левой панели «Типы инфоблоков» (рис. 2.3.2.2 ).



**Рис. 2.3.3.1 Инфоблоки, входящие в тип инфоблока «Страны и области»**

Эти инфоблоки реализуют административно-территориальное деление страны. Так, для России структура трехуровневая: Округ – Области – Районы. Элементы инфоблока «Страна», то есть страны, являются родительскими верхнего уровня по отношению к входящим в страну административным единицам, поэтому не имеют принадлежности к структуре более высокого уровня и, следовательно, не имеют свойств.

Для ввода, редактирования и настройки свойств инфоблоков следует нажать кнопку редактирования, показанную на рис. 2.3.3.1. На рис. 2.3.3.2 показана открытая форма настройки свойств для инфоблока «Округа», который содержит список Федеральных округов страны.

#### Настройка свойств элементов

| ID  | Название | Тип                | Акт.                                | Множ.                    | Обяз.                               | Сорт. | Код        | Изм. | Удал.                    |
|-----|----------|--------------------|-------------------------------------|--------------------------|-------------------------------------|-------|------------|------|--------------------------|
| 822 | Страна   | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | 500   | id_country | ...  | <input type="checkbox"/> |
| 821 | Округ    | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | 500   | id_region  | ...  | <input type="checkbox"/> |

**Рис. 2.3.3.2 Свойства инфоблока «Округи»**

Для инфоблока «Округи» имеется только одно свойство, по которому округа «привязаны» к определенной стране – Название свойства «Страна», код свойства «id\_country». Свойство привязано к элементу из инфоблока «Страны».

На рис. 2.3.3.3 в форме настройки свойств инфоблока «Области» показана «привязка» элементов инфоблока к элементам двух родительских инфоблоков: «Страна» и «Округ», так как области входят в округа (в РФ), которые принадлежат стране.

| ID  | Название | Тип                | Акт.                                | Множ.                    | Обяз.                               | Сорт. | Код        | Изм. | Удал.                    |
|-----|----------|--------------------|-------------------------------------|--------------------------|-------------------------------------|-------|------------|------|--------------------------|
| 477 | Страна   | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | 500   | id_country | ...  | <input type="checkbox"/> |
| 821 | Округ    | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | 500   | id_region  | ...  | <input type="checkbox"/> |

**Рис. 2.3.3.3 Свойства инфоблока «Области»**

На рис. 2.3.3.4 представлена форма настройки свойств инфоблока «Районы». Элементы этого инфоблока в порядке административного подчинения «привязаны» к стране, округу и области. Поэтому здесь уже имеются три свойства.

| Настройка свойств элементов |          |                    |                                     |                          |                                     |       |            |      |                          |
|-----------------------------|----------|--------------------|-------------------------------------|--------------------------|-------------------------------------|-------|------------|------|--------------------------|
| ID                          | Название | Тип                | Акт.                                | Множ.                    | Обяз.                               | Сорт. | Код        | Изм. | Удал.                    |
| 485                         | Страна   | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | 500   | id_country | ...  | <input type="checkbox"/> |
| 486                         | Область  | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | 500   | id_region  | ...  | <input type="checkbox"/> |
| 1007                        | Округ    | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/>            | 500   | okrug      | ...  | <input type="checkbox"/> |

**Рис. 2.3.3.4 Свойства инфоблока «Районы»**

## 2. Тип инфоблока «Организации»

На рис. 2.3.3.5 представлены инфоблоки, составляющие тип инфоблока «Организации»

| Организации: Информационные блоки ☆ |   |                               |       |      |           |          |                        |                     |      |
|-------------------------------------|---|-------------------------------|-------|------|-----------|----------|------------------------|---------------------|------|
| Фильтр                              |   |                               |       |      |           |          |                        |                     |      |
|                                     |   | + Добавить инфоблок           |       |      |           |          |                        |                     |      |
| <input type="checkbox"/>            |   | Название                      | Сорт. | Акт. | Элементов | Разделов | Сайт                   | Дата изм.           | ID ▲ |
| <input type="checkbox"/>            | ≡ | Типы организаций              | 500   | Да   | 8         | 0        | ab / ks / ru / te / vi | 29.03.2016 14:56:01 | 88   |
| <input type="checkbox"/>            | ≡ | Организации                   | 500   | Да   | 233       | 0        | ab / ks / ru / te / vi | 29.03.2016 14:55:48 | 89   |
| <input type="checkbox"/>            | ≡ | Подразделения первого уровня  | 500   | Да   | 82        | 0        | ab / ks / ru / te / vi | 29.03.2016 14:56:24 | 90   |
| <input type="checkbox"/>            | ≡ | Подразделения второго уровня  | 500   | Да   | 4         | 0        | ab / ks / ru / te / vi | 03.02.2015 16:16:38 | 118  |
| <input type="checkbox"/>            | ≡ | Подразделения третьего уровня | 500   | Да   | 0         | 0        | ab / ks / ru / te / vi | 11.04.2016 14:38:03 | 119  |

**Рис. 2.3.3.5 Инфоблоки, входящие в состав типа инфоблока «Организации»**

Инфоблок «Типы организаций» является справочником, в котором записаны возможные типы организаций для возможности выбора организаций по их типам. Инфоблок не имеет свойств, так как записи справочника не имеют атрибутов, а только имеют названия.

На рис. 2.3.3.6 представлен вид экрана при нажатии кнопки «Изменить» в строке инфоблока «Организации» в форме, показанной на рис. 2.3.3.5.

**Настройка свойств элементов**

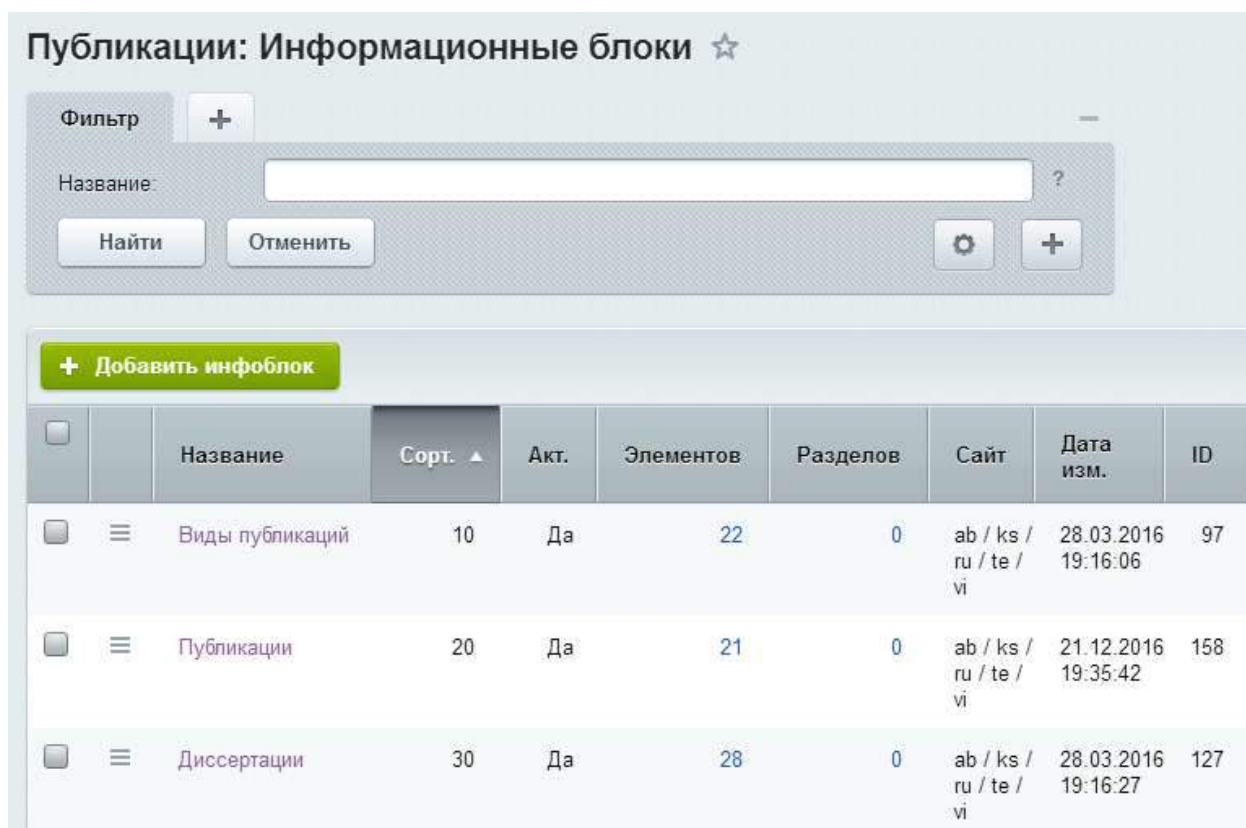
| ID   | Название                | Тип                 | Акт.                                | Множ.                               | Обяз.                               | Сорт. | Код            | Изм. | Удал.                    |
|------|-------------------------|---------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------|----------------|------|--------------------------|
| 481  | Страна                  | Привязка к элемент  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 10    | country        | ...  | <input type="checkbox"/> |
| 1008 | Округ                   | Привязка к элемент  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 15    | okrug          | ...  | <input type="checkbox"/> |
| 483  | Область                 | Привязка к элемент  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 20    | region         | ...  | <input type="checkbox"/> |
| 487  | Район                   | Привязка к элемент  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 30    | id_district    | ...  | <input type="checkbox"/> |
| 482  | Ведомство               | Привязка к элемент  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 40    | vedomstvo      | ...  | <input type="checkbox"/> |
| 484  | Департамент ведомства   | Привязка к элемент  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 50    | division       | ...  | <input type="checkbox"/> |
| 488  | Тип организации         | Привязка к элемент  | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 60    | id_tip         | ...  | <input type="checkbox"/> |
| 806  | Отрасль                 | Привязка к разделу  | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 70    | otrasl         | ...  | <input type="checkbox"/> |
| 492  | Руководитель            | Строка              | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 80    | rukovodit      | ...  | <input type="checkbox"/> |
| 580  | Краткое название        | Строка              | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 85    | name_brief     | ...  | <input type="checkbox"/> |
| 489  | Адрес                   | Строка              | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 90    | adress         | ...  | <input type="checkbox"/> |
| 490  | Электронная почта       | Строка              | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 100   | email          | ...  | <input type="checkbox"/> |
| 491  | Телефоны                | HTML/текст          | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 110   | telefon        | ...  | <input type="checkbox"/> |
| 613  | Ссылка на сайт          | Строка              | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 120   | url_site       | ...  | <input type="checkbox"/> |
| 590  | Логотип организации     | Файл                | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | logotip        | ...  | <input type="checkbox"/> |
| 612  | Ссылка на типовой сайт  | Строка              | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | tipov_url_site | ...  | <input type="checkbox"/> |
| 903  | Владелец (пользователь) | Привязка к пользов- | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 500   | user           | ...  | <input type="checkbox"/> |

**Рис. 2.3.3.6 Свойства инфоблока «Организации»**

Информация по каждой организации записывается в инфоблок с указанием представленных значений свойств записи (атрибутов организации). В верхней части списка свойств указаны свойства «привязки» организации к стране, округу, области, району и т.п. Свойство с ID=806 «Отрасль» устанавливается выбором из Общероссийского классификатора видов экономической деятельности (ОКВЭД), структура которого показана на рис. 2.3.1.3. Это свойство допускает множественность привязки вида деятельности организации к нескольким разделам классификатора.

### 3. Тип инфоблока «Публикации»

На рис. 2.3.3.7 представлены инфоблоки, входящие в тип инфоблока «Публикации».



**Рис. 2.3.3.7 Инфоблоки, входящие в состав типа инфоблока «Публикации»**

Инфоблок «Виды публикаций» является справочником, в котором введены возможные виды публикаций для установления этого признака при введении работ различного характера в список. Справочник видов публикаций можно просмотреть при открытии инфоблока «Виды публикаций» на форме рис. 2.3.3.7. Атрибут «Виды публикаций» используется для выбора из списка публикаций работ определенного типа: статьи, брошюры, книги, отчеты о НИР и т.д.

Список атрибутов (свойств) инфоблока «Публикации» представлен на рис. 2.3.3.8.



## Настройка свойств элементов

| ID   | Название               | Тип                | Акт.                                | Множ.                               | Обяз.                               | Сорт. | Код                | Изм. | Удал.                    |
|------|------------------------|--------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------|--------------------|------|--------------------------|
| 1009 | Авторы                 | Привязка к элемент | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 10    | author             | ...  | <input type="checkbox"/> |
| 1010 | Ведомство              | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 20    | vedomstvo          | ...  | <input type="checkbox"/> |
| 1011 | Организация-автор      | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 30    | organization       | ...  | <input type="checkbox"/> |
| 1012 | Вид публикации         | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 40    | type               | ...  | <input type="checkbox"/> |
| 1013 | Подразделение          | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 50    | podrazdelenie      | ...  | <input type="checkbox"/> |
| 1014 | Подразделение_1        | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 55    | podrazdelenie_1    | ...  | <input type="checkbox"/> |
| 1015 | Подразделение_2        | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 60    | podrazdelenie_2    | ...  | <input type="checkbox"/> |
| 1016 | Департамент ведомства  | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 70    | division           | ...  | <input type="checkbox"/> |
| 1017 | ГРНТИ                  | Привязка к разделу | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 80    | grnti              | ...  | <input type="checkbox"/> |
| 1018 | Издательство           | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 90    | publisher          | ...  | <input type="checkbox"/> |
| 1019 | Страна                 | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 100   | id_country         | ...  | <input type="checkbox"/> |
| 1020 | Год                    | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 110   | year               | ...  | <input type="checkbox"/> |
| 1066 | Авторы-коллектив       | HTML/текст         | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 120   | collectiv          | ...  | <input type="checkbox"/> |
| 1021 | Код ISBN               | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | isbn               | ...  | <input type="checkbox"/> |
| 1022 | Объем                  | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | volume             | ...  | <input type="checkbox"/> |
| 1023 | Тираж                  | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | tiraj              | ...  | <input type="checkbox"/> |
| 1024 | Ключевые слова         | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | keywords           | ...  | <input type="checkbox"/> |
| 1025 | Литература             | Привязка к элемент | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 500   | citation           | ...  | <input type="checkbox"/> |
| 1026 | Описание для SEO       | HTML/текст         | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | DESCRIPTION        | ...  | <input type="checkbox"/> |
| 1027 | Заголовок для SEO      | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | TITLE              | ...  | <input type="checkbox"/> |
| 1028 | Прикрепленный файл     | Файл               | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | file               | ...  | <input type="checkbox"/> |
| 1029 | Адрес сайта (страницы) | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | internet           | ...  | <input type="checkbox"/> |
| 1030 | Издание                | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | izdanie            | ...  | <input type="checkbox"/> |
| 1031 | PDF-файл               | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | a_link             | ...  | <input type="checkbox"/> |
| 1032 | Организация-место      | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | organization_place | ...  | <input type="checkbox"/> |
| 1069 | УДК                    | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | udk                | ...  | <input type="checkbox"/> |

Рис. 2.3.3.8 Свойства инфоблока «Публикации»



Свойства «Описание для SEO» и «Заголовок для SEO» используются для записи ключевых символов, определяющих тематику публикации. SEO – Search Engine Optimization (поисковая оптимизация). Чем точнее и полнее значения этих полей используются при поиске материалов в Интернете через браузеры, тем «выше» в списке найденных он будет находиться. Поиск в системе ЕИПАЗ построен по другим полям, поэтому эти свойства не представляют собой большого значения в этой системе.

Свойство «Прикрепленный файл» типа “file” предназначено для указания файла из библиотеки системы, если требуется добавить какую-либо иллюстрацию к публикации. Свойство «PDF-файл» используется для подключения текстового файла. Свойство «Адрес сайта (страницы)» используется в том случае, когда текст публикации физически размещен на сайте у собственника этой публикации, в этом случае нет необходимости переносить текст в хранилище к провайдеру, где расположена основная БД. Этим обеспечивается распределение информационных ресурсов по источникам их образования.

Особенностью набора свойств этого инфоблока является наличие двух полей (свойств) для привязки к организации: «Организация-автор» и «Организация-место». Свойство «Организация-автор» определяет авторство публикации, а в значении свойства «Организация-место» указывается адрес сайта, где размещена информация об этой публикации. Значения этих двух полей могут не совпадать, так как допускается возможность опубликования сведения о работе на сайте организации, не являющейся ее автором. Так, например, на сайтах информационно-консультационных служб (ИКС) размещаются сведения о работах организаций определенного региона, хотя сама ИКС не является автором этих работ. Также могут публиковать «не свои» работы и любые другие организации. Это расширяет возможность публикации работ в интернете.

#### 4. Тип инфоблока «Разработки»

На рис. 2.3.3.9 представлены инфоблоки, входящие в тип инфоблока «Разработки». Это инфоблоки «Разработки» и «Виды разработок».

Разработки: Информационные блоки ☆

Фильтр

+

Название:

?

Найти

Отменить

⚙

+

+ Добавить инфоблок

| <input type="checkbox"/> |   | Название        | Сорт. | Акт. | Элементов | Разделов | Сайт                   | Дата изм. ▾         | ID  |
|--------------------------|---|-----------------|-------|------|-----------|----------|------------------------|---------------------|-----|
| <input type="checkbox"/> | ≡ | Разработки      | 500   | Да   | 1024      | 0        | ab / ks / ru / te / vi | 13.02.2017 16:58:12 | 159 |
| <input type="checkbox"/> | ≡ | Виды разработок | 500   | Да   | 40        | 0        | ab / ks / ru / te / vi | 28.03.2016 19:15:34 | 131 |

**Рис. 2.3.3.9 Инфоблоки, входящие в состав типа инфоблока «Публикации»**

Список атрибутов (свойств) инфоблока «Разработки» представлен на рис. 2.3.3.10

## Настройка свойств элементов

| ID   | Название               | Тип                | Акт.                                | Множ.                               | Обяз.                               | Сорт. | Код                | Изм. | Удал.                    |
|------|------------------------|--------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------|--------------------|------|--------------------------|
| 1038 | Авторы                 | Привязка к элемент | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 15    | author             | ...  | <input type="checkbox"/> |
| 1039 | Организация-автор      | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 20    | organization       | ...  | <input type="checkbox"/> |
| 1040 | Вид разработки         | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 25    | vid_razrab         | ...  | <input type="checkbox"/> |
| 1041 | Подразделение          | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 30    | podrazdelenie      | ...  | <input type="checkbox"/> |
| 1042 | Подразделение_1        | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 35    | podrazdeleni_1     | ...  | <input type="checkbox"/> |
| 1043 | Подразделение_2        | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 37    | podrazdelenie_2    | ...  | <input type="checkbox"/> |
| 1044 | Ведомство              | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 40    | vedomstvo          | ...  | <input type="checkbox"/> |
| 1045 | Департамент ведомства  | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 50    | division           | ...  | <input type="checkbox"/> |
| 1046 | ГРНТИ                  | Привязка к разделу | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | 80    | grnti              | ...  | <input type="checkbox"/> |
| 1047 | Литература             | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 90    | literatura         | ...  | <input type="checkbox"/> |
| 1048 | Год                    | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 100   | year               | ...  | <input type="checkbox"/> |
| 1064 | Авторы-коллектив       | HTML/текст         | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 120   | collectiv          | ...  | <input type="checkbox"/> |
| 1049 | Ключевые слова         | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 200   | key_words          | ...  | <input type="checkbox"/> |
| 1050 | Прикрепленный файл     | Файл               | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 300   | file               | ...  | <input type="checkbox"/> |
| 1051 | Адрес сайта (страницы) | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 400   | internet           | ...  | <input type="checkbox"/> |
| 1052 | Организация-место      | Привязка к элемент | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 500   | organization_place | ...  | <input type="checkbox"/> |
| 1063 | Издание                | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | izdanie            | ...  | <input type="checkbox"/> |
| 1065 | Научная тема           | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | theme              | ...  | <input type="checkbox"/> |
| 1067 | Номер патента          | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | numpatent          | ...  | <input type="checkbox"/> |
| 1068 | Дата регистрации       | Дата               | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | dateend            | ...  | <input type="checkbox"/> |
| 1070 | Руководитель работы    | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | rukovodit          | ...  | <input type="checkbox"/> |
| 1071 | УДК                    | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | udk                | ...  | <input type="checkbox"/> |
| 1072 | Дата начала работы     | Дата               | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | datestart          | ...  | <input type="checkbox"/> |
| 1073 | Заказчик               | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | employer           | ...  | <input type="checkbox"/> |
| 1074 | Международный код      | Строка             | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   | worldcode          | ...  | <input type="checkbox"/> |
| 1075 | Дата окончания работы  | Дата               | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            | 500   |                    | ...  | <input type="checkbox"/> |

Рис. 2.3.3.10 Свойства инфоблока «Разработки»

Набор свойств инфоблока «Разработки» в основном такой же, как и для инфоблока «Публикации» за исключением нескольких специфических свойств, таких как «Научная тема», «Номер патента» и др., свойственных для этого типа записей.

Те же принципы размещения информации (наличие свойств «Организация-автор» и «Организация-место») используются и для других инфоблоков: разработок, нормативно-правовых документов, продукции, работ и услуг, баз данных, прикладных программ, дистанционного обучения. Для инфоблока «Персона» также введен атрибут «Организация-место» ввиду того, что человек, помимо работы в основной организации, может являться консультантом и может быть занесен как консультант в справочник консультантов на сайте информационно-консультационной службы, для чего эта ИКС должна быть указана в значении поля «Организация-место».

Настройки и составы всех остальных типов инфоблоков можно посмотреть непосредственно через административную панель на экране в соответствии с описанными выше процедурами работы.

## **2.4 Сбор данных в интернете для формирования баз данных публикаций и разработок**

Проблемы сбора информации в Интернете (в англоязычных источниках за этой областью программирования прочно закрепилось название «Web scraping») возникла практически одновременно с возникновением Интернета.

Однако в последнее время эта область программирования получила значительное развитие. Причин такому стремительному развитию этой сферы несколько:

- продолжающийся экспоненциальный рост объемов и скоростей обработки информации в Интернете;
- появление и развитие ряда специализированных и универсальных языков обработки данных;
- появление и развитие средств работы с «большими данными»;
- появление и развитие средств «параллельного» программирования;
- новые форматы и средства хранения, передачи и обработки данных;
- и т.п.

Таким образом, проблема извлечения данных из Интернета превратилась в достаточно широкую и быстро развивающуюся отрасль программирования, средства и методы которой постоянно совершенствуются и достойны изучения, как для теоретического, так и для практического применения.

Традиционно проблема «извлечения» (scraping) данных из Интернета делится на ряд стандартных (типовых) этапов:

- Поиск релевантной (адекватной) информации;
- Анализ структуры найденных данных;
- Поиск и/или создание средств извлечения необходимых данных;
- Извлечение и очистка необходимых данных;
- Запись данных в удобном для их последующего использования формате.

Опишем ниже каждый из этих этапов подробнее и в применении к нашей конкретной задаче: извлечению данных по публикациям и разработкам для сайтов аграрных НИУ экономической направленности и сайтов НИУ Тамбовской области.

#### **2.4.1 Поиск релевантной (адекватной) информации**

В соответствии с постановкой задачи был составлена следующая «поисковая матрица» (таб. 2.4.1.1) организаций, их сайтов и информации, размещенной на сайтах различных источников.

Следует заметить, что ряд источников требовал последующей дополнительной обработки. Так, например, далеко не все из 12625 разработок и публикаций, найденных на сайте ВНИИЭСХ, требовались для решения нашей задачи.

Ряд источников также пересекается по содержанию. Так, например, публикации на сайтах и в РИНЦ обычно довольно сильно пересекаются (но никогда не совпадают!)

Эта проблема решалась на следующем этапе, который описан далее. Окончательно дубликаты публикаций и разработок находились по полному совпадению названий, и в базу данных заносился только один экземпляр из всех обнаруженных дубликатов.

Таблица 2.4.1.1

## Источники для сбора информации по публикациям и разработкам НИУ аграрного профиля экономической направленности

| Ном.<br>п/п | Название организации                                                                                                                      | Сайт                                                                            | Публикации |           | Разработки |                    | Патенты  |             |         |                       |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|------------|-----------|------------|--------------------|----------|-------------|---------|-----------------------|
|             |                                                                                                                                           |                                                                                 | На сайте   | e-library | На сайте   | ЕГИСУ<br>rosrid.ru | На сайте | patscape.ru | fips.ru | ru.espace<br>net.com/ |
| 1           | Сибирский научно-исследовательский институт экономики сельского хозяйства                                                                 | <a href="http://www.sibniiesh.sorashn.ru/">http://www.sibniiesh.sorashn.ru/</a> | 138        | 278       | 81         | 22                 | -        | -           | -       | -                     |
| 2           | Поволжский научно-исследовательский институт экономики и организации агропромышленного комплекса                                          | <a href="http://niisar.ru/">http://niisar.ru/</a>                               | 90         | 702       | -          | -                  | -        | -           | -       | -                     |
| 3           | Всероссийский научно-исследовательский институт экономики и нормативов                                                                    | <a href="http://vniiein.ru/">http://vniiein.ru/</a>                             | 71         | 2039      | 9          | 17                 | -        | -           | -       | -                     |
| 4           | Северо-Западный научно-исследовательский институт экономики и организации сельского хозяйства                                             | <a href="http://szniiesh.ru/">http://szniiesh.ru/</a>                           | 10         | 975       | 0          | 8                  | -        | -           | -       | -                     |
| 5           | Научно-исследовательский институт экономики и организации агропромышленного комплекса Центрально-Черноземного района Российской Федерации | <a href="http://www.niieoapk.ru">http://www.niieoapk.ru</a>                     | 34         | 861       | 0          | 35                 | -        | -           | -       | -                     |
| 6           | Всероссийский институт аграрных проблем и информатики имени А.А. Никонова                                                                 | <a href="http://www.viapi.ru">www.viapi.ru</a>                                  | 2637       | 2694      | -          | 66                 | -        | -           | -       | -                     |

|              |                                                                                                                         |                                                                    |             |              |              |            |            |             |            |           |
|--------------|-------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------|-------------|--------------|--------------|------------|------------|-------------|------------|-----------|
| <b>7</b>     | Всероссийский научно-исследовательский институт организации производства, труда и управления в сельском хозяйстве       | <a href="http://www.vniioptush.ru">http://www.vniioptush.ru</a>    | 105         | 4058         | -            | 65         | -          | 44          | -          | -         |
| <b>8</b>     | Всероссийский научно-исследовательский институт экономики сельского хозяйства                                           | <a href="http://vniiesh.ru">http://vniiesh.ru</a>                  | 128         | 4115         | 12625        | 133        | -          | -           | -          | -         |
| <b>9</b>     | ФГБНУ "Дальневосточный научно-исследовательский институт экономики агропромышленного комплекса"                         | сайта нет                                                          | -           | 30           | -            | -          | -          | 292         | -          | -         |
| <b>10</b>    | ФГБ образовательное учреждение высшего профессионального образования "Мичуринский государственный аграрный университет" | <a href="http://www.mgau.ru">www.mgau.ru</a>                       | 31          | 3994         | 70           | 78         | 171        | 114         | 126        | 46        |
| <b>11</b>    | ГНУ «Всероссийский НИИ садоводства им. И.В. Мичурина» ФГБНУ "ФНЦ им. И.В. Мичурина"                                     | <a href="http://vniismich.ru">http://vniismich.ru</a>              | -           | 888          | 20           | 3          | -          | 555         | 9          | -         |
| <b>12</b>    | ГНУ «Всероссийский НИИ генетики и селекции плодовых растений им. И.В. Мичурина»                                         | <a href="http://www.vniigispr.narod.ru">www.vniigispr.narod.ru</a> | -           | 533          | 15           | 115        | -          | 3           | 3          | -         |
| <b>ИТОГО</b> |                                                                                                                         |                                                                    | <b>3244</b> | <b>21167</b> | <b>12820</b> | <b>542</b> | <b>171</b> | <b>1008</b> | <b>138</b> | <b>46</b> |

#### 2.4.2 Анализ структуры найденных данных

Как и следовало ожидать, представление одинаковой по смыслу информации на разных сайтах было совершенно разным. Даже база данных разработок на сайте ВНИЭСХ на проверку оказалась плохо упорядоченным массивом из 12625 (не всегда одинаково отформатированных!) страничек типового для Интернета HTML-формата.

Информацию на других сайтах анализируемых НИУ также можно было найти в самых разных форматах:

- HTML-формате типичных Web-страниц;
- PDF-формате – универсальном формате публикации документов от Adobe;
- DOC-формате документов Word от Microsoft;
- XLS-формате электронных таблиц Excel от Microsoft.

Ни на одном из обследованных файлов не встретилось файлов с информацией в стандартных и популярных для Интернета форматах передачи данных, таких как XML или JSON. Типовое API для целей извлечения информации на сайтах НИУ также отсутствует.

Проблема сбора информации, однако, облегчалась тем обстоятельством, что и публикации и разработки на сайтах исследуемых НИУ представлены весьма фрагментарно и могут быть достаточно быстро извлечены стандартными средствами, без необходимости написания для этого специальных программ. Исключения составили разработки на сайте ВНИИЭСХ, большое количество которых (12625) и их слабая структурированность, потребовали написания специальной программы для извлечения, очистки и удобного представления извлечённых данных для их последующего использования во вновь создаваемой базе данных публикаций и разработок НИУ аграрного профиля.

На сторонних сайтах с публикациями и разработками, информацию о которых можно найти в приведённой выше таблице 2.4.1.1 типовых API для комфортного (программируемого) извлечения данных с этих сайтов не обнаружено. Ряд сайтов предлагали данные в рамках, так называемой, политики «открытых данных». Однако эти «открытые данные» были, как правило, весьма фрагментарны, что лишало смысла их использование. Так, например, на сайте ЕГИСУ можно найти открытые данные по разработкам в формате XML. Однако, эти данные представлены только за 2016 год, что лишает смысла их использования для наших целей.

Не найдено было также какого-либо удобного API на сайте РИНЦ (e-library.ru). Более того, этот (коммерческий) сайт открыто и письменно на страницах своего сайта угрожает репрессиями (блокировкой и отключением без предупреждения) всем



желающим извлекать информацию со страниц этого сайта программным способом. Таким образом, здесь возникла не столько программистская, сколько морально-этическая проблема с извлечением нужной нам информации с нужных нам страниц.

#### **2.4.3 Поиск и/или создание средств извлечения необходимых данных**

Программных средств и инструментов извлечения информации в Интернете в настоящее время можно найти довольно много. По уровню разработки они ранжируются от простых вспомогательных утилит и плагинов браузеров для помощи в извлечении данных с конкретной странички до достаточно мощных и хорошо развитых программных комплексов для масштабного извлечения информации со страниц сайтов на определённую тематику, в необходимом объёме и временной последовательности. Примеры некоторых таких средств будут приведены ниже.

Разумеется, среди таких средств извлечения информации из Интернета встречаются как бесплатные разработки, так и коммерческие (платные) разработки. В большинстве своём, однако, эти разработки бесплатные.

Наконец, для решения задач по извлечению очень больших массивов данных (так называемых «больших данных») требуется привлечение средств работы с такими данными и, как следствие, средств параллельной обработки таких данных. Для наших целей привлечения таких средств не потребовалось в силу (сравнительно) небольшого объёма собранных данных.

Отдельного рассмотрения достойна тема использования языков программирования для решения данных задач. Здесь, в силу достаточно широкой постановки первоначальной задачи, ограничения на использование конкретных языков практически отсутствуют.

Однако в последнее время, наиболее популярными в данной сфере стали язык программирования Python и программная среда R. Причин такого положения дел несколько. Так, например, для обоих языков существует несколько (!) средств и комплексов решения задач извлечения данных из Интернет. Все эти средства (что немаловажно!) бесплатны и достаточно просты в настройке и использовании. Появление большого массива учебно-практической литературы на эту тему также добавляет популярности упомянутым решениям и языкам.

В практических целях извлечения информации с сайтов НИУ по публикациям и разработкам выбор был сделан в пользу программной среды R. Полные тексты комплекса программ для извлечения информации с сайта ВНИИЭСХ, написанных, отлаженных и исполненных в среде R, приведены в Приложении 2.

Аналогичный комплекс программ был написан для извлечения информации с сайта ЕГИСУ. Однако для экономии места он в отчёте не приводится.

В случае, когда написание соответствующего программного кода по каким-либо причинам (морально-этическим или с целью экономии времени и сил) не осуществлялось, информация с соответствующих страниц извлекалась при помощи двух плагинов для браузера:

Плагина **Selectorgadget** для полуавтоматического формирования необходимого XPath – паттерна (<https://chrome.google.com/webstore/detail/selectorgadget/mhjhnkcfbdhnjickkkdbjoemdmdbfginb>)

Плагина **XPath Helper** для извлечения информации в соответствии с заданным XPath-паттерном (<https://chrome.google.com/webstore/detail/xpath-helper/hgimnogjllphhhkhlmebbmlgjojdpjl>).

Процедуры поиска информации для извлечения данных и их извлечения при помощи двух вышеупомянутых плагинов описаны ниже.

#### 2.4.4 Извлечение и очистка необходимых данных

В случае, когда извлечение данных проводится программным способом естественно здесь же, в программе, произвести очистку и последующую запись очищенных данных в удобной для их последующей обработке форме. В тексте программ, приведённых в Приложении 2, эти этапы обозначены достаточно точно и к ним даны комментарии.

В случае, когда данные со страниц извлекаются «вручную» (постранично и при помощи определённых сервисов и средств), данный этап превращается в отдельную процедуру.

Процедуры очистки данных не поддаются какой-либо полной систематизации. Однако можно выделить ряд этапов и процедур, которые встречаются достаточно часто и регулярно. К таковым относятся:

- преобразование чисел и дат;
- удаление системной и технической информации (специальных символов форматирования и печати, пустых строк и символов, табуляторов, переносов строк и т.п.);
- поиск в извлечённом тексте необходимых фрагментов и их преобразование к нужному формату.

В случае, когда написание соответствующих программ было по той или иной причине невозможно, данные с нужных страниц, как уже было упомянуто выше, извлекались при помощи плагинов Selectgadget и XPath Helper. Первый плагин предназначен для поиска

нужного xPath – паттерна. Второй плагин, получая в своём входном окне заданный xPath – паттерн, выдаёт в другом своём рабочем окне все подходящие под этот паттерн текстовые блоки открытой страницы. Эти тексты извлекались из данного окна и в последующем обрабатывались в текстовых редакторах и/или электронных таблицах. Пример извлечения конкретной информации с конкретной страницы в Интернете приведён в Приложении 2.

#### **2.4.5 Запись данных в удобном формате для их последующего использования**

Данный этап – в случае программного решения задачи извлечения данных из Интернета – также является органической частью соответствующего комплекса задач и следует сразу же за полным выполнением предыдущего этапа. В противном случае – когда данные извлекаются из Интернета «вручную» – данный этап выделяется в отдельный. В первом случае, конечно, формат вывода ограничивается только программистскими способностями. Во втором случае, как правило, речь идёт о каких-то достаточно простых форматах. В нашем случае было принято решение формировать файлы вывода в виде файлов CSV-формата – так называемых форматах вывода данных «разделённых запятой» (от англ. Comma-Separated Values — значения, разделённые запятыми). Данный формат не только прост в понимании и формировании, но и распознаётся практически всеми популярными программами работы с табличными данными, в том числе и самой популярной из них – Excel.

В последующей обработке данных, которая описана ниже, входные данные как раз предполагалось получать в формате файлов Excel. Таким образом, решение об использовании файлов csv-формата, как выходных файлов на данном этапе, было достаточно логичным и очевидным.

Заметим, наконец, что предложенный выше алгоритм, конечно, не претендует на законченность и строгую обусловленность. Как уже указывалось, существует огромное количество вариантов решения данной конкретной задачи – извлечения конкретной информации с сайтов аграрной тематики, и многие из описанных выше этапов носят творческий характер и с большим трудом подвергаются формализации. Здесь приведён только один из возможных методов. Этот метод, однако, достаточно эффективен, и, что самое главное, приводит к хорошим практическим результатам (Приложение 2).

## 2.4.6 Пример извлечения информации плагинами Selectorgadget и XPath Helper

Извлечём со страницы поиска ЕГИСУ <https://www.rosrid.ru/search> (см. скриншот этой страницы – рис. 2.4.6.1 - ниже) названия разработок, которые на ней представлены (всего таковых на странице по умолчанию десять штук).

Для этого установим и запустим плагин Selectorgadget. Сразу же после запуска этого плагина появляется возможность выделить мышью нужный объект на странице. В нашем случае, мы выделяем в первую очередь, конечно, название первой записи с разработками. Результат этой операции приводится ниже (рис. 2.4.6.2). Зелёным цветом подсвечен элемент, который мы выделили – заголовок первой записи каталога. Жёлтым цветом выделены все подходящие элементы. Как видно, вместе с нужными нам десятью заголовками статей, выделился также и ряд ненужных элементов (в правом нижнем углу видно, что всего выделено 24 элемента).

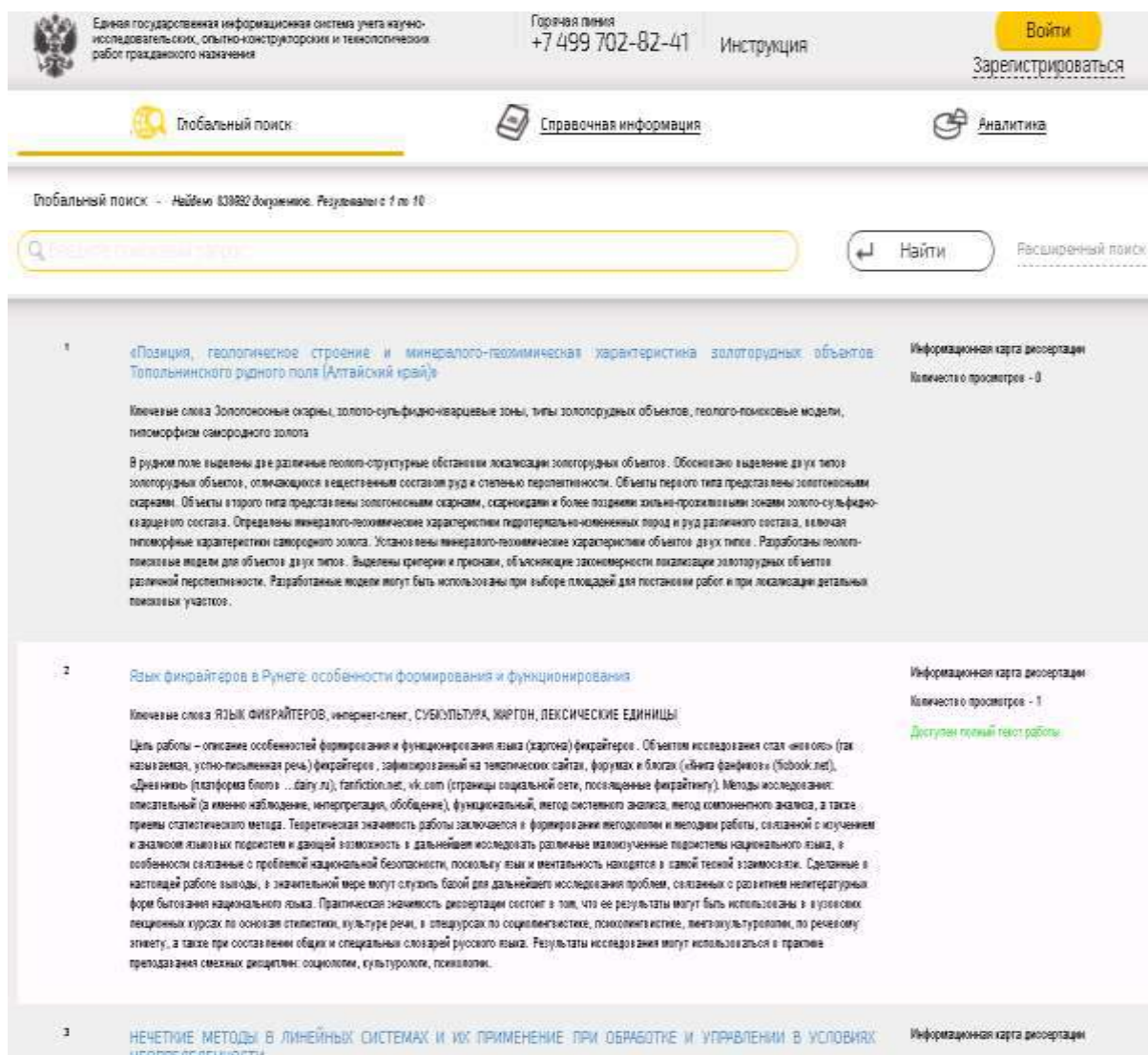
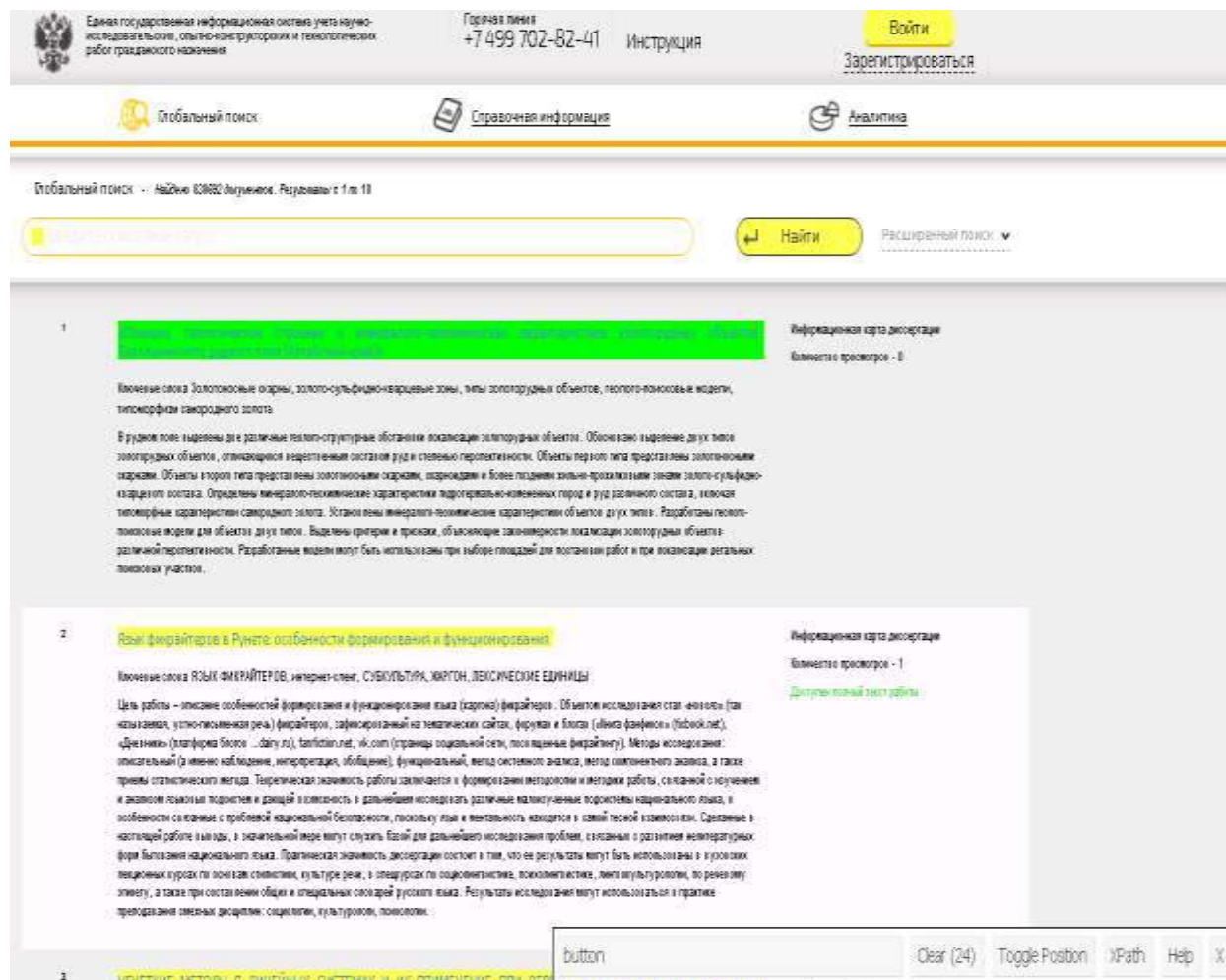


Рис. 2.4.6.1 Страница поиска информации на сайте ЕГИСУ

(<https://www.rosrid.ru/search>)

Для избавления от ненужных элементов в плагине Selectorgadget используется достаточно простая, интуитивно понятная и очень комфортная процедура: мы просто становимся на НЕНУЖНОМ нам элементе (в нашем случае – это кнопка с надписью «Найти» в верхней части экрана) и выделяем его! Результат такого действия виден на следующем рис. 2.4.6.3 .



**Рис. 2.4.6.2 Результат работы плагина Selectorgadget (первый этап)**

Как видно, мы получили в точности то, что нам нужно: на странице жёлтым (или зелёным) цветом выделено в точности 10 нужных нам элементов – заголовков соответствующих страниц. То, что мы не ошибаемся с нашими представлениями и ожиданиями, подтверждает число 10 в правом нижнем углу экрана в записи о количестве выделенных элементов, теперь их осталось (из 24-х первоначальных) только десять.

Нажмём на кнопку с надписью xPath в правом нижнем углу экрана и скопируем нужный нам xPath – патерн. В данном случае это

```
//*[@contains(concat( " ", @class, " " ), concat( " ", "col-lg-8", " " ))]/button
```

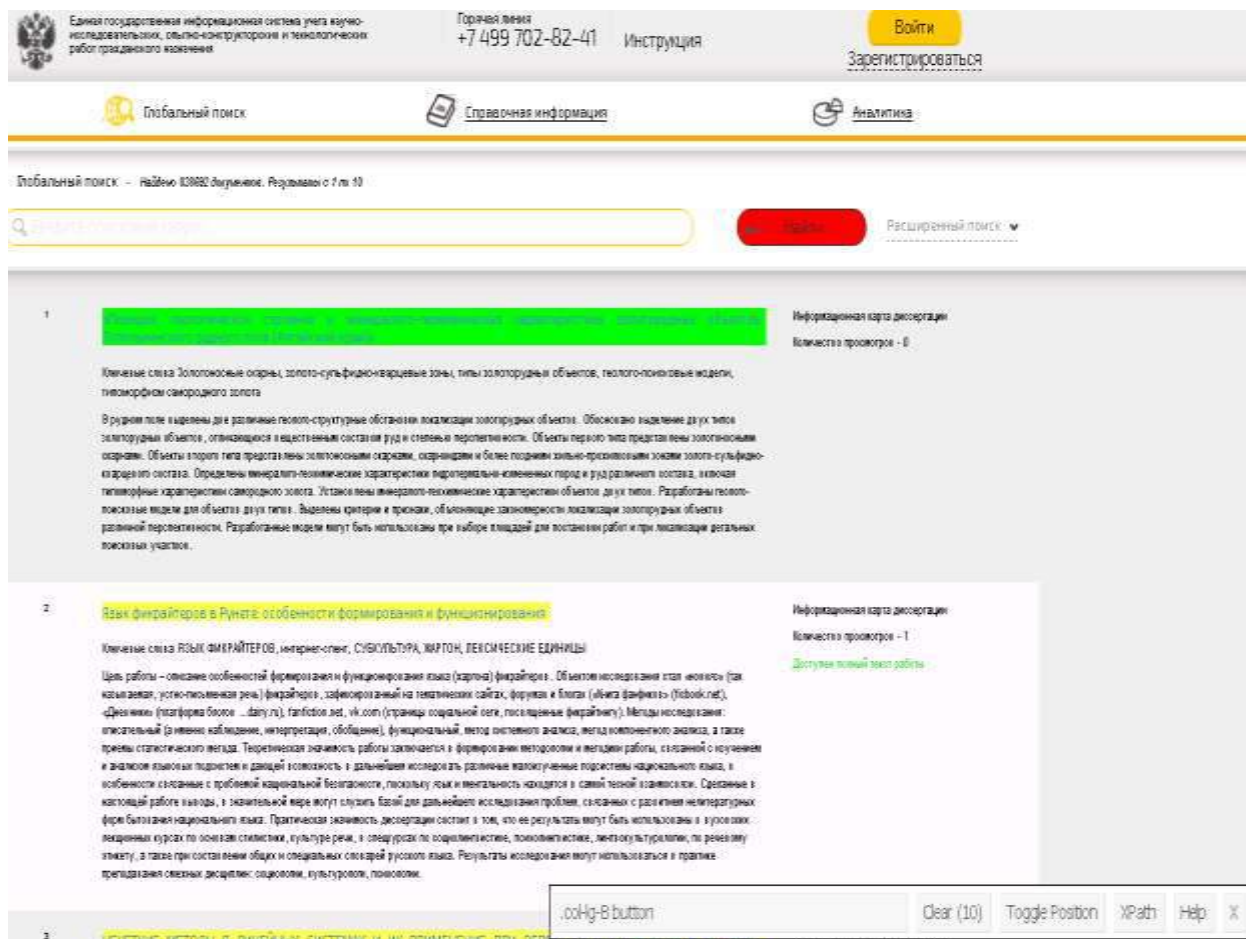


Рис. 2.4.6.3 Результат работы плагина Selectorgadget (второй этап)

Теперь используется плагин **xPath Helper**. Для того, чтобы это сделать, запустим его, нажав на соответствующую кнопку на панели браузера. Откроется два вспомогательных окна в верхней части браузера (рис. 2.4.6.4). Одно окон имеет заголовок «QUERY» и предназначено для ввода XPath – паттерна. Второе окно имеет заголовок «RESULTS» и служит для выдачи всех найденных на странице фрагментов, которые соответствуют введённому в предыдущее окно XPath-паттерну.

В первое окно вводится найденный ранее XPath-паттерн:

```
//*[contains(concat( " ", @class, " " ), concat( " ", "col-lg-8", " " ))]/button
```



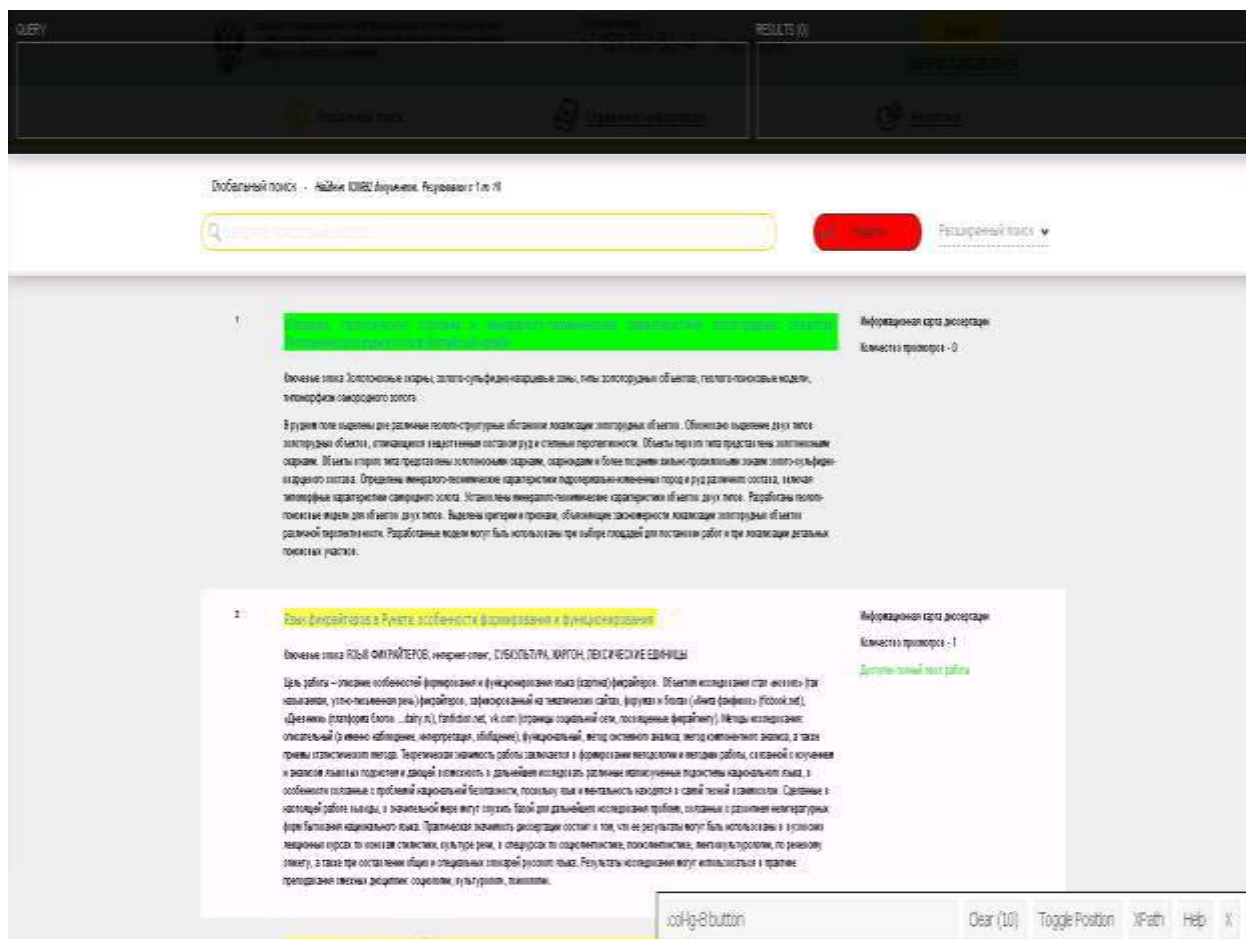


Рис. 2.4.6.4 Рабочие окна плагина xPath Helper

Сразу в окне RESULTS получаются нужные 10 строчек с названиями статей (рис. 2.4.6.5).



Рис. 2.4.6.5 Результат работы плагина xPath Helper

То, что их действительно нужно количество (10 штук), подтверждает изменение заголовка окна вывода, теперь этот заголовок выглядит как «RESULTS (10)», где число в скобках как раз и указывает на количество найденных фрагментов страницы, соответствующих заданному xPath-паттерну. Осталось извлечь эти заголовки из окна результата обыкновенными операциями Copy-Paste.

Результат (десять заголовков разработок) приводится ниже (показано наклонным текстом):

*«Позиция, геологическое строение и минералого-геохимическая характеристика золоторудных объектов Топольнинского рудного поля (Алтайский край)»*

*Язык фикрайтеров в Рунете: особенности формирования и функционирования  
НЕЧЕТКИЕ МЕТОДЫ В ЛИНЕЙНЫХ СИСТЕМАХ И ИХ ПРИМЕНЕНИЕ ПРИ  
ОБРАБОТКЕ И УПРАВЛЕНИИ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ*

*Экспериментальные исследования композитов на основе магнитомягкого порошка  
(промежуточный) этап 4*

*Разработка российского информационно-образовательного портала Новый мир.  
Этап 2*

*Текст в культуре этноса как фактор сохранения идентичности народов сибирско-  
дальневосточного региона*

*Актуальные проблемы и направления развития банковской системы Российской  
Федерации*

*Разработка, изготовление и лабораторные испытания макетного образца  
аппарата для лечения холодом.*

*Разработка компонент молекулярно диагностических систем для выявления  
клинически значимых маркеров генетических, онкологических и инфекционных  
заболеваний человека*

*"Проведение поисковых исследований и создание перспективных технологий по  
защите от поражения электрическим током на автомобилях с комбинированной  
энергетической установкой и электромобилях", Шифр "Изолятор-1"*



## 2.5 Занесение данных в систему ЕИПАЗ для формирования баз данных публикаций и разработок

В системе ЕИПАЗ наполнение баз данных может осуществляться в «ручном», в индивидуальном режиме, при котором каждая новая запись заносится через сервис административной панели. Администратор или авторизованный редактор открывает форму занесения данных и вводит значения всех свойств данной записи. В Приложении 3 представлен фрагмент формы занесения записи публикации.

Для первоначального создания баз данных, интегрированных с различными источниками данных в Интернете, используется методика массированного занесения собранных данных из различных источников и сформированных в итоговые таблицы Excel в записи инфоблоков «Публикации» и «Разработки». В программном комплексе «1С-Битрикс» имеется функция импорта данных в инфоблоки данных, представленных в форматах XML или CSV (формата данных «разделённых запятой», от англ. Comma-Separated Values — значения, разделённые запятыми).

Данные, собираемые и Интернета, оформляются в таблицы Excel, затем эти таблицы переформатируются в таблицы CSV-формата, которые переносятся в соответствующую папку портала ЕИПАЗ, расположенную по пути: **«Панель «Администрирование» - Контент -Структура сайта - Файлы и папки- папка /home/v/viapi/eipaz.ru/public\_html»**. Затем из этой папки выполняется импорт (перенос) данных в соответствующий инфоблок. Ниже показана последовательность переноса данных.

В табл. 2.5.1 представлен фрагмент данных публикаций, собранных из Интернет в формате Excel.

Таблица 2.5.1

## Данные в формате Excel

| Поволжский научно-исследовательский институт экономики и организации |                                                                                                                                        |                                                                 |                                                                                                             |      |                                                                           |  |
|----------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|------|---------------------------------------------------------------------------|--|
| № п/п                                                                | Наименование                                                                                                                           | Авторы                                                          | Издание                                                                                                     | Год  | Ссылка                                                                    |  |
| 1                                                                    | Модель функционирования региональных агропромышленных кластеров с элементами частно-государственного партнерства                       | Сердобинцев Д.В.                                                | Вестник Московского университета. Серия 21: Управление (государство и общество). – 2015. – № 1. – С. 77–93. | 2015 | <a href="http://niisar.ru/publications">http://niisar.ru/publications</a> |  |
| 2                                                                    | Формирование и функционирование региональных мясопродуктовых кластеров с использованием механизмов частно-государственного партнерства | Сердобинцев Д.В.                                                | Научное обозрение. – 2015. – № 4. – С. 220–229.                                                             | 2015 | <a href="http://niisar.ru/publications">http://niisar.ru/publications</a> |  |
| 3                                                                    | Направления организационно-экономических проблем развития АПК на современном этапе                                                     | Черняев А.А.                                                    | Научное обозрение: теория и практика. – 2015. – № 4. – С. 14–24.                                            | 2015 | <a href="http://niisar.ru/publications">http://niisar.ru/publications</a> |  |
| 4                                                                    | Теория и практика современных процессов агропромышленной интеграции в Поволжье                                                         | Сердобинцев Д.В., Стукалина О.М., Соловьева Е.Б., Касатова Т.М. | Научное обозрение: теория и практика. – 2015. – № 4. – С. 109–115.                                          | 2015 | <a href="http://niisar.ru/publications">http://niisar.ru/publications</a> |  |

Данная таблица дополняется свойствами, необходимыми для переноса в инфоблок «Публикации»: каждая запись должна иметь признаки организации-автора, организации-места, вида публикации. Эти свойства, согласно логической структуре (рис. 2.2.1), связаны с записями родительских таблиц (и, соответственно, родительских инфоблоков в реализации БД), поэтому в соответствующие дополнительные ячейки таблицы должны быть записаны идентификационные номера значений из родительских инфоблоков. На рис. 2.5.1 показан фрагмент значений записей инфоблока «Виды публикаций» с ID – номерами.

| + Добавить вид публикации |   |                   |            |         |                     |        |
|---------------------------|---|-------------------|------------|---------|---------------------|--------|
| Добавить раздел           |   |                   |            |         |                     |        |
| <input type="checkbox"/>  |   | Название          | Активность | Сорт. ▲ | Дата изм.           | ID     |
| <input type="checkbox"/>  | ≡ | Статья            | Да         | 10      | 26.09.2017 13:17:24 | 31817  |
| <input type="checkbox"/>  | ≡ | Книга, монография | Да         | 20      | 26.09.2017 13:18:11 | 31809  |
| <input type="checkbox"/>  | ≡ | Доклад            | Да         | 30      | 26.09.2017 13:18:43 | 137847 |
| <input type="checkbox"/>  | ≡ | Отчет о НИР       | Да         | 40      | 26.09.2017 13:19:08 | 31811  |

Рис. 2.5.1 Фрагмент значений инфоблока «Виды публикаций»

Аналогично имеются ID и у записей в инфоблоке организаций. В табл. 2.5.2 представлена расширенная таблица с дополнительными колонками: «Вид публикации»,

«Организация-автор», «Организация-место», в которых записаны ID-номера свойств в родительских инфоблоках. В табл. 2.5.2 в колонке «Вид публикации» стоит значение ID=31817 вида публикации «Статья», в колонках «Организация-автор» и «Организация-место» записан ID = 137241, это признак Поволжского научно-исследовательского института экономики и организации агропромышленного комплекса. Значения автора и места в данном случае совпадают, так как публикации сотрудников института взяты с сайта этого же института. Однако значения в колонках «Организация-автор» и «Организация-место» могут не совпадать, если работы, имеющие авторство одного учреждения были размещены на сайте другого ведомства или института, например, на сайте ЦНСХБ.

Далее, сформированная таблица Excel переводится путем сохранения в табличный CSV-формат. Данные, сформированные таким образом, используются для автоматизированного импорта в соответствующие инфоблоки функцией «1С-Битрикс». Для записи данных CSV-формата в директории сайта ЕИПАЗ создана папка **/home / v / viapi / eipaz.ru / public\_html / upload / import\_files**. В этой папке созданы две подпапки загрузок: публикаций – подпапка **import\_files / public** и разработок - подпапка **import\_files / razrabotki**.

На рис. 2.5.2 показана подпапка **import\_files / public** с частью загруженных файлов CSV-формата публикаций научных учреждений для импорта данных. На рис. 2.5.3 показана подпапка **import\_files / razrabotki** с частью загруженных файлов CSV-формата разработок научных учреждений.

Таблица 2.5.2

## Пример «привязки» записей к значениям родительских инфоблоков

**Поволжский научно-исследовательский институт экономики и организации агропромышленного комплекса**

| № п/п | Наименование                                                                                                                           | Авторы           | Издание                                                                                                     | Год  | Ссылка                                                                    | Вид публикации | Организация-автор | Организация-место |
|-------|----------------------------------------------------------------------------------------------------------------------------------------|------------------|-------------------------------------------------------------------------------------------------------------|------|---------------------------------------------------------------------------|----------------|-------------------|-------------------|
| 1     | Модель функционирования региональных агропромышленных кластеров с элементами частно-государственного партнерства                       | Сердобинцев Д.В. | Вестник Московского университета. Серия 21: Управление (государство и общество). – 2015. – № 1. – С. 77–93. | 2015 | <a href="http://niisar.ru/publications">http://niisar.ru/publications</a> | 31817          | 137241            | 137241            |
| 2     | Формирование и функционирование региональных мясопродуктовых кластеров с использованием механизмов частно-государственного партнерства | Сердобинцев Д.В. | Научное обозрение. – 2015. – № 4. – С. 220–229.                                                             | 2015 | <a href="http://niisar.ru/publications">http://niisar.ru/publications</a> | 31817          | 137241            | 137241            |

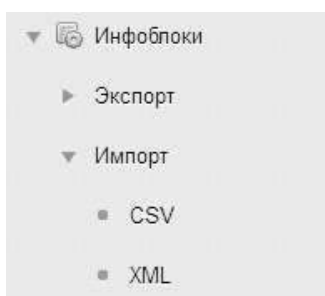
| Путь: /upload/import_files/public |                                                                  | OK           |                     | Поиск    |  |
|-----------------------------------|------------------------------------------------------------------|--------------|---------------------|----------|--|
|                                   | Имя                                                              | Размер файла | Изменен             | Тип      |  |
|                                   | ..                                                               |              |                     |          |  |
|                                   | Публикации из РИНЦ-ВНИИПТУСХ-1 часть_csv.csv                     | 385.87 КБ    | 01.06.2017 16:13:49 | Файл CSV |  |
|                                   | Публикации-РИНЦ-НИИИ-экон-организ-АПК-Центр-Черноз-861p_cvs.csv  | 310.39 КБ    | 18.05.2017 14:38:35 | Файл CSV |  |
|                                   | Публикации-РИНЦ-Ростов Институт эконом и нормативов-1290-cvs.csv | 482.01 КБ    | 25.05.2017 17:54:13 | Файл CSV |  |
|                                   | Публикации-Поволжский-НИИ-экон-orga-APK-702p-cvs.csv             | 229.32 КБ    | 16.05.2017 14:20:20 | Файл CSV |  |
|                                   | Публикации-РИНЦ-Daln-НИИ-экон-APK-csv.csv                        | 8.43 КБ      | 15.05.2017 12:59:30 | Файл CSV |  |
|                                   | Публикации-РИНЦ-ВНИИЭСХ-4099p-Часть 1_csv.csv                    | 625.41 КБ    | 06.06.2017 18:27:36 | Файл CSV |  |

**Рис. 2.5.2** Файлы публикаций для импорта данных

| Путь: /upload/import_files/razrabotki |                                                     | OK           |                     | Поиск    |  |
|---------------------------------------|-----------------------------------------------------|--------------|---------------------|----------|--|
|                                       | Имя                                                 | Размер файла | Изменен             | Тип      |  |
|                                       | ..                                                  |              |                     |          |  |
|                                       | 20170507-Patent-MichGAU-2007-2016-171.csv           | 31.06 КБ     | 10.05.2017 15:27:50 | Файл CSV |  |
|                                       | Разработки-СибНИИЭСХ-с-сайта-81-csv_2.csv           | 25.64 КБ     | 05.09.2017 15:34:09 | Файл CSV |  |
|                                       | Разработки_из-VNIIESH-csv.csv                       | 564.33 КБ    | 01.05.2017 14:39:18 | Файл CSV |  |
|                                       | Разработки_из_РИНЦ_Мич_аграрный_университет_CSV.csv | 26.68 КБ     | 22.06.2017 17:20:34 | Файл CSV |  |
|                                       | Разработки из-Rosrid-143-CSV.csv                    | 231.13 КБ    | 27.07.2017 15:32:11 | Файл CSV |  |

**Рис. 2.5.3** Файлы разработок для импорта данных

В левом меню административной панели системы «1С-Битрикс» имеется пункт «Инфоблоки».



Здесь выбирается подпункт «Импорт» и формат файла, откуда будут импортироваться данные в инфоблок, в данном случае – CSV. При этом открывается форма выбора файлов для импорта данных, показанная на рис. 2.5.4.

**Рис. 2.5.4 Форма выбора файла для импорта**

На этой форме выбирается тип информационного блока и сам инфоблок, куда будут импортированы данные, в нашем случае это типы инфоблоков «Публикации» и «Разработки» и соответствующие инфоблоки. Выбирается также и файл, откуда будут импортированы данные. При нажатии на кнопку «Открыть» открывается возможность выбора папок, где имеется список файлов для импорта, в данном случае: **/home / v / viapi / eipaz.ru / public\_html / upload / import\_files / public** (или **import\_files / razrabotki**). Списки файлов представлены на рис. 2.5.2 и 2.5.3.

Выбирается нужный файл, который указывается в поле «Файл данных» на форме рис. 2.5.4. В данном примере выбирается CSV-файл, сформированный на основе таблицы 2.5.2 публикаций Поволжского НИИ экономики и организации агропромышленного производства - **/upload/import\_files/public/Публикации Поволжского НИИ экон и орган-с сайта-090-CSV.csv**. При нажатии кнопки «Далее» переходим к форме установления соответствия между колонками таблицы 2.5.2 (полями файла данных) и полями инфоблока «Публикации», представленной на рис. 2.5.5. Для каждого поля файла данных кнопками открытия списка полей инфоблока раскрывается список его полей и из списка выбирается соответствующее поле, куда при импортировании будут записываться значения полей их файла данных. После этого нажимается кнопка «Загрузить данные», и программа выполняет перенос данных в инфоблок. На этом процесс импорта данных заканчивается. На рис. 2.5.5 представлены те поля данных, которые были представлены в базе данных источников. В поле «Адрес сайта (страницы)» указывается ссылка на тот сайт, на котором размещен объект информации. Таким образом, в БД ЕИПАЗ создается каталог публикаций и разработок организаций, а более подробные сведения о работах (полные тексты, аннотации) физически в БД ЕИПАЗ не

переносятся, но их можно посмотреть непосредственно у источников этих данных, пройдя по ссылке на сайт или страницу, если они указаны.

**Загрузка информационного блока: шаг 3** ☆

Файл данных | Формат | Поля | Результат

Установка соответствия полей в файле данных и полей в базе

Задайте соответствие полей в файле полям в базе

|                          |                                   |  |
|--------------------------|-----------------------------------|--|
| Поле 1 (№ п/п):          | -                                 |  |
| Поле 2 (Наименование):   | Название (B_IBLOCK_ELEMENT_NAME)  |  |
| Поле 3 (Авторы):         | -                                 |  |
| Поле 4 (Издано):         | Свойство "Издание"                |  |
| Поле 5 (Год):            | Свойство "Год"                    |  |
| Поле 6 (Ссылка):         | Свойство "Адрес сайта (страницы)" |  |
| Поле 7 (Вид публикации): | Свойство "Вид публикации"         |  |
| Поле 8 (Организация):    | Свойство "Организация-автор"      |  |
| Поле 9 (Страна):         | -                                 |  |
| Поле 10 (Орган-место):   | Свойство "Организация-место"      |  |

**Рис. 2.5.5 Установка соответствия полей**

Следует отметить, что при занесении данных для одной организации из разных источников элементы информации могут частично совпадать. При описанной процедуре занесения проверяются совпадения работ по названию, и работы с одинаковым написанием названий заносятся в БД только один раз. Так, например, названия диссертаций и авторефератов совпадают, поэтому, чтобы и те, и другие были перенесены в инфоблок, приходится «вручную» в названиях тех или других вставлять произвольно пробел в каком-нибудь месте названия. В этом случае программа считает их за разные работы и заносит оба вида записей.

В результате, были созданы в рамках ЕИПАЗ две базы данных: БД данных публикаций (объемом 21994 записи) и разработок (объемом 1024 записи).

## 2.6 Работа с базами данных публикаций и разработок в ЕИПАЗ

Доступ к созданным базам данных публикаций и разработок осуществляется через портал ЕИПАЗ, где в главном меню имеются отдельные пункты «Публикации» и «Разработки». Порядок работы с этими базами данных показан на примере публикаций. На рис. 2.6.1 представлен вид страницы «Публикации», которая представляется пользователю при нажатии на соответствующий пункт меню. Внизу страницы вынесен единый список всех публикаций БД. Этот список размещен на нескольких страницах по двадцать записей на каждой странице. Имеется механизм отбора из всего списка нужных записей по заданным (свойствам) признакам, показанным на рис. 2.6.1: названию, авторам, организациям, ГРНТИ, году, виду публикации и другим.

**Публикации**

Фильтр

Тип организации:

Фильтр

Название:

Авторы:

Ведомство:

Организация-автор:

Вид публикации:

ГРНТИ:

Год:

Авторы-коллектив:

Ключевые слова:

Логин:

Пароль:

☐ Запомнить меня

[Регистрация](#)

[Забыли свои данные?](#)

Войти как

[@](#)

**ОРГАНИЗАЦИОННО-ЭКОНОМИЧЕСКИЕ ОСНОВЫ ЭФФЕКТИВНОГО ИСПОЛЬЗОВАНИЯ ПРОИЗВОДСТВЕННОЙ МОЩНОСТИ СКОТООТКОРМОЧНЫХ ПРЕДПРИЯТИЙ**

Организация-автор: [ФГБНУ Всероссийский научно-исследовательский институт организации производства, труда и управления в сельском хозяйстве](#)

Вид публикации: [Автореферат](#)

Год: 1982

Рис. 2.6.1 Страница с началом списка публикаций

При выборе из раскрывающегося списка показывается список из соответствующего родительского инфоблока, из которого выбирается нужное значение для фильтрации. Окно «Тип организации» можно использовать для предварительного выбора типа организаций. Тогда при выборе «Организации-автора» в раскрывающемся списке организаций будут представлены только организации отобранного ранее типа. Это сделано для упрощения выбора организации из всего списка, который может быть довольно большим.



Окно «Авторы» заполняется также выбором из раскрывающегося списка авторизованных в системе авторов (занесенных в инфоблок «Персоналии»). Однако, при массовом занесении записей в базу данных не выполнялось разделение авторов на авторизованных и неавторизованных, так как, во-первых, это очень трудоемкий процесс, и, во-вторых, не все авторы работ авторизованы. Поэтому для работ, у которых авторы даются списком, при импорте этот список авторов переносится в поле «Авторы-коллектив». Для отбора работ по этому признаку следует с клавиатуры занести фамилию нужного автора.

Также выполняется поиск и для окон «Название», «Год», «Ключевые слова». Следует отметить, что во многих работах, размещенных у источников, отсутствуют заполненные некоторые признаки, например, не выбран код ГРНТИ, не указан вид публикации, не указаны ключевые слова, иногда даже и год и т.д. В результате поиск в базе данных будет не совсем корректный. Поэтому следует проводить поиск многократно по различным признакам. Так, поиск по названию работы может дать хорошие результаты, если выполнять его, вводя различные слова, или сочетания слов, которые могут быть в названии предполагаемой работы.

## **ЗАКЛЮЧЕНИЕ**

Рассмотрена общая проблема формирования единого информационного Интернет-пространства аграрных знаний для АПК. Описаны общие подходы к проблеме интеграции информационных ресурсов в единой базе данных из разрозненных гетерогенных баз данных, размещенных на сайтах различных агентств и научно-исследовательских организаций. Разработана база данных для представления информационных ресурсов и описан разработанный аппарат для размещения их в Интернете в рамках портала [www.eipaz.ru](http://www.eipaz.ru).

На основании представленных методов и программных средств поиска и сбора информации в Интернете и методов построения таблиц для переноса в базу данных средствами комплекса «1С-Битрикс» в рамках портала [www.eipaz.ru](http://www.eipaz.ru) созданы базы данных информационных научных ресурсов “Публикации научных работ” и “Прикладные разработки и инновационные проекты” аграрной науки научно-исследовательских учреждений аграрно-экономического направления.

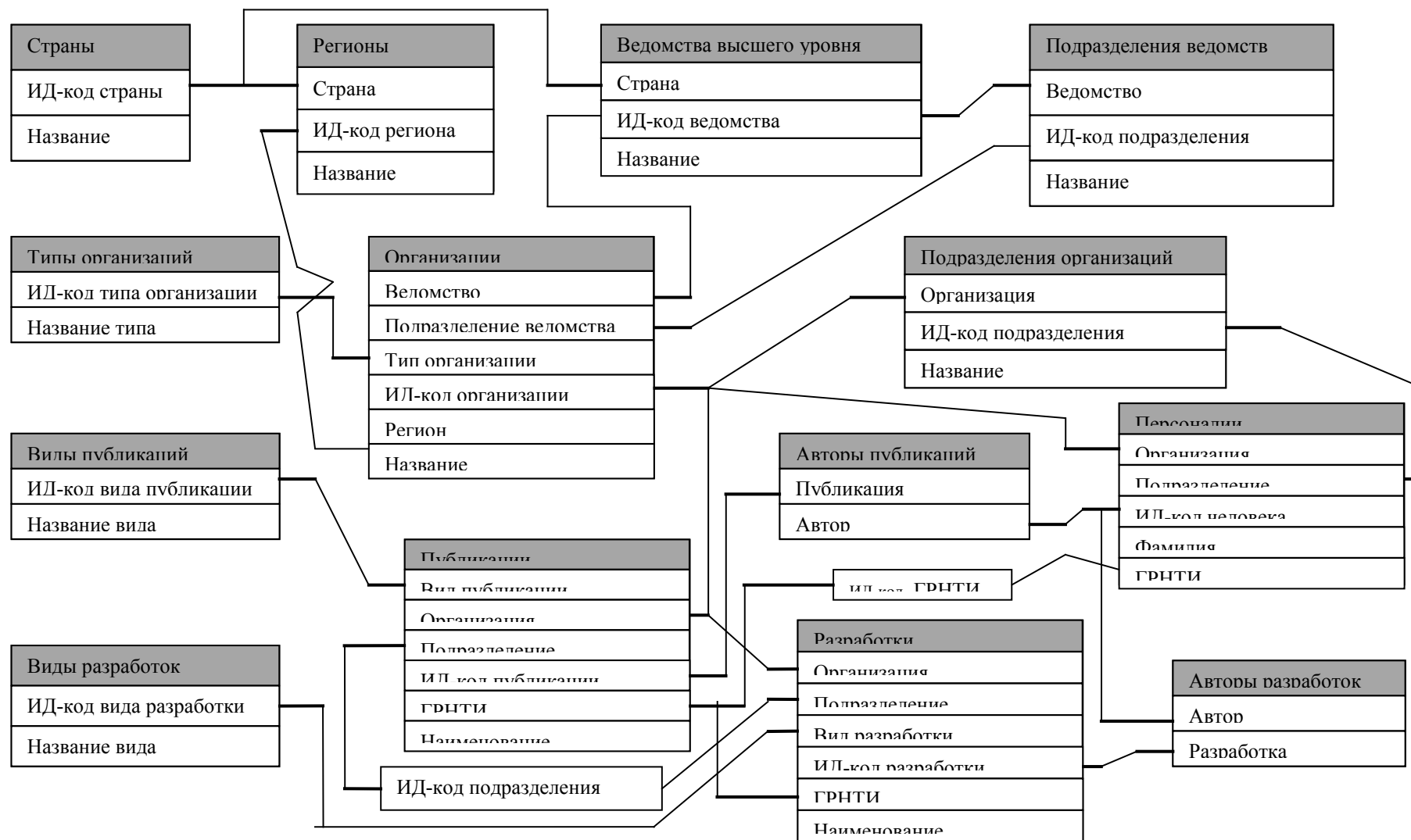
Представленные подходы к формированию информационных ресурсов и доступу к ним на портале средствами комплекса «1С-Битрикс» закладывают основу для дальнейшего развития портала [www.eipaz.ru](http://www.eipaz.ru) и расширения единого информационного пространства аграрных знаний.

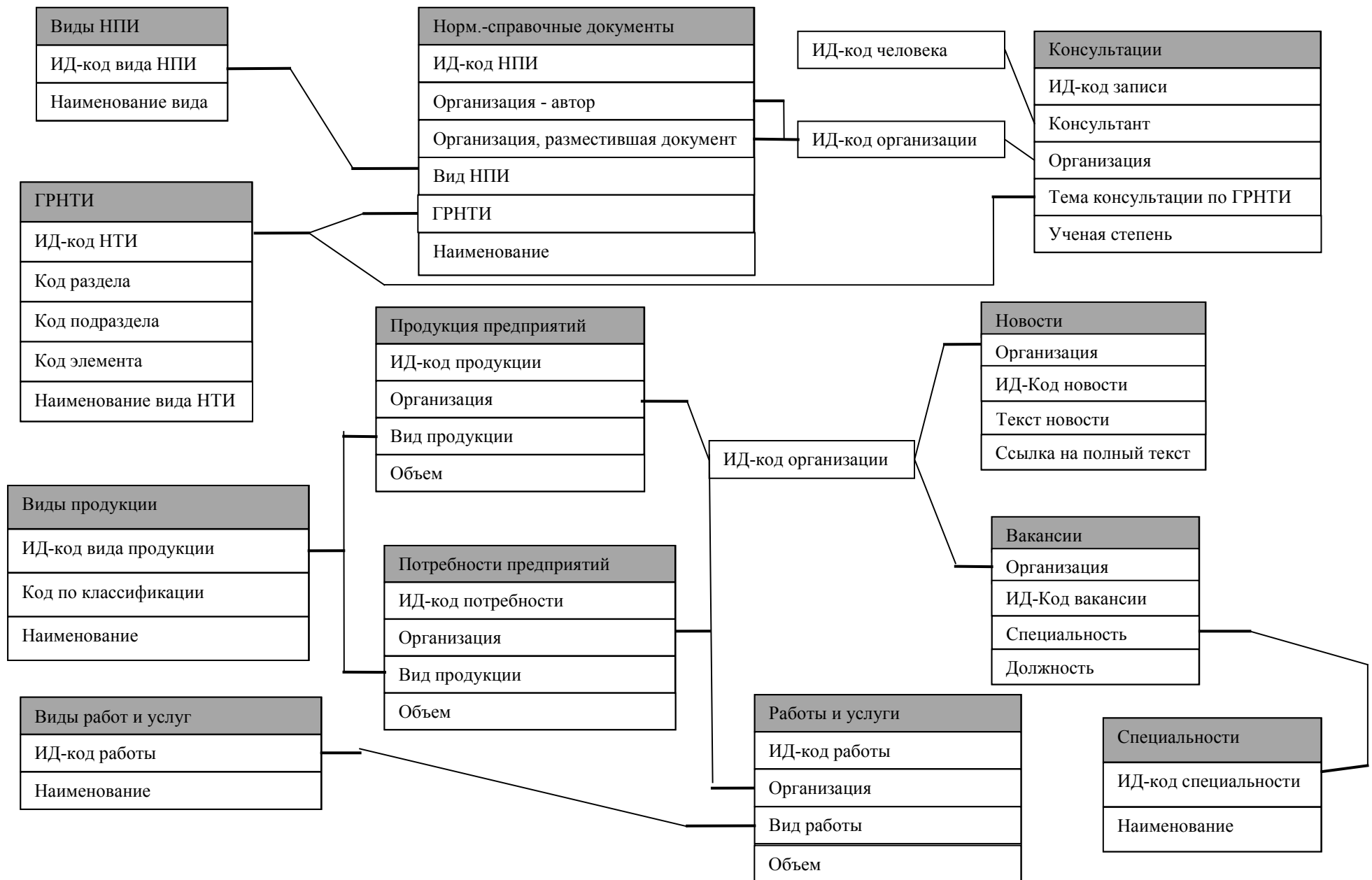
## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

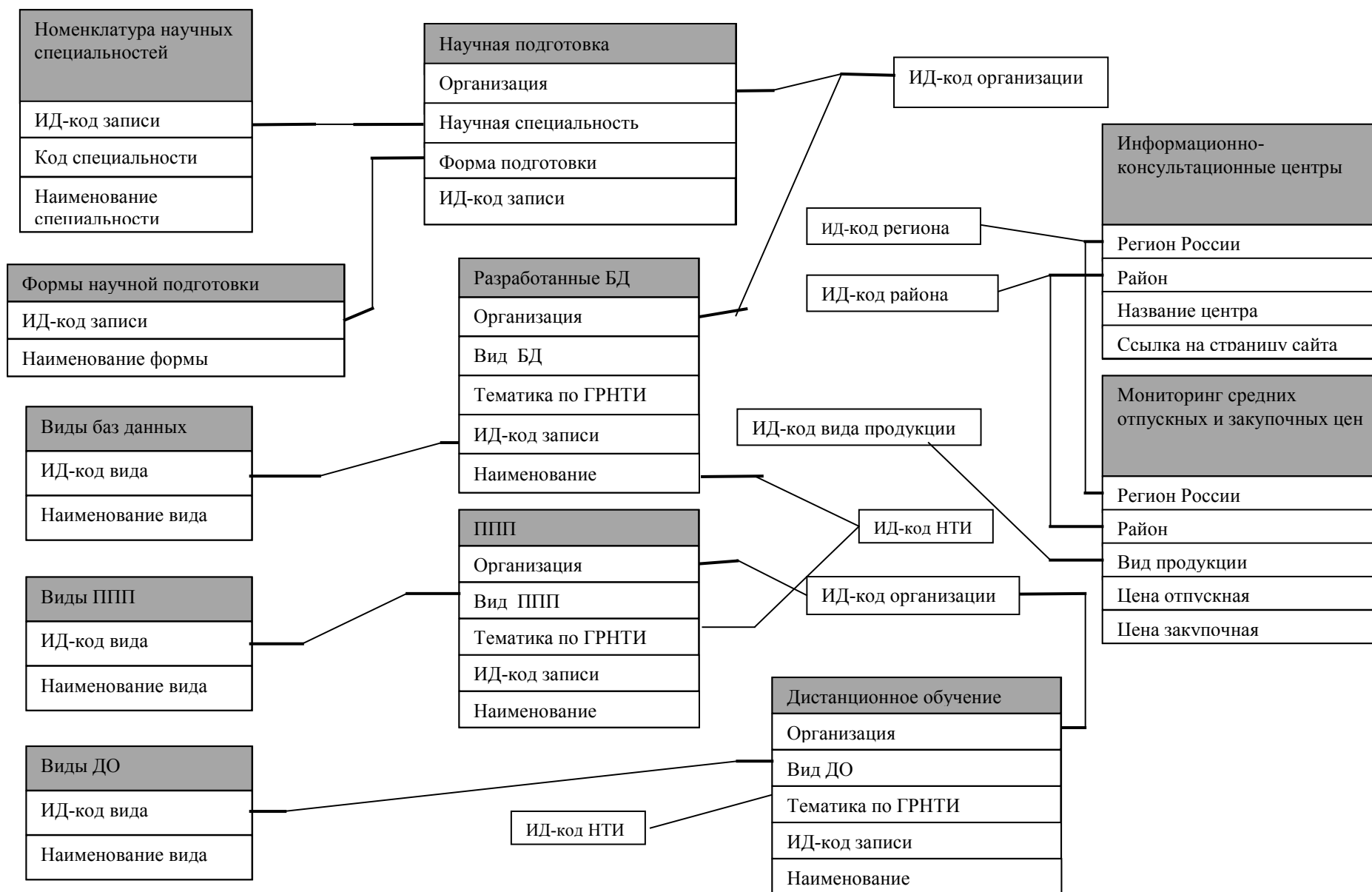
1. Тевс Д.П., Климентьева В.А., Петров В.А., Злобина Т.Н., Стекина Г.В., Пантюшина Т.В., Рыбина Л.И. Информационные технологии для пользователей библиотеки образовательного учреждения. Барнаул: Изд-во БГПУ, 2004. – 133 с.
2. Жлобинская О.Н. Российская национальная библиотека: [http://www.rusmarc.ru/publish/MARC\\_now.pdf](http://www.rusmarc.ru/publish/MARC_now.pdf).
3. Кашников А., Лядова Л. Интеграция гетерогенных источников данных на основе рекурсивной декомпозиции. - International Journal "Information Technologies & Knowledge" Vol.5, Number 3, 2011.
4. Википедия <http://dic.academic.ru/dic.nsf/ruwiki/1506712>
5. Когаловский М.Р. Методы интеграции данных в информационных системах - Сб. трудов Третьей Всероссийской конференции "Стандарты в проектах современных информационных систем". Москва, 23-24 апреля 2003 г.
6. Гарсиа-Молина Г., Ульман Дж., Уидом Дж. Системы баз данных. Полный курс = Database Systems: The Complete Book. — Вильямс, 2003. — 1088 с.
7. Гюнтер Зауфер, Мэй Сельваж, Эойн Лейн, Билл Мэтьюс. Шаблоны для информационного сервиса.
8. William Kent. Solving Domain Mismatch and Schema Mismatch Problems with an Object-Oriented Database Programming Language. Proceedings of the International Conference on Very Large Data Bases (1991).
9. Erhard Rahm, Philip A. Bernstein. A Survey of Approaches to Automatic Schema Matching. VLDB JOURNAL(2001).
10. Ahmed K. Elmagarmid, Panagiotis G. Ipeirotis, Vassilios S. Verykios. Duplicate Record Detection: A Survey. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 19, NO. 1, JANUARY 2007.
11. [Электронный ресурс]. - Режим доступа: <http://golosnoy.com.ua/2015/%D0%BA%D0%B0%D0%BA-egov-%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%B0%D0%B5%D1%82-%D0%B2-%D1%8D%D1%81%D1%82%D0%BE%D0%BD%D0%B8%D0%B8-x-road> (дата обращения: 10.09.2017).
12. Меденников В.И., Муратова Л.Г., Сальников С.Г. и др. Отчет о НИР «Разработать модель интеграции оптимальных информационных систем для сельскохозяйственных предприятий». - ВИАПИ РАСХН. 1999 - 76с.
13. Муэрс Р. Эффективное управление. Пер. с англ. - М. : Финпресс, 1998.

14. АСУ на промышленном предприятии: Методы создания: Справочник/ Михалев С.Б., Седегов Р.С., Гринберг А.С. и др.- 2 изд. перераб. и доп.-М.: Энергоатомиздат, 1986.
15. Кульба В.В., Ковалевский С.С., Косяченко С.А., Сиротюк В.О. Теоретические основы проектирования оптимальных структур распределенных баз данных. - М.: Синтег, 1999.
16. Меденников В.И. Теоретические аспекты синтеза структур компьютерного управления агропромышленным производством. // Аграрная наука, 1993, N 2.
17. Меденников В.И., Муратова Л.Г., Сальников С.Г. и др. «Разработать технико-экономическое обоснование проекта единого информационного Интернет-пространства знаний агронауки», отчет о НИР, ВИАПИ им. Никонова, 2010г.
18. Меденников В.И., Муратова Л.Г. Трансформация ИКС в условиях создания Единого Интернет-Пространства аграрных знаний. Материалы Международной научно-практической конференции "Развитие институтов инновационной экономики в условиях интеграции России в мировое экономическое пространство", Ярославль, 2016.
19. Меденников В.И., Муратова Л.Г., Сальников С.Г. и др. Отчет о НИР «Разработать концепцию единого информационного Интернет-пространства знаний агронауки». - ВИАПИ РАСХН. 2009 - 130с.
20. Меденников В.И., Ерешко Ф.И., Сальников С.Г. Интернет-технологии в экономике знаний. Материалы девятой международной конференции «Управление развитием крупномасштабных систем» MLSD'2016, г. Москва, ИПУ РАН, том 1, стр. 178-181.
21. Меденников В.И., Сальников С.Г. Единое информационное Интернет-пространство научно-образовательных ресурсов. Информатизация образования и науки, № 3(35)/2017, стр. 3-13.
22. [Электронный ресурс] - Режим доступа: [http://rcdl.ru/doc/2014/paper/RCDL2014\\_331-338.pdf](http://rcdl.ru/doc/2014/paper/RCDL2014_331-338.pdf) (дата обращения: 10.09.2017).
23. Меденников В.И., Муратова Л.Г., Сальников С.Г., Личман А.А., Луппов В.В. и др. «Мониторинг состояния и объемов информационных ресурсов информационно-консультационной службы АПК в Интернет-пространстве», Отчет о НИР, ВИАПИ, 2012 г.

## ПРИЛОЖЕНИЕ 1 ЛОГИЧЕСКАЯ СТРУКТУРА БД ЕИПАЗ







## ПРИЛОЖЕНИЕ 2 ИЗВЛЕЧЕНИЕ ДАННЫХ СО СТРАНИЦ САЙТОВ (ТЕКСТ ПРОГРАММЫ НА ЯЗЫКЕ R).

```
getUrlFromPages<-function(pages, cookie, sleep=4)
{
  urls<-
  paste("http://fano.gov.ru/ru/about/sub_organizations/?count_per_page_4=30&DEPT_TYPE_ID_4=3&page_4=",
  pages, sep="")
  ret<-data.frame()
  for(url in urls)
  {
    ret<-rbind(ret, getNIUs(url, cookie))
    Sys.sleep(sleep)
  }
  return(ret)
}
```

```
getNIUs<-function(url, cookie)
{
  cat("Грузим-список: ", url,"\n")
  html<-getURL(url, cookie=cookie)
  doc<-htmlParse(html)
  href<-xpathSApply(doc, "//a[@href]", xmlGetAttr,'href')
  posts<-data.frame(href)
  return(posts)
}
```

```
getInfoFromUrls<-function(URLs, cookie, sleep=4)
{
  ret<-data.frame()
  URLs1<-URLs[,1]
  for(urls in URLs1)
  {
    ret<-rbind(ret, getNIUsInfo(urls, cookie))
    Sys.sleep(sleep)
  }

  return(ret)
}
```

```
getNIUsInfo<-function(url, cookie)
```



```

{
  cat("Грузим-инфо: ", url,"\n")
  html<-getURL(url, cookie=cookie)
  doc<-htmlParse(html)

  nameorg<-xpathSApply(doc, "/html/body/div[1]/div[3]/div[2]/div[2]/div/h2", xmlValue)
  rukorg<-xpathSApply(doc, "//span[@class='structure-item__name']", xmlValue)
  joborg<-xpathSApply(doc, "//span[@class='structure-item__job']", xmlValue)
  postorg<-xpathSApply(doc, "//*[@id='js-tab-common']/text()[5]", xmlValue)
  teleorg<-xpathSApply(doc, "//*[@id='js-tab-common']/text()[7]", xmlValue)
  faxorg<-xpathSApply(doc, "//*[@id='js-tab-common']/text()[8]", xmlValue)
  siteorg<-xpathSApply(doc, "//*[@id='js-tab-common']/a[2]", xmlValue)
  mailorg<-xpathSApply(doc, "//*[@id='js-tab-common']/a[1]", xmlValue)

  if (!is.character(nameorg)) nameorg<-"нет данных"
  if (!is.character(rukorg)) rukorg<-"нет данных"
  if (!is.character(joborg)) joborg<-"нет данных"
  if (!is.character(postorg)) postorg<-"Нет данных"
  if (!is.character(teleorg)) teleorg<-"нет данных"
  if (!is.character(faxorg)) faxorg<-"нет данных"
  if (!is.character(siteorg)) siteorg<-"нет данных"
  if (!is.character(mailorg)) mailorg<-"нет данных"

  posts<-data.frame(url,rukorg,postorg)
  # posts<-data.frame(url,nameorg,mailorg,siteorg,joborg,teleorg,faxorg)

  return(posts)
}

```

```

getKatVNIESH<-function()
{
  # вытащим список подкаталогов каталога НТД сайта ВНИЭСХ
  # загрузим нужные библиотеки
  library(curl)
  library(XML)
  library(RCurl)
  # качаем содержимое страницы каталога НТД
  html<-getURL("http://vniiesh.ru/results/katalog/", cookie=cookie)
  # парсим полученные данные
  doc<-htmlParse(html)
  # извлекаем все ссылки

```

```

href<-xpathSApply(doc, "//a[@href]", xmlGetAttr,'href')
# извлекаем имена ссылок
nams<-xpathSApply(doc, "//*[@id='content']/div/ul/li/a", xmlValue)
# оставим только страницы подкаталогов
urls<-href[grepl('/results/katalog/',href)]
# формируем абсолютные адреса страниц подкаталога
urls<-paste("http://www.vniiesh.ru",urls,sep="")
# соединение имен и ссылок в один датафрейм
df<-data.frame(nams,urls)
# возвращаем датафрейм подкаталогов
return(df)
}

getNTDurl<-function(urls, cookie, sleep=4)
{
# по списку адресов подкаталогов собираем подкаталоги разработок
ret<-data.frame() # создаем пустой датафрейм
for(url1 in urls) # перебираем в цикле все адреса подкаталогов
{
cat("Грузим-инфо: ", url1,"\n") # сообщаем о действии в консоль
html<-getURL(url1, cookie=cookie) # качаем содержимое страницы каталога НТД
doc<-htmlParse(html) # парсим полученные данные
href<-xpathSApply(doc, "//a[@href]", xmlGetAttr,'href') # извлекаем все ссылки
href<-paste("http://www.vniiesh.ru",href,sep="") # добавим адрес спереди
href<-paste(href,"/",sep="") # и черту сзади
# качаем имена разработок
#
# nams<-xpathSApply(doc, "//*[@id='content']/div/ul[1]/li/a", xmlValue)
#
#
# к сожалению, эта штука возвращает неправильный результат - кашу из символов -на некоторых
# специфических строковых данных, потому пока что только выкачиванием
# гиперссылок
#

urls2<-href[grepl(url1,href)] # оставим только нужные ссылки
urls2<-urls2[grepl("html",urls2)] # и только ссылки на НТД
# df<-data.frame(nams,urls2) # формируем датафрейм
df<-data.frame(urls2) # формируем датафрейм
ret<-rbind(ret, df) # пишем в датафрейм
Sys.sleep(sleep) # задержка системы
}

```

```
return(ret)
```

```
return()
```

```
}
```

```
getNTDdata<-function(urls, cookie, sleep=4)
```

```
{
```

```
# по списку адресов НТД вытаскиваем всю подробную информацию
```

```
ret<-data.frame() # создаем пустой датафрейм
```

```
for(url1 in urls) # перебираем в цикле все адреса подкаталогов
```

```
{
```

```
html<-getURL(url1, cookie=cookie) # качаем содержимое страницы каталога НТД
```

```
while (length(grep("\xd0",html))) {html<-sub("\xd0"," ",html)}
```

```
Encoding(html)<-"UTF-8"
```

```
doc<-htmlParse(html) # парсим полученные данные
```

```
# полный состав НТД
```

```
ntdall<-xpathSApply(doc, "//tr[(((count(preceding-sibling::*) + 1) = 2) and parent::*)//td", xmlValue)
```

```
ntdall<-as.character(ntdall)
```

```
if (length(grep("\u009d",ntdall)))
```

```
{
```

```
cat("Плохая структура: ", url1,"\n")
```

```
ntd_vidr<-"bad page"
```

```
ntd_name<-"bad page"
```

```
ntd_krat<-"bad page"
```

```
ntd_razr<-"bad page"
```

```
ntd_avto<-"bad page"
```

```
ntd_kont<-"bad page"
```

```
ntd_adre<-"bad page"
```

```
ntd_secr<-"bad page"
```

```
ntd_tele<-"bad page"
```

```
ntd_spra<-"bad page"
```

```
ntd_regn<-"bad page"
```

```
}
```

```
else
```

```
{
```

```
cat("Хорошая структура: ", url1,"\n")
```

```
while (length(grep("\r",ntdall))) {ntdall<-sub("\r"," ",ntdall)}
```

```
while (length(grep("\n",ntdall))) {ntdall<-sub("\n"," ",ntdall)}
```

```
while (length(grep(" ",ntdall))) {ntdall<-sub(" ","",ntdall)}
```

```

ntdall<-sub("Вид результата:", "#Вид результата:", ntdall)
ntdall<-sub("Наименование результата:", "#Наименование результата:", ntdall)
ntdall<-sub("Краткое описание результата:", "#Краткое описание результата:", ntdall)
ntdall<-sub("Разработчик:", "#Разработчик:", ntdall)
ntdall<-sub("азработчик:", "#Разработчик:", ntdall)
ntdall<-sub("Авторы:", "#Авторы:", ntdall)
ntdall<-sub("Адрес:", "#Адрес:", ntdall)
ntdall<-sub("Контактна", "#Контактна", ntdall)
ntdall<-sub("Ученый секретарь:", "#Ученый секретарь:", ntdall)
ntdall<-sub("Телефон/Е-Mail:", "#Телефон/Е-Mail:", ntdall)
ntdall<-sub("Справочно:", "#Справочно:", ntdall)
ntdall<-sub("Регистрационный номер результата в базе данных", "#Регистрационный номер результата в базе
данных", ntdall)
ntdall<-sub("егистрационный номер результата в базе данных", "#Регистрационный номер результата в базе
данных", ntdall)
ntdall<-sub(" оссельхозакадемии", " Россельхозакадемии", ntdall)
ntdall<-sub(" АСХН", " РАСХН", ntdall)
while (length(grep(" ", ntdall))) {ntdall<-sub(" ", " ", ntdall)}
list_ntd<-strsplit(as.character(ntdall), "#")
names(list_ntd)<- "namentd"
ntd_vidr<-list_ntd$namentd[2]
ntd_name<-list_ntd$namentd[3]
ntd_krat<-list_ntd$namentd[4]
ntd_razr<-list_ntd$namentd[5]
ntd_avto<-list_ntd$namentd[6]
ntd_kont<-list_ntd$namentd[7]
ntd_adre<-list_ntd$namentd[8]
ntd_secr<-list_ntd$namentd[9]
ntd_tele<-list_ntd$namentd[10]
ntd_spra<-list_ntd$namentd[11]
ntd_regn<-list_ntd$namentd[12]
ntd_vidr<-sub("Вид результата: ", "", ntd_vidr)
ntd_name<-sub("Наименование результата: ", "", ntd_name)
ntd_krat<-sub("Краткое описание результата: ", "", ntd_krat)
ntd_razr<-sub("Разработчик: ", "", ntd_razr)
ntd_avto<-sub("Авторы: ", "", ntd_avto)
ntd_kont<-sub("Контактна", "", ntd_kont)
ntd_kont<-sub("информаци", "", ntd_kont)
ntd_adre<-sub("Адрес: ", "", ntd_adre)
ntd_secr<-sub("Ученый секретарь: ", "", ntd_secr)
ntd_tele<-sub("Телефон/Е-Mail: ", "", ntd_tele)
ntd_spra<-sub("Справочно: ", "", ntd_spra)

```

```

ntd_regn<-sub("Регистрационный номер результата в базе данных Россельхозакадемии: ","",ntd_regn)
ntd_vidr<-sub("Вид результата:", "",ntd_vidr)
ntd_name<-sub("Наименование результата:", "",ntd_name)
ntd_krat<-sub("Краткое описание результата:", "",ntd_krat)
ntd_razr<-sub("Разработчик:", "",ntd_razr)
ntd_avto<-sub("Авторы:", "",ntd_avto)
ntd_adre<-sub("Адрес:", "",ntd_adre)
ntd_secr<-sub("Ученый секретарь:", "",ntd_secr)
ntd_tele<-sub("Телефон/Е-Mail:", "",ntd_tele)
ntd_spra<-sub("Справочно:", "",ntd_spra)

while (length(grep("Тема НИ :",ntd_spra))) {ntd_spra<-sub("Тема НИ :", "Тема НИР:",ntd_spra)}
while (length(grep(" Тема НИР:",ntd_spra))) {ntd_spra<-sub(" Тема НИР:", "Тема НИР:",ntd_spra)}

ntd_regn<-sub("Регистрационный номер результата в базе данных Россельхозакадемии:", "",ntd_regn)
ntd_regn<-sub("Регистрационный номер результата в базе данных РАСХН:", "",ntd_regn)
ntd_regn<-sub("Регистрационный номер результата в базе данных:", "",ntd_regn)
ntd_regn<-sub("Регистрационный номер результата в базе данных", "",ntd_regn)

ntd_regn<-sub("егистрационный номер результата в базе данных Россельхозакадемии:", "",ntd_regn)
ntd_regn<-sub("егистрационный номер результата в базе данных Р
АСХН:", "",ntd_regn)
ntd_regn<-sub("егистрационный номер результата в базе данных:", "",ntd_regn)
ntd_regn<-sub("егистрационный номер результата в базе данных", "",ntd_regn)
ntd_regn<-sub(" оссельхозакадемии: ", "",ntd_regn)
ntd_spra<-sub(" азработать", " Разработать",ntd_spra)

if (nchar(as.character(ntd_kont))<5) {ntd_kont<-"} else {ntd_kont<-substr(ntd_kont,5,9999)}
}
#df<-data.frame(url1,ntdall,ntdorg,ntdraz)      # формируем датафрейм
df<-
data.frame(url1,ntd_vidr,ntd_name,ntd_krat,ntd_razr,ntd_avto,ntd_kont,ntd_adre,ntd_secr,ntd_tele,ntd_spra,ntd_regn)
# формируем датафрейм
ret<-rbind(ret, df)      # пишем в датафрейм
Sys.sleep(sleep)      # задержка системы
}
return(ret)
}

```

## ПРИЛОЖЕНИЕ 3 ФРАГМЕНТ ФОРМЫ ЗАНЕСЕНИЯ ЗАПИСИ ПУБЛИКАЦИИ

Публикации: Элемент: Добавление ☆

Элементы

Элемент | Анонс | Подробно | SEO | Разделы

### Элемент

Активность: ☒

Начало активности:

Окончание активности:

Название:

Символьный код:

Сортировка:

Значения свойств:

Авторы:

Добавить Персону

Ведомство:

Организация-автор:

Добавить организацию

Вид публикации:

Подразделение:

Подразделение\_1:

Подразделение\_2:

Департамент ведомства:

ГРНТИ:   
... 68.01.33-Терминология. Справочная литература. Учебная литература  
... 68.01.37-Стандартизация  
... 68.01.39-Пропаганда и популяризация знаний по сельскому хозяйству  
... 68.01.51-Комплексные вопросы в сельском хозяйстве  
... 68.01.63-Производство товаров народного потребления

Издательство:

Страна:

Меденников Виктор Иванович  
Луппов Вячеслав Владимирович  
Муратова Людмила Георгиевна  
Сальников Сергей Георгиевич  
Личман Аэлита Афанасьевна  
Тухина Наталья Юрьевна

*Научное издание*

Методы и модели интеграции аграрных научных информационных  
ресурсов в Интернете

Монография

ISBN 978-5-6040748-9-3

Отпечатано в типографии ООО «Аналитик»  
Заказ № 870. Бумага офсетная. Тираж 300 экз.  
Формат 60х90/16. Усл. печ. л. 6,38  
г. Москва, ул. Клары Цеткин, д.18, стр. 3 . Тел. 495 617-09-24

