

# Открытый архив научных публикаций ОИЯИ

## JINR Document Server

Б.Ф. Борисовский, Т.Н. Заикина, С.В. Куняев, Ж.Ж. Мусульманбеков, П. Устенко,  
И.А. Филозова, Г. В. Шестакова

Объединенный институт ядерных исследований, г. Дубна

vborisov@jinr.ru, hostilemouse@mail.ru, kouniaev@jinr.ru, genis@jinr.ru, omega@jinr.ru,  
Irina.Filozova@jinr.ru, shestakova@jinr.ru

### Аннотация

Излагаются особенности создания в рамках Open Access Initiative репозитория научных публикаций и документов Объединенного института ядерных исследований (ОИЯИ) Jinr Document Server (JDS) на базе программного обеспечения CDS Invenio. Обсуждаются проблемы наполнения репозитория существующим контентом и возможность организации научной социальной сети на базе JDS в сочетании с другими информационными ресурсами ОИЯИ.

### 1 Введение

Извечная проблема скорейшего доведения авторами своих идей и новых результатов до научного сообщества способствовала постепенному переходу от распространения печатной продукции к оцифрованной электронной форме и, соответственно, к созданию многими институциональными и академическими организациями и университетами архивных баз данных библиографической и фактографической информации. В настоящее время многие научные и образовательные организации во всем мире создают в рамках ОАИ собственные репозитории, размещая в них различные документы как научного, так и организационного характера и предоставляя к ним открытый доступ для всего мирового сообщества. Репозитории открытого доступа существенно сокращают путь идеи автора до широкой научной аудитории, повышают ее цитируемость и престиж автора и, в конечном итоге, способствуют ускорению процесса ее внедрения и реализации, т. е. ускоряют научно-технический прогресс. В таком электронном архиве-репозитории документы хранятся в полнотекстовом формате (обычно PDF), а в случае регистрации репозитория в сети ОАИ-репозиториях (OpenDOAR, ROAR) обмен между ними осуществ-

ляется посредством протокола ОАИ-PMH. Пользователь репозитория, дислоцированного в каком-либо институте (университете), имеет доступ к документам любого репозитория, зарегистрированного в сети. В зависимости от профиля института/организации репозитории могут различаться тематической направленностью: фундаментальная или прикладная физика, астрономия, математика, химия, медицина и т. п. Институциональные репозитории, обладающие средствами электронных библиотек, в совокупности с другими информационными ресурсами данной организации расширяют возможности для интеграции всех этих ресурсов в единое (локальное) информационное пространство. Подобная интеграция позволяет создать виртуальную научную среду, в которой появляется возможность связывать информационные профили авторов – место работы, участие в различных проектах и другие виды научной деятельности – с традиционными научными публикациями и другими документами, хранящимися в репозиториях. Такие возможности при наличии средств визуального поиска и навигации являются предпосылками для организации научных социальных сетей как внутри одного института, так и в рамках сообщества организаций, объединенных родственной тематикой.

В разделе 2 даны структура репозитория JDS, инструментальные средства, предоставляемые программным обеспечением и источники наполнения. Перспективы интеграции информационных ресурсов ОИЯИ обсуждаются в разделе 3. Раздел 4 посвящен возможности создания социальной научной сети.

### 2 JDS

#### 2.1 Мотивация

Сотрудники ОИЯИ публикуют свои статьи в виде препринтов – в Издательском отделе ОИЯИ, электронных препринтов – в базах arXiv.org, MathSciNet и др., журнальных статей и монографий – в различных издательствах. Препринты Издательского отдела, изданные до 1991 года, хранятся в институтской библиотеке только в печат-

ной форме. Среди них немало работ выдающихся российских ученых, посвященных фундаментальным проблемам физики и являющихся актуальными по сей день. Перевод их в электронную форму сделал бы их доступными широкому кругу заинтересованных читателей, в числе которых могут быть не только сотрудники ОИЯИ, но и ученые из других институтов и стран. До выхода в печать принятых к публикации работ в реферируемых журналах проходят месяцы, что значительно ограничивает их доступность. Кроме того, доступ к полным текстам предоставляется только по подписке на соответствующие журналы, стоимость которой растет с каждым годом на 10–12%. В настоящее время с некоторыми издательствами достигнута договоренность о размещении в институциональных репозиториях статей авторов данного института, прошедших рецензирование и принятых к печати. В академических институтах России становится актуальной проблема оценки эффективности деятельности научных организаций и их сотрудников на основе показателей результативности научной деятельности (ПРНД). Одним из основных показателей является цитируемость работ автора. Как показывает анализ, проведенный в работах [1, 2], размещение препринтов в репозиториях многократно повышает их цитируемость в сравнении с их журнальными (пост-принт) версиями. Кроме того, посещаемость электронных архивов с полнотекстовыми статьями гораздо выше посещаемости сайта с соответствующими журнальными статьями. В 2009 г. были начаты работы по созданию репозитория ОИЯИ на базе сервера научных документов JINR Document Server (JDS) [3]. Основными направлениями деятельности ОИЯИ являются физика частиц высоких энергий и ядерная физика низких и промежуточных энергий. Вследствие тесного сотрудничества между ОИЯИ и ЦЕРН в области физики высоких энергий в качестве программного обеспечения сервера использован разработанный в ЦЕРН пакет CDS Invenio [4]. Он эксплуатируется в ряде родственных по тематике ОИЯИ институтов и университетов Европы

## 2.2 Содержание фонда JDS

На предыдущих этапах работы были определены цели создания JDS, сформулированы требования, исследованы программные продукты, применяемые для организации подобных систем [3]. С учетом результатов предпроектного исследования JINR Document Server реализуется как институциональный репозиторий, в фонде которого размещаются следующие информационные объекты:

1. Материалы научно-исследовательской и научно-организационной деятельности ОИЯИ, представленные любым типом произведения научной литературы:

- ✓ публикации сотрудников ОИЯИ – электронные коллекции материалов, авторами или соавторами которых являются сотрудники ОИЯИ, опубликованных в печатных изданиях;

- ✓ другие электронные документы, описывающие все существенные этапы научно-исследовательских работ, проводимых в ОИЯИ или с участием ОИЯИ.

2. Материалы, обеспечивающие информационную поддержку научных исследований ОИЯИ по основным и смежным направлениям деятельности;

Информационные потребности пользователей научных электронных библиотек (ЭБ), электронных информационных фондов (ЭИФ) определяются задачами, которые они решают в процессе своей профессиональной деятельности. Это в полной мере относится и к потребителям продукции JDS. Структура деятельности ученых и исследователей неоднородна и сочетает в себе несколько компонентов: научно-исследовательскую деятельность, научно-информационную работу, научное руководство и педагогическую работу, редакционно-издательскую деятельность, экспертизу, организационно-административную работу, материальное обеспечение исследований и пр. Безусловно, научно-исследовательская деятельность является основной. Поэтому центральное место в архиве JDS занимают публикации сотрудников ОИЯИ. Но по мере необходимости и пожеланий пользователей в репозиторий размещается любая полезная информация, необходимая для поддержки научного исследования, проекта, эксперимента и т. п. ОИЯИ участвует в нескольких крупных международных экспериментах, требующих современной информационной поддержки. В репозитории могут быть размещены коллекции документов, отражающих процесс подготовки, создания экспериментальной установки и получение физических результатов.

## 2.3 Программное обеспечение

В результате проведенного анализа программного обеспечения, применяемого для организации электронных архивов, был выбран пакет CDS Invenio [4]. Он является интегрированной цифровой библиотечной системой, предоставляющей среду и инструментарий для создания и поддержки цифрового библиотечного сервера. Инструментарий этой системы учитывает все возможные аспекты управления цифровой библиотекой. Кроме того, можно отметить некоторые особенности этой системы:

- мощная поисковая машина (специально разработанные индексы обеспечивают Google-подобные скорости для архивов до 1500000 записей); поисковый интерфейс доступен на 20-ти языках;
- одновременный поиск по метаданным, полным текстам и цитатам; результаты группируются по коллекциям;
- автоматизированное извлечение ключевых слов и ссылок из полнотекстовых документов; затем эти ключевые слова могут быть использованы для создания индексов и классификации документа, а ссылки – для оценки индекса цитирования автора или статьи и импакт-фактора журнала.

JDS развернут на базе следующего комплекта ПО: Linux SL5, Apache 2.2.3, Mysql 5.0.45, Python 2.4.3, CDS Invenio 0.99.1. Определенные усилия потребовались для адаптации и настройки соответствующих узлов и модулей CDS Invenio.

## 2.4 Источники наполнения JDS

После установки и настройки программного обеспечения встает вопрос о переводе существующего контента в создаваемый репозиторий. Одним из основных направлений научных исследований ОИЯИ является физика высоких энергий. Прежде чем посылать свои работы в реферируемые журналы, авторы помещают их в таких архивах, как arXiv.org, Mathathical Physics Preprint Server. Самым популярным среди этих полнотекстовых архивов является архив arXiv.org, предназначенный изначально для физики частиц высоких энергий, а затем расширивший свою тематику, включая весь спектр естественных наук и технологий. Вся библиографическая информация о публикациях со ссылками на полные тексты собирается в информационно-поисковых системах SPIRES (InSPIRE), CDS, ADS, MathSciNet, Zentralblatt Math и других. SPIRES (Stanford Physics Information Retrieval System) – это база данных, очень популярная среди физиков-теоретиков, специалистов в области физики высоких энергий, астрофизиков. Большое внимание уделяется созданию и комплектации системы цитирования работ. SPIRES содержит порядка 1 млн наименований, которые группируются в базы данных статей, конференций и книг. В настоящее время она преобразуется в информационную систему нового поколения iNSPIRE, которая сочетает базу данных SPIRES с развитыми технологиями электронных библиотек. Примерно такой же объем записей содержится в базе данных CDS (CERN Document Server), поскольку библиографические записи из SPIRES регулярно закладываются в CDS. Библиографическая информационно-поисковая система MathSciNet является дополнительным источником библиографических записей по математике и математической физике. В ОИЯИ ведутся исследования и в области радиохимии и радиобиологии, и библиографические данные о публикациях в этой области могут быть размещены в других базах данных, таких например, как INIS (International Nuclear Information System). После предварительного анализа контента этих баз данных основными источниками для наполнения JDS были выбраны arXiv.org, CDS и MathSciNet, поскольку в этих базах реализован максимальный охват статей, опубликованных когда-либо сотрудниками ОИЯИ. Загрузка полнотекстовых статей и библиографических записей из arXiv.org и CDS осуществляется в автоматическом режиме по протоколу OAI-PMH. Эта процедура осуществляется с помощью модуля BibHarvest, который настраивается на периодическую загрузку данных. В настройках указывается, откуда загружать данные, с какой периодичностью, каким образом преобразо-

вывать формат данных перед окончательным размещением их в репозитории. Автор, посылая новую работу в базу arXiv.org или в какой-нибудь реферируемый журнал, обязан депонировать рукопись в репозиторий. Если статья получила одобрение для публикации ее в журнале, он имеет возможность вносить изменения в библиографические данные и поместить в репозиторий журнальную версию статьи. Для реализации такой возможности были расширены и доработаны шаблоны форм пользовательского интерфейса и выходных форматов.

В ОИЯИ функционирует информационная система PIN (Персональная ИНформация), предназначенная для распределенного сбора и анализа информации о результатах научной деятельности сотрудников. Одними из показателей, характеризующих результаты научной деятельности сотрудника, являются количество публикаций и их качество, определяемое индексом цитирования. Между системой PIN и сервером JDS разработан канал обмена данными (импорт/экспорт) по протоколу OAI-PMH, что избавляет авторов от дублирования действий при внесении библиографических данных. Относительно полным источником библиографических данных работ сотрудников ОИЯИ является электронный каталог Научно-технической библиотеки ОИЯИ (JINR Library). Однако возможность автоматизированного обмена данными с сервером JDS в настоящий момент отсутствует, поскольку программное обеспечение АИБС Libex не поддерживает протокол OAI-PMH. Поэтому взаимное согласование библиографических данных сотрудников ОИЯИ осуществляется пока в ручном режиме.

Издательский отдел ОИЯИ (JINR Publishing Department) публикует препринты, монографии, авторефераты диссертаций, материалы конференций, проводимых в ОИЯИ, учебные материалы. Не все эти материалы доступны в электронной форме, особенно препринты, изданные до середины девяностых годов. В ближайшее время планируется организация работ по оцифровке этих документов, которые включают сканирование, распознавание и собственно оцифровку. Таким образом, основными способами наполнения и актуализации JDS являются (рис.1):

- ✓ автоматический сбор данных посредством Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) из arXiv.org, CDS;
- ✓ полуавтоматический режим сбора данных из информационно-поисковых баз INIS, MathSciNet;
- ✓ депонирование документов в режиме “самоархивирование”;
- ✓ импорт данных из ПИН как в автоматическом, так и полуавтоматическом режимах;
- ✓ загрузка оцифрованных ретроспективных материалов (препринтов, авторефератов, отчетов, монографий), сопровождаемая их библио-

графическими записями, если таковые в репозитории отсутствуют.

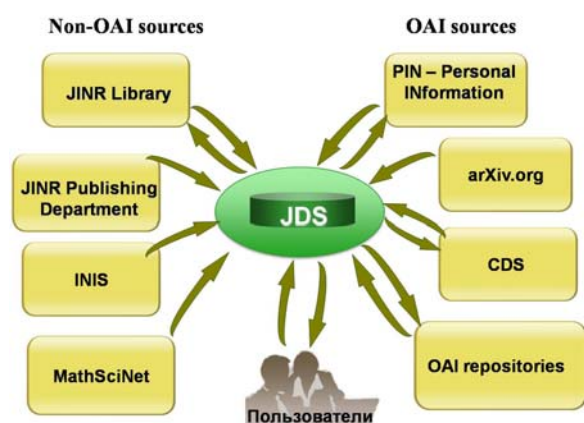


Рис. 1. Потенциальные источники наполнения JDS

Полуавтоматический режим предполагает сбор документов в ручном режиме, где в поисковом запросе присутствует название института (JINR), затем осуществляются конвертация в пакетном режиме этого набора во внутренний формат CDS Invenio и загрузка в архив. К настоящему моменту выполнена тестовая загрузка данных из ПИИ в JDS. Осуществляется загрузка на сервер JDS библиографических данных публикаций авторов ОИЯИ из баз arXiv.org, CDS и MathSciNet. На момент подготовки статьи общее количество загруженных документов составляло около 70000.

## 2.5 Дерево коллекций

Навигационный механизм CDS Invenio построен на иерархии (деревьях) коллекций, что позволяет создавать деревья со сложной многоуровневой структурой и в конечном итоге сужает и ускоряет поиск. Предусмотрено использование двух деревьев – базового и виртуального. За основу формирования основного дерева в JDS был взят классификационный признак – жанр (вид) научной публикации, виртуального дерева – тематика исследований и разработок.

Основное дерево:

### Articles & Preprints

- Articles
- Preprints

### Books & Reports

- Books
- Reports
- Conferences Proceedings

### Theses

### Presentations & Talks

- Talks
- Conferences Announcements
- Lectures
- Notes of Schools and Seminars

### Handbooks & Manuals

### Multimedia & Arts

- Pictures
- Posters
- Videos
- Audio

### Press

- "Dubna" Weekly
- Bulletins
- STL Bulletins

В ОИЯИ проводится широкий спектр исследований по различной тематике: теоретическая и экспериментальная физика высоких энергий, ядерная физика, физика ускорителей, конденсированные среды, нанотехнологии, информационные технологии и др. Виртуальное дерево позволяет выделить и сгруппировать документы по определенной тематике, по принадлежности к конкретному эксперименту, коллаборации и т. п.

Вот как выглядит один из фрагментов тематического дерева:

### Articles & Preprints

- Theoretical Physics
- Experimental Physics
- Nuclear Physics
- Accelerators
- Condensed Matter
- Information Technologies
- Nanotechnologies /NANO
- Computational Physics /CP

### Experiments in JINR

- NICA/MPD
- DELTA
- NIS
- MED-NUCLOTRON
- BECQUEREL
- STRELA
- MARUSYA
- LNS
- IREN

## 2.6 Пользователи и потребители JDS

Пользователю для осуществления поиска и/или просмотра документов в JDS не требуется регистрации. Однако для депонирования публикации нужно пройти процедуру аутентификации. Кроме того, доступ к некоторым коллекциям ограничен (Диссертации и авторефераты/Theses). Для этого используется механизм ролей, где пользователи принадлежат нескольким группам согласно их роли в системе:

- читатель (аноним) – пользователь с минимальным набором прав для поиска/просмотра коллекций;

- автор – зарегистрированный пользователь (сотрудник ОИЯИ), которому доступна процедура депонирования в режиме самоархивирования;
- каталогизатор – пользователь-эксперт, осуществляющий контроль с правами отклонения загрузки и удаления некорректных документов;
- менеджер данных осуществляет депонирование публикаций по доверенности;
- администратор архива осуществляет общее администрирование.

Кроме того, сбор данных из JDS может быть осуществлен внешними пользователями по протоколу OAI-PMH.

Пользователи имеют возможность создавать в среде JDS свои индивидуальные коллекции, организовываться в группы по интересам, участвовать в обсуждении работ авторов, предлагающих новые, неординарные решения актуальных проблем.

### 3 Перспективы развития JDS

Изначально репозиторий на сервере JDS предназначался для депонирования рукописей статей, которые в кратчайшие сроки становились доступными заинтересованным читателям для обсуждения и критических замечаний. По мере опубликования этих рукописей (препринтов) в реферируемых журналах библиографические записи модифицируются и дополняются ссылками и полными текстами опубликованных статей. В дальнейшем типы и тематика загружаемых документов будут расширены, и помимо препринтов и статей в журналах в состав архива будут включаться диссертации, книги, годовые отчеты, материалы семинаров, тексты докладов, презентаций и материалы видеоконференций. Наличие развитого пользовательского интерфейса и необходимого библиотечного сервиса, обеспечиваемого пакетом CDS Invenio, превращает репозиторий ОИЯИ на JDS в электронную библиотеку. В ближайшее время на сервере будет функционировать архив конференций, который создается с помощью пакета Indico, разработанного в ЦЕРН. Архив будет включать информацию и материалы состоявшихся и будущих рабочих совещаний, официальных встреч, конференций, комитетов и советов. Согласованный формат метаданных в обоих архивах позволит обмениваться документами по протоколу OAI-PMH.

Один из аспектов анализа результатов научной работы – публикации научных сотрудников, частота обращения к ним читателей, цитирование их в других работах. Эти данные и их статистический анализ могут быть получены средствами сервера JDS. Для применения такого анализа между базой данных ПИН и JDS разработан канал обмена метаданными по протоколу OAI-PMH, являющемуся стандартом для обмена между репозиториями открытого доступа. Модуль RefExtract извлекает ссылки из полного текста статьи, которые далее заносятся в соответствующие поля библиографического описания и затем анализируются на пред-

мет цитирования интересующего вас сотрудника или его статьи (рис. 2). Эти ссылки в сочетании с количеством обращений к данной статье и числом загрузок читателями, фиксируемыми модулем Bibrank, и являются одними из важных показателей творческой активности и популярности автора. Размещение публикации в репозитории открытого доступа способствует более быстрому продвижению новых идей, открытой дискуссии и обсуждению ее заинтересованными читателями, большей цитируемости и т. п.

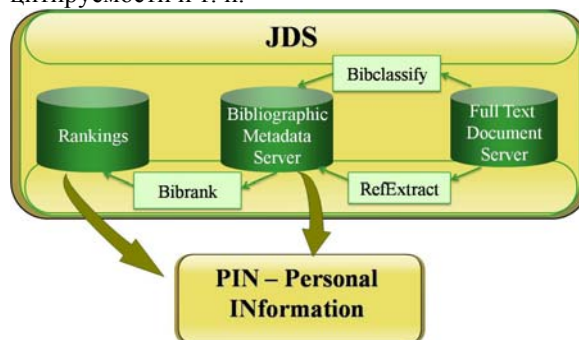


Рис. 2. Извлечение ссылок из полного текста статьи и числа обращений к ней для последующей оценки ПРНД

В перспективе планируется интеграция информационных ресурсов ОИЯИ, разнородных по структуре и контенту, в единое информационное пространство с развитой моделью структуры данных (на уровне метаданных) и средствами визуализации поиска и навигации.

### 4 Социальная научная сеть на базе информационных ресурсов ОИЯИ

Наличие множества информационных ресурсов, объединенных в единое информационное пространство, позволяет авторам устанавливать связи между объектами (данными), содержащимися в них. Конфигурации связей могут быть следующих типов:

- между ученым (его персональным профилем) и организацией – местом его работы;
- между ученым и его авторскими электронными материалами (в институтских ЭБ, научных журналах и т. п.);
- между авторскими электронными материалами ученого и другими материалами, которые были им использованы (процитированы) при подготовке своих работ, что позволяет устанавливать содержательные связи между различными учеными и научными материалами и подсчитывать уровень их использования в научном сообществе.

CDS Invenio имеет настраиваемый модуль WebStat, обеспечивающий сбор статистических данных о таких параметрах, как количество обращений, количество загрузок, наиболее популярные документы и пр. Есть возможность оценки документов читателями (модуль WebComment). Таким образом, модуль статистики WebStat совместно с

другим инструментарием – модулями WebBasket, WebGroup, WebMessage, WebComment – лежит в основе социальной функции сети, позволяет учитывать ее социальные особенности. Модуль WebComment обеспечивает социально-ориентированный инструмент ранжирования документов читателями и передачи своих замечаний, обмен мнениями по документам. WebMessage реализует кластеризацию (возможно, анонимных) пользователей в группы через веб-форумы. Пользователи JDS имеют возможность взаимодействовать между собой посредством сервиса для обсуждения публикации, образуя таким образом сеть коммуникаций в научном сообществе в рамках информационной системы. Анализируя эти связи, можно получать дополнительную важную информацию о характере этих взаимодействий. Например, по интенсивности связей между авторами, работающими в одном научном направлении (теме, проекте), можно судить о степени активности развития данного направления.

Для такого анализа становится актуальной задача визуализации социальной научной сети, поскольку визуализация позволяет делать быстрый и наглядный вывод о характере взаимодействия авторов, не прибегая к другим методом анализа сети. В настоящее время для решения этой задачи проводится тестирование свободно распространяемого пакета Prefuse, обладающего большой функциональностью и широкими возможностями визуализации социальных сетей. Prefuse поддерживает богатый набор функций для передачи данных моделирования, визуализации и взаимодействия, предоставляет оптимизированные структуры данных для представления таблиц, графиков, а также деревьев, поддержку анимации, динамические запросы, комплексный поиск и подключение базы данных. К настоящему моменту разработана и опробована модель пользовательского интерфейса к JDS со средствами визуальной навигации и поиска, позволяющая строить графы, связывающие между собой соавторов и их публикации. В дальнейшем предполагается подключить систему PIN для создания профилей авторов в узлах научной социальной сети.

## Литература

- [1] Antelman K. Do open-access articles have a greater research impact// *College and Research Libraries*. – 2004. – V. 65. – P. 372-382.
- [2] Gentil-Beccot A., Mele S., Brooks T.C. Citing and reading behaviors in high-energy physics// *Scientometrics*. – 2010. – No 84. – P. 345.
- [3] Борисовский В.Ф., Кореньков В.В., Куняев С.В., Мусульманбеков Ж.Ж., Никонов Э.Г., Филозова И.А. Организация открытого архива научных публикаций сотрудников ОИЯИ // Труды XI Всерос. конф. «Электронные библиотеки: перспективные методы и технологии,

электронные коллекции», RCDL'2009 – Петрозаводск: КарНЦ РАН, 2009. – С. 451-458.

- [4] CDSware Overview. – <http://cdsware.cern.ch/invenio/index.html>.
- [5] Prefuse: Information Visualization Toolkit. – <http://prefuse.org/>.

## Open Access Archive of scientific publications: JINR Document Server

V.F. Borisovski, I.A. Filozova, S.V. Kuniaev,  
G. Musulmanbekov, G. Shestakova, P. Ustenko,  
T. Zaikina

The paper concerns with the problems of building the OAI-compliant archive (repository) of the JINR staff members publications — JINR Document Server (JDS).