

Интеграция научных информационных ресурсов ОИЯИ на платформе JDS

© Т.Н. Заикина © С.В. Куняев © Р.Н. Семенов
© П.В. Устенко © И.А. Филозова © Г.В. Шестакова
Объединенный институт ядерных исследований,
Дубна
ztanya@jinr.ru svk@jinr.ru roman@jinr.ru
omega@jinr.ru fia@jinr.ru shestakova@jinr.ru

Аннотация

Опыт показывает, что информационные потребности конечных пользователей не покрываются одной информационной системой, и в одной организации, как правило, поддерживается несколько систем. В рамках деятельности по информационному сопровождению научных исследований в ОИЯИ поддерживаются и развиваются сервер научных документов JINR Document Server (JDS), информационно-аналитическая система «Персональная ИНформация о сотрудниках ОИЯИ» / Personal INformation (PIN), система управления научными мероприятиями Indico. В работе рассматриваются подходы для обеспечения интеграции контента этих систем, их интероперабельности.

1 Введение

Публикационная активность — один из основных показателей эффективности научно-исследовательской работы ученого. Но это лишь один из элементов, которые должны быть выявлены и оценены [3, 4]. При оценке эффективности вспомогательную роль могут играть формальные библиометрические параметры: характер публикаций (монографии, реферируемые журналы, труды конференций, прочее); полная цитируемость и цитируемость за указанный период; число и «география» соавторов публикаций. Также представляют интерес такие показатели как участие сотрудников в научных мероприятиях, грантах, проектах и пр.

Для представления своей интеллектуальной продукции научные и образовательные организации создают и поддерживают институциональные репозитории в форме открытого доступа (ОД), обеспечивающие сбор и хранение информационных ресурсов, а также набор сервисов для доступа и управления этими информационными ресурсами.

Труды 16-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL-2014, Дубна, Россия, 13–16 октября 2014 г.

Открытые архивы строятся по принципу распределенной системы с общим сетевым протоколом и программным обеспечением, что позволяет говорить о единой информационной системе [8, 11, 12, 17]. Таким образом, формируется единое научное и образовательное информационное пространство [5, 7, 9]. Это позволяет привлечь большее число научных сообществ, занимающихся решением общих проблем.

Значительный рост использования научных информационных систем иллюстрирует потребность в них научного сообщества. Эти системы получили название CRIS (Current Research Information System) и представляют собой информационные системы для доступа к научной и академической информации, включая персональные профили исследователей, их научно-организационную работу, педагогическую и инновационную деятельность; используемые в исследованиях экспериментальные установки и оборудование и пр. [6, 16]. На институциональном уровне CRIS система является также инструментом аналитической, управленческой и организаторской деятельности в науке [3, 4, 10, 16].

2 Обзор программно-технических средств и информационных ресурсов ОИЯИ

В ОИЯИ, как и во многих других организациях, созданы и успешно используются системы обработки информации, связанные и пересекающиеся по наполнению с CRIS — экспериментальные данные и результаты исследований, библиографическая информация, системы управления проектами, информация о персонале и т.д. Однако эти системы были созданы с определенной целью — для сопровождения библиотеки, управления проектами, учета финансовой деятельности, управления персоналом. Наиболее ценным контентом, объемлющим научную и научно-организационную деятельность, обладают сервер научных документов JINR Document Server (JDS), информационно-аналитическая система «ПИН».

2.1 ПИН (Персональная ИНформация)

ПИН — информационная система, обеспечивающая: сбор первичной информации о результатах научной деятельности, включая публикации, научно-организационную работу, педагогическую и инновационную деятельность; авторизованный доступ для ввода и просмотра

информации; вычисление индекса ПРНД, подготовка пакета документов для аттестации и для оформления заявок на стимулирующие выплаты; формирование сводных отчетов по ПРНД и по эффективности научной деятельности по подразделениям и по организации в целом [1, 3]. На рисунке 1 представлен пользовательский интерфейс системы ПИН.

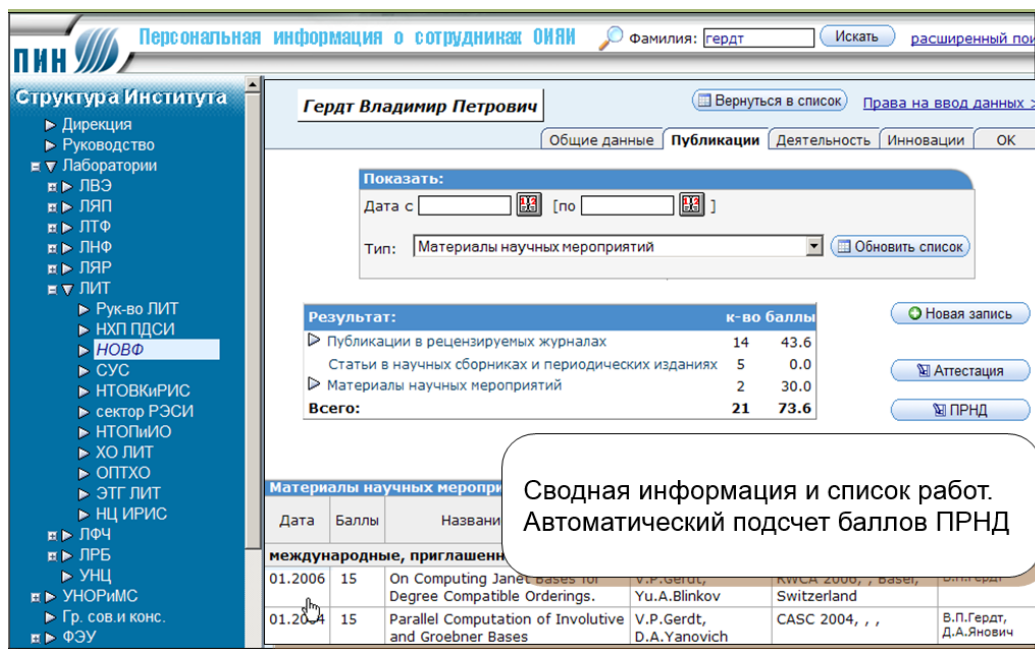


Рисунок 1. Пользовательский интерфейс ПИН

В системе предусмотрена гибкая настройка параметров, возможность расширения перечня собираемых данных и правил расчета индекса ПРНД.

В разработке PIN применены технологии и инструментальные средства, разработанные в сотрудничестве с ЦЕРН и успешно использовавшиеся при создании систем автоматизации административно - управленческой деятельности в ЦЕРН и ОИЯИ: JAVA Servlet, AJAX, DHTML, WALT (Web Application Lego Toolkit). Эти средства позволяют создавать в короткие сроки сложные и гибкие web-приложения с дружественным интерфейсом пользователя.

Эффективная эксплуатация системы предполагает, что научный сотрудник ОИЯИ своевременно размещает информацию о результатах своей работы в базе данных ПИН ОИЯИ. Ответственность за своевременное предоставление сведений и их наличие в базе данных ПИН несет оцениваемый сотрудник и его непосредственный руководитель.

Основные достоинства системы ПИН:

- Быстрый и удобный web-интерфейс для ввода, поиска и просмотра информации;

- Гибкая настройка реквизитов для различных типов публикаций и для других событий, включая правила расчета индекса ПРНД;

- Гибкий механизм задания и контроля прав доступа на ввод и на просмотр информации, сохранение истории изменения данных;

- Использование нечеткого поиска по базе публикаций для уменьшения количества дублирующих записей, возможность автоматизированного поиска и исключения дублей;

- Возможность ввода математических формул в формате TeX с последующим автоматическим конвертированием в HTML;

- Автоматическая подготовка пакета документов для аттестации и для оформления заявок на стимулирующие выплаты по индексу ПРНД.

Пакет документов для аттестации сотрудника, формируемой системой ПИН, представлен на рисунке 2.

В качестве дальнейшего развития системы ПИН планируется добавление полезного функционала: отображение финансовой информации о расходах; формирование сводных отчетов по подразделениям и по организации в целом.

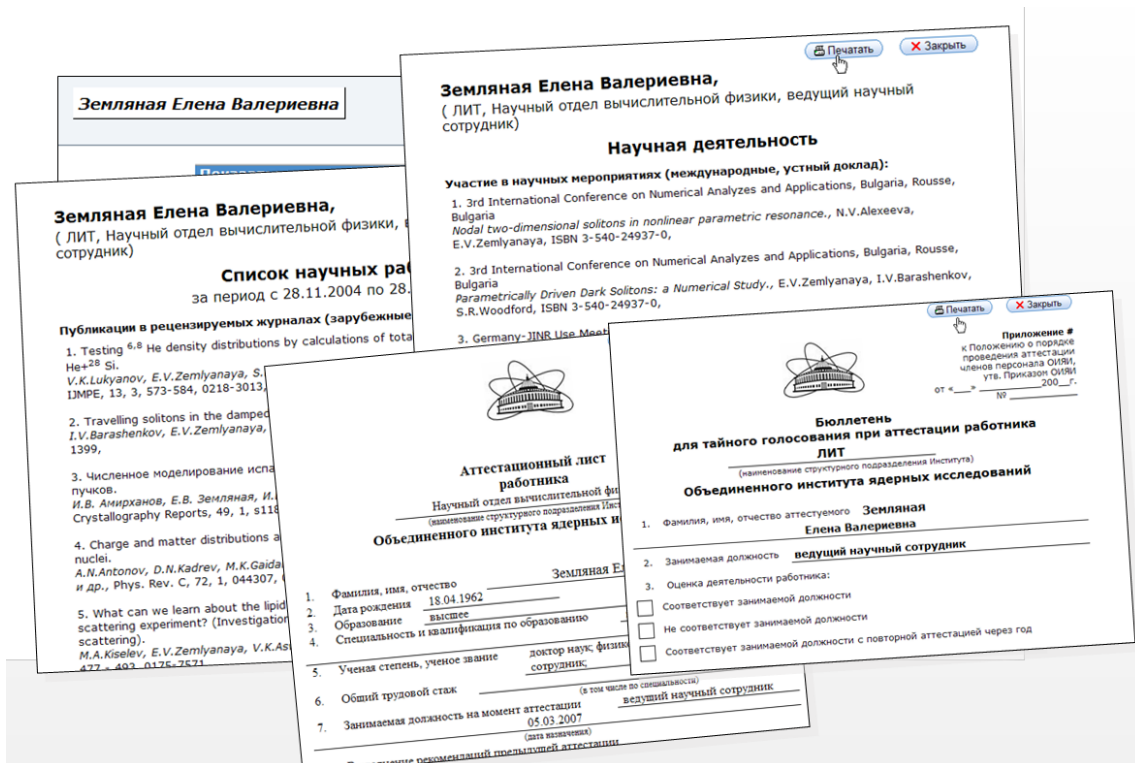


Рисунок 2. Пакет документов для аттестации персонала

2.2 JDS

Репозиторий научных публикаций ОИЯИ JDS создан на основе программного обеспечения Invenio, разработанного в ЦЕРН. JDS является репозиторием открытого доступа со следующим содержанием [3, 11]:

1. Научные и исследовательские документы:

– Публикации, изданные в соавторстве с учеными ОИЯИ;

– Архивные документы, описывающие все необходимые этапы исследовательской деятельности ОИЯИ;

2. Документы, обеспечивающие информационную поддержку научно-технических исследований, проводимых в ОИЯИ;

3. Аудио- и видеолекции, встречи, конференции.

Центральное место в архиве JDS занимают публикации сотрудников ОИЯИ. Но по мере необходимости и пожеланий пользователей в репозитории размещается любая полезная информация, необходимая для поддержки научного исследования, проекта, эксперимента и т. п. ОИЯИ участвует крупных международных экспериментах, требующих современной информационной поддержки. В репозитории могут быть размещены коллекции документов, отражающих процесс подготовки, создания экспериментальной установки и получение физических результатов [3].

Invenio, на основе которой создан сервер документов ОИЯИ JDS, имеет настраиваемый модуль WebStat, обеспечивающий сбор

статистических данных о таких параметрах как количество обращений, количество загрузок, наиболее популярные документы, интенсивность поисковых запросов, интенсивность пополнения коллекций и пр. [2, 15]. WebStat позволяет получить статистические отчеты в виде двухмерной графики. Модуль WebComment позволяет читателям оценивать документы. Таким образом, имеющийся инструментарий позволяет учитывать особенности социальной сети.

2.3 ПИН&JDS — на пути к CRIS

Рассмотрим перечень сформулированных требований к CRIS, работающих с научными документами [4, 6, 10, 16]:

- Полный охват информационных источников.
- Релевантность документов.
- Актуальность, полнота, достоверность происхождения документов.
- Наличие интеллектуальных служб обработки пользовательских запросов.
- Поддержка распределенных архитектур информационных систем.
- Структурирование и классификация контента.
- Предоставление информации пользователю в различных выходных форматах.
- Историчность информации.
- Поддержка различных уровней абстракции для представления информации.
- Архив.

Требованиями, предъявляемыми к CRIS, в условиях работы в распределенной среде, являются:

- Поддержка принятых стандартов метаданных для экспорта и импорта данных.
- Поддержка протоколов обмена информацией с другими информационными системами.
- Возможность проверки, экспертизы приходящей информации.
- Поддержка возможности ссылки на внутренние ресурсы.

Проанализировав на соответствие этим требованиям системы, разрабатываемые в ОИЯИ (ПИН и JDS), можно сделать вывод о том, что требования к CRIS, работающим с научными документами [5], реализуются в этих системах достаточно полно. Распределение контента выглядит следующим образом. В ПИН сформированы персональные профили сотрудников, доступна информация о проблемно-тематическом планировании, финансовых и материальных затратах. В систему ПИН внесены монографии и статьи сотрудников из русскоязычных журналов. На сервере JDS наиболее полно представлены публикации из зарубежных источников. Важно и то, что JDS интегрирован с разработанным в ЦЕРН пакетом Indico (INtegrated DIgital Conferencing), позволяющим управлять всем процессом создания научных мероприятий (конференции, семинары, лекции) и автоматизировать множество функций их организаторов в течение жизненного цикла конференции [13]. Таким образом, имеет место типичная проблемная ситуация — необходимая информация присутствует в информационном пространстве организации, но доступ к ней

возможен посредством соответствующих интерфейсов этих целевых систем.

Принимая во внимание вышесказанное, представляется целесообразным обеспечить интеграцию контента систем ПИН и JDS для реализации доступа к максимально полной информации как об авторах и сотрудниках института, так и об институте как организации.

3 Интеграция информационных ресурсов

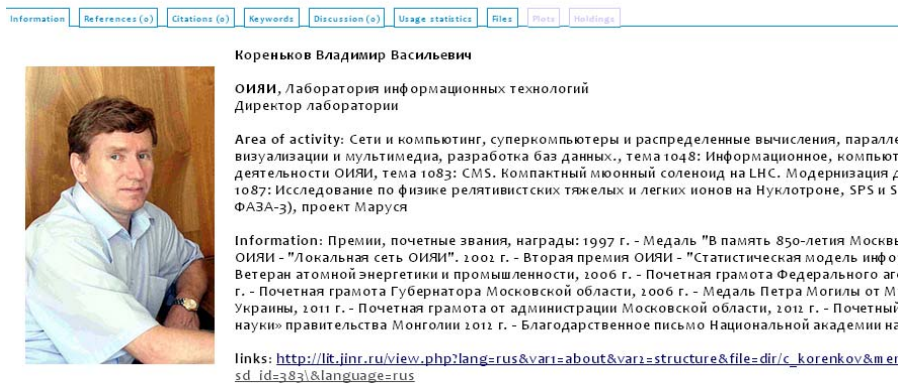
Далее рассмотрим пути решения сформулированной выше проблемной ситуации.

В ЛИТ ОИЯИ активно развивается и создается облачная инфраструктура, на одном из таких облачных серверов в начале года был размещен тестовый сервер репозитория JDS, который используется как полигон для наших исследований.

Представляются целесообразными для рассмотрения изложенные ниже варианты работы с библиографическими данными авторов репозитория.

3.1 Загрузка авторитетных файлов в JDS в формате MARCXML

Формат MARCXML предусматривает хранение авторитетных/нормативных файлов. Для обеспечения полноты авторитетных сведений в репозитории JDS (на тестовом сервере) создана коллекция «Персоналии», в которую данные авторов репозитория подгружаются в формате MARCXML из системы PIN. Описан соответствующий выходной формат, отображающий профиль автора с персональной информацией (рис. 3).



Information References (0) Citations (0) Keywords Discussion (0) Usage statistics Files Photo Holdings

Кореньков Владимир Васильевич

ОИЯИ, Лаборатория информационных технологий
Директор лаборатории

Area of activity: Сети и компьютеринг, суперкомпьютеры и распределенные вычисления, параллельная визуализация и мультимедиа, разработка баз данных., тема 1048: Информационное, компьютерное, сетевое взаимодействие, тема 1083: CMS. Компактный мюонный соленоид на LHC. Модернизация LHC, тема 1087: Исследование по физике релятивистских тяжелых и легких ионов на Нуклотроне, SP5 и S FA3A-3), проект Маруся

Information: Премии, почетные звания, награды: 1997 г. - Медаль "В память 850-летия Москвы" ОИЯИ - "Локальная сеть ОИЯИ". 2002 г. - Вторая премия ОИЯИ - "Статистическая модель инфо-коммуникационной энергии и промышленности, 2006 г. - Почетная грамота Федерального агентства по атомной энергии и промышленности, 2006 г. - Почетная грамота Московской области, 2006 г. - Медаль Петра Могилы от МГУ Украины, 2011 г. - Почетная грамота от администрации Московской области, 2012 г. - Почетный знак «Почетный работник» правительства Монголии 2012 г. - Благодарственное письмо Национальной академии наук Украины

links: http://lit.jinr.ru/view.php?lang=rus&var1=about&var2=structure&file=dir/c_korenkov&msd_id=3831&language=rus

Рисунок 3. Пользовательский интерфейс JDS, профиль автора

3.2 Настройка поисковой машины Invenio в JDS

Invenio – система, на базе которой основан репозиторий научных публикаций ОИЯИ JDS, позволяет осуществлять поиск и выводить результаты поисковых запросов из других информационных систем, подключая их к JDS как внешние коллекции. В настоящее время адаптируется и тестируется модуль внешних коллекций, настраиваемый на поиск информации в

системе ПИН. Это обеспечит доступ к необходимой информации о сотрудниках и их деятельности, не выходя из JDS.

3.3 Доработка авторитетных данных с интеграцией профиля авторов из InSpire

Информационная система InSpire, востребованная в физике высоких энергий, сочетает в себе содержимое баз данных SPIRES, DESY, Fermilab и SLAC [14, 15]:

- Быстрый многокритериальный поиск,
- Полнотекстовый поиск,
- Персональные настройки,
- Устранение омонимии по автору и т.д.

В системе реализована коллекция HEPNames — нормативные файлы авторов. На данный момент в InSPIRE описаны профили 900 авторов ОИЯИ. Профиль автора содержит всю необходимую информацию об авторе, его публикациях, информация о соавторах, метрические показатели и т.д. (рис. 4).

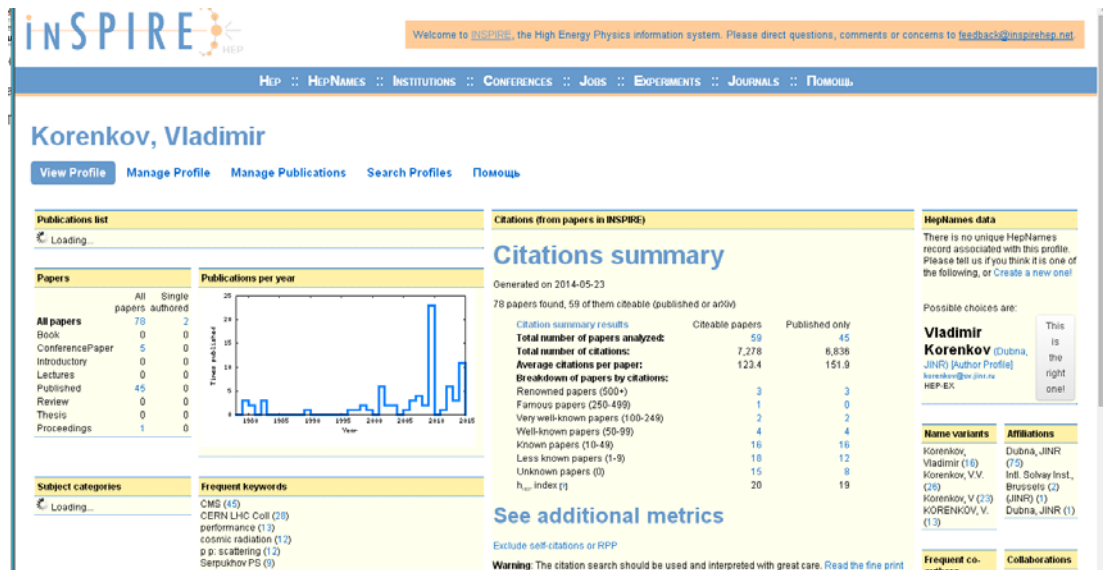


Рисунок 4. Пользовательский интерфейс InSPIRE, профиль автора

Модуль Citation Metrics позволяет получать и отображать следующую статистику цитирования для любой группы работ: общее количество публикаций, общее количество цитат, среднее число цитат на статью, общее количество цитат без самоцитирования. Так как InSPIRE разработана на базе того же программного продукта, что и JDS (Invenio), то в дальнейшем планируется доработка модулей JDS и перенесение информации о сотрудниках ОИЯИ из InSPIRE в JDS.

Работы авторов ОИЯИ в области физики высоких энергий в InSPIRE представлены достаточно полно, но в ней слабо представлены или отсутствуют работы авторов, чья научная деятельность выходит за рамки тематики InSPIRE. В настоящее время начата работа по верификации авторских профилей: добавление отсутствующих авторов, слияние различных вариантов имени, добавление метаданных.

4 Заключение

Конечной целью данной разработки является комплексная автоматизация научно-исследовательской деятельности на основе интернет-технологий.

Ожидаемые результаты:

- Отображение в открытом доступе текущих показателей деятельности научного сообщества, обновляемые в реальном времени.
- Оперативный мониторинг научной деятельности.

- Повышение мотивированности сотрудников профессионального коллектива.
- Повышение эффективности управленческих решений.

CRIS как современная модель комплексной научно-организационной информационной системы обеспечит интероперабельность, единый доступ к разнородным источникам, позволит управлять комплексом информационных объектов, охватывающих исследователей и результаты их деятельности.

Литература

- [1] Автоматизированная система ПИН (Персональная ИНформация) ОИЯИ [Электронный ресурс]. URL: <https://pin.jinr.ru/pin/pin>.
- [2] Архив публикаций CERN Document Server Европейского центра ядерных исследований [Электронный ресурс]. URL: <http://cds.cern.ch/>.
- [3] Заикина Т.Н., Филозова И.А. Система мониторинга и сбора статистики для оценки результативности научной деятельности в ОАИ- репозитории JINR Document Server // Труды 15-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» — RCDL'2013, Ярославль, Россия, 14–17 окт. 2013 г. – Ярославль: ЯрГУ, 2013. – С. 128–132.

- [4] Зелепухина, В.А. Концепция информационно-аналитической системы для сбора и анализа научной и наукометрической информации в организации [Текст] / В.А. Зелепухина, Ю.Ю. Тарасевич // Информатизация образования и науки. – 2013. – № 2. – С. 133–144. – ISSN 2073–7572.
- [5] Лопатенко А.С. Современные Научные Информационные Системы. Перспективы использования [Электронный ресурс]. URL: http://derpi.tuwien.ac.at/~andrei/CRIS_DOC.htm.
- [6] Паринов С.И. Развитие электронных библиотек – путь к Открытой Науке // Труды 11-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2009, Петрозаводск, Россия, 2009. – Петрозаводск: КарНЦ РАН, 2009. – С. 225–234.
- [7] Проект ЕНИП РАН. [Электронный ресурс]. URL: <http://enip.ras.ru/>.
- [8] Реестр репозитариев открытого доступа / Registry of Open Access Repositories. [Электронный ресурс]. URL: <http://roar.eprints.org/>.
- [9] Сысоев Т.М., Дьяконов И.А. Интеграция научной информации в ЕНИП // Новые технологии в информационном обеспечении науки: сб. науч. тр. / Каленов Н.Е. (ред.). – М.: Научный Мир, 2007. – С. 259–267.
- [10] DSPACE-CRIS. Technical Documentation Version 3.x.- URL:<https://cineca.github.io/dspace-cris>
- [11] Filozova I.A., Korenkov V.V., Musulmanbekov G. Towards Open Access Publishing at JINR. – Proceedings of XXII International Symposium on Nuclear Electronics and Computing (NEC'2009), Varna, Bulgaria, Sept. 7–14, 2009. – Dubna: JINR, 2010. – P. 124–128. – (JINR; E10,11-2010-22).
- [12] Horstmann Wolfram, Repositories for research information management, CERIF-CRIS and Repositories, Brussels, 12/13-oct-2011.
- [13] Indico JINR. [Электронный ресурс]. URL: <http://indico-new.jinr.ru/>
- [14] InSpire. [Электронный ресурс]. URL: <http://inspirehep.net/>.
- [15] Invenio-software. <https://Invenio-software.com>.
- [16] Jeffery K., Asserson A. CRIS and Institutional Repositories // Data Science Journal, Vol. 9, 24 July 2010.
- [17] OpenDOAR – Directory of Open Access Repositories. [Электронный ресурс]. URL: <http://www.opendoar.org/>.

Integration of JINR Scientific Information Resources on the JDS Platform

Tatiana N. Zaikina, Sergey V. Kunyaev,
Roman N. Semenov, Pavel V. Ustenko,
Irina A. Filozova, Galina V. Shestakova

Experience shows that the information needs of end-users are not covered by a single information system and in one organization typically multiple systems are supported.

Several information systems are provided for the information support of researches at JINR. They include server of the scientific documents JINR Document Server (JDS), Information System "Personal Information JINR employees", tools for the scientific activities management Indico. The paper considers the approaches providing for the content integration of these systems and their interoperability.