

Федеральное государственное бюджетное научное учреждение  
«Федеральный исследовательский центр информационных и вычислительных технологий»  
Федеральное государственное автономное образовательное учреждение высшего образования  
«Новосибирский национальный исследовательский государственный университет»  
Федеральное государственное бюджетное образовательное учреждение высшего образования  
«Новосибирский государственный технический университет»  
Федеральное государственное бюджетное образовательное учреждение высшего образования  
«Сибирский государственный университет телекоммуникаций и информатики»  
Федеральное государственное бюджетное научное учреждение  
Институт вычислительного моделирования Сибирского отделения Российской академии наук  
Федеральное государственное бюджетное учреждение науки  
Институт динамики систем и теории управления имени В.М. Матросова  
Сибирского отделения Российской академии наук



**Российская конференция с международным участием**

**РАСПРЕДЕЛЕННЫЕ  
ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫЕ  
РЕСУРСЫ  
(DICR-2022)**

**Сборник трудов**

*Россия, г. Новосибирск, 5 – 8 декабря 2022 г.*

Под редакцией  
С.А. Рылова, Ю.И. Молородова, А.А. Жирнова, Ю.Н. Синявского

Новосибирск – 2022

Federal Research Center for Information and Computing Technologies

Novosibirsk State University

Novosibirsk State Technical University

Siberian State University of Telecommunications and Information Sciences

Institute of Computational Modelling of the Siberian Branch of the RAS - Division of Federal Research  
Center «Krasnoyarsk Scientific Center of the Siberian Branch of the RAS»

Matrosov Institute for System Dynamics and Control Theory of the Siberian Branch of the RAS



## **XVIII Russian Conference with International Participation**

# **DISTRIBUTED INFORMATION AND COMPUTATIONAL RESOURCES (DICR-2022)**

## **Proceedings**

*Russia, Novosibirsk, December 5 – 8, 2022*

Edit by

S.A. Rylov, Yu.I. Molorodov, A.A. Zhirnov, Yu.N. Sinyavskiy

Novosibirsk – 2022

**Распределенные информационно-вычислительные ресурсы (DICR-2022) [Электронный ресурс]: Сборник трудов XVIII Российской конференции с международным участием (Новосибирск, 5-8 декабря 2022 г.) / Под ред. С.А. Рылова, Ю.И. Молородова, А.А. Жирнова, Ю.Н. Синявского. - Новосибирск: ФИЦ ИВТ, 2022. - 249 с. – ISBN: 978-5-905569-21-0.**

Тематика конференции ориентирована на проблемы разработки новых информационных технологий и созданию на их основе информационных систем. Обсуждались вопросы обеспечения информационной безопасности, решения задач с интенсивным использованием цифровых данных и других технологий, направленных на совершенствование методов и средств поддержки принятия решений. Особое внимание уделено технологиям создания цифровых двойников и связанных с ними задач анализа больших данных. Рассмотрены проблемы эволюции понятия «цифровой двойник», роли методов и технологий распределенной обработки и анализа разнородных данных. Обсуждались задачи компьютерного моделирования с усвоением данных мониторинга и наблюдений, теории управления и оптимизации в решении комплексной проблемы создания цифровых двойников для объектов и систем различной природы.

Предназначено для ученых, специалистов и преподавателей, работающих в области информатики и вычислительной техники.

**Distributed Information and Computational Resources (DICR-2022) [Electronic Resource]: Proceedings of the XVIII Russian Conference with International Participation (Novosibirsk, December 5-8, 2022) / S.A. Rylov, Yu.I. Molorodov, A.A. Zhirnov, Yu.N. Sinyavskiy (Eds.). - Novosibirsk: FRC ICT, 2022 - 249 p. - ISBN 978-5-905569-21-0.**

The theme of the conference is focused on the problems of development of new information technologies and creation of information systems based on them. The issues of ensuring information security, solving problems with intensive use of digital data and other technologies aimed at improving methods and means of decision support were discussed. Special attention is paid to the technologies of creating digital doubles and related tasks of big data analysis. The problems of the evolution of the concept of "digital double", the role of methods and technologies of distributed processing and analysis of heterogeneous data are considered. The problems of computer modeling with the assimilation of monitoring and observation data, control theory and optimization in solving the complex problem of creating digital twins for objects and systems of different nature were discussed.

It is intended for the scientists, experts and teachers working in the field of informatics and computer technology.

## ТЕХНОЛОГИЧЕСКИЙ КОМПЛЕКС ПОДГОТОВКИ И ПРЕДСТАВЛЕНИЯ МЕТАДАННЫХ О ПУБЛИКАЦИЯХ НАУЧНЫХ ОРГАНИЗАЦИЙ

*Мазов Н.А., Гуреев В.Н.*

Институт нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН, Новосибирск,  
Россия

Государственная публичная научно-техническая библиотека СО РАН, Новосибирск, Россия  
E-mail: MazovNA@ipgg.sbras.ru

Рассматриваются современные требования к подготовке, хранению и представлению метаданных о публикациях сотрудников научных организаций. Подчеркивается необходимость постоянной модернизации внутренних баз данных в ответ на перманентный рост числа новых элементов описания; обоснована целесообразность поддержания связей метаданных внутренней системы с элементами описания во внешних библиографических указателях. Сделан акцент на важности хранения и представления метаданных в разных форматах – как в целях снижения информационных рисков потери данных, так и в целях расширения доступа к ним.

*Ключевые слова:* публикационная активность, база данных, репозиторий, научная организация, метаописание.

DOI: 10.25743/DIR.2022.36.22.024

## TECHNOLOGICAL COMPLEX OF PREPARATION AND PRESENTATION OF METADATA ON ACADEMIC PAPERS OF SCIENTIFIC ORGANIZATION

*Mazov N.A., Gureyev V.N.*

Trofimuk Institute of Petroleum Geology and Geophysics SB RAS, Novosibirsk  
State Public Scientific Technological Library SB RAS, Novosibirsk

E-mail: MazovNA@ipgg.sbras.ru

The report discusses current requirements to preparation, storing and presentation of metadata on scholarly output of researchers at scientific organization. The authors (a) underline the need for permanent modernization of internal databases in response to growth in the number of new data fields describing the paper; (b) justify the necessity to provide links between internal metadata elements with that of external bibliographic databases; (c) highlight the importance of storage and presentation of metadata in different formats for decreasing information risks of data loss and enhancing open access to those data.

*Keywords:* scholarly output, databases, repository, research organization, meta description.

Наблюдаемый в последние десятилетия рост объемов научной информации затронул практически все формы ее функционирования и представления: ежегодно увеличивается количество названий периодических изданий, среднее число статей в журнале, объем аннотаций и списков литературы, число узких тематических направлений, стран и коллективов, участвующих в исследованиях. С учетом того, что широкое распространение в оценке научной информации приобрели формальные количественные подходы, включая анализ публикационной активности и цитируемости, все большую роль играют полнота и точность библиографической информации, форматы ее представления, «видимость» метаданных для внешних пользователей и прочие параметры, требования к которым постоянно повышаются.

В научных и образовательных организациях библиографические сведения о публикационной активности их сотрудников наиболее оптимально реализовывать в виде внутренних баз данных, что, согласно недавнему исследованию [1], в российских учреждениях происходит далеко не во всех случаях. На практике зачастую формируются

обычные списки публикаций в текстовом или табличном форматах. Чаще всего используется не отвечающая современным требованиям форма 3, в которой отсутствует большинство ныне необходимых сведений, может меняться порядок следования авторов, что затрудняет восстановление оригинальной записи. Этот подход полностью исключает функциональный поиск и какие бы то ни было аналитические возможности. Также организации могут прибегать к использованию внешних библиографических систем (РИНЦ, Scopus и пр.), аккумулирующих информацию по широкому спектру наук и источников. Преимущества и недостатки каждого из этих подходов к представлению данных о публикационной активности организации и ее оценке суммированы в таблице.

Таблица. Основные подходы к формированию библиографических сведений о публикуемости сотрудников организации

Подход к формированию библиогр. данных Критерий	Внутренние базы данных	Внешние базы данных	Списки публикаций
Полнота совокупности записей	высокая	средняя	средняя
Полнота одной записи	высокая	высокая	низкая
Точность данных	высокая	средняя (проблема однозначной идентификации)	низкая
Оперативность представления данных	высокая	средняя	высокая
Доступность данных	преимущественно открытый доступ	преимущественно платный доступ	преимущественно открытый доступ
Видимость данных для внешних пользователей	средняя	высокая	низкая
Поисковые возможности	высокие	высокие	отсутствуют
Набор библиометрических индикаторов	средний	расширенный	отсутствует
Аналитические возможности / библиометрический анализ	высокие	высокие	отсутствуют
Трудовые затраты	высокие	средние	средние
Временные затраты	высокие	средние	средние

Как видно из таблицы, наибольшими прикладными преимуществами обладают внутренние базы данных публикаций сотрудников научных и образовательных организаций, хотя они и сопряжены с высокими временными и трудовыми затратами, а также имеют меньшую в сравнении с внешними системами аудиторию. Хорошо зарекомендовавшими себя функциональными системами учета и представления публикаций сотрудников организаций являются разработки Межведомственного суперкомпьютерного центра РАН [2, 3], Института вычислительного моделирования СО РАН [4], Института катализа СО РАН [5, 6], БЕН РАН [7], в которых базы данных трудов сотрудников характеризуются исчерпывающей полнотой информации, взаимосвязью основных элементов описания, возможностью выгрузки данных в различных форматах. Ведение баз данных реализуется как за счет собственных разработок, так и с использованием специализированных систем, среди которых наиболее распространена автоматизированная библиотечно-информационная система «Ирбис», настраиваемая под задачи пользователей [8].

Широкие возможности формирования списков публикаций того или иного отдельного учреждения в настоящее время предполагают и внешние библиографические системы, что в итоге положительно влияет на узнаваемость и имидж организаций.

Однако почти во всех случаях эти системы будут уступать внутренним в оперативности отражения библиографических данных и их полноте, даже с учетом кропотливой корректировки данных представителями научных организаций в этих базах данных. Кроме того, ежегодно возрастающие требования к учету публикационной деятельности и отдельных сотрудников, и организаций в целом предполагают расширение перечня индексируемых сведений – наличия экспертных заключений, аффилированности с организацией, финансирования и пр., что в полной мере реализуемо лишь во внутренних системах.

Тем не менее, представляется оптимальным обеспечение связей внутренних систем с внешними базами данных, что дает несколько преимуществ. Так, заимствование библиографической информации из внешних систем существенно сокращает время ввода метаданных во внутреннюю систему [1, 9], программный интерфейс API позволяет отображать во внутренних системах необходимые библиометрические показатели публикаций, а точность и полнота сведений во внутренней системе могут служить основой для внесения своевременных корректировок во внешние базы данных [5].

Далее описан опыт ведения базы данных публикаций Института нефтегазовой геологии и геофизики им. А.А. Трофимука СО РАН (ИНГГ СО РАН), сотрудниками которого являются авторы доклада. База данных «Труды сотрудников ИНГГ СО РАН» (<http://ibc.ipgg.sbras.ru/>) реализована в среде автоматизированной библиотечно-информационной системы CDS-ISIS, включающей полные сведения о публикациях научных сотрудников института со времени его основания – 1957 г. При создании базы данных за основу была взята библиотечная картотека научных публикаций, преобразованная в 1996 г. в машиночитаемый вид. Тогда же данные картотеки были существенно дополнены публикациями из внешних баз данных РЖ ВИНТИ и Current Contents, а также на основе ежегодных списков публикаций, предоставляемых сотрудниками института в службу ученого секретаря. Для поиска публикаций во внешних системах использовались фамилии сотрудников института на основе данных из отдела кадров, а для поиска в ретроспективу – данные о сотрудниках из библиотечной картотеки.

База данных состоит из трех связанных между собой модулей – публикации, авторы и источники публикаций (рис. 1), зарегистрированных в Роспатенте [10, 11].

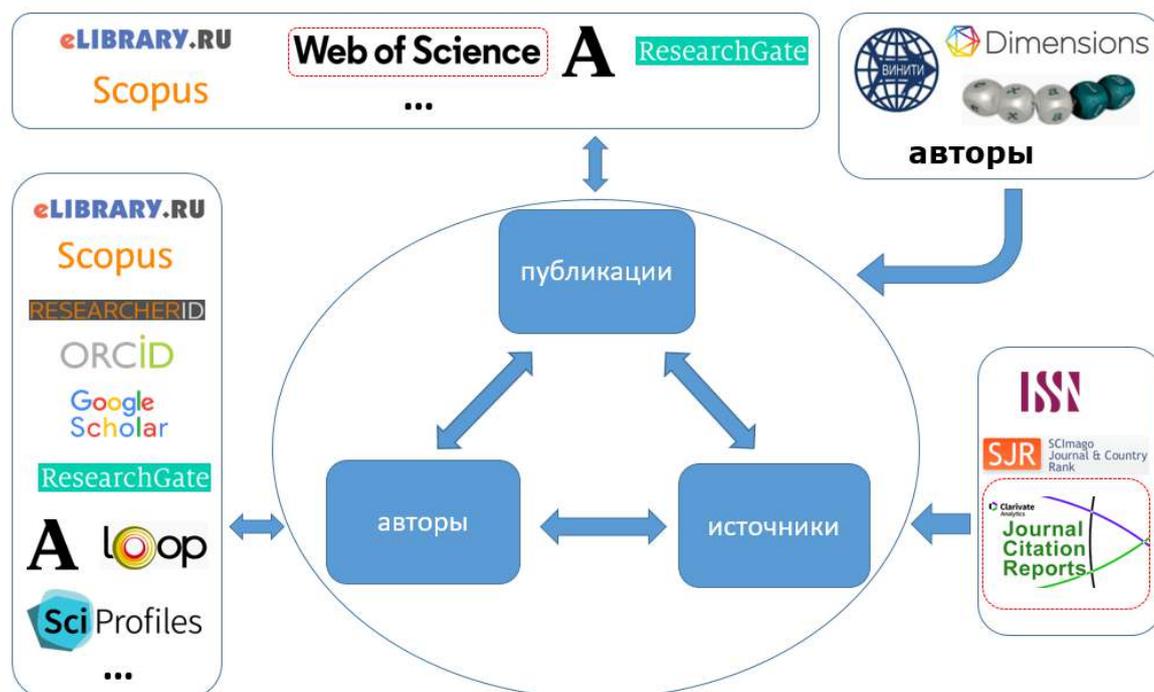


Рисунок 1. Комплекс баз данных «Труды сотрудников ИНГГ СО РАН», взаимосвязи между ними и источники наполнения.

1. Модуль описания публикаций содержит более 40 тыс. записей всех типов документов, включая исчерпывающий список работ, индексируемых в Научной электронной библиотеке eLibrary, Scopus и Web of Science. К традиционному набору обязательных элементов описания публикаций по ГОСТу поэтапно добавлялись такие элементы, как:

- аннотация и ее перевод (при наличии);
- ключевые слова и их перевод (при наличии);
- сведения об идентификационных номерах публикаций в РИНЦ (включая новый идентификатор EDN), Scopus и Web of Science, через которые реализован переход к описанию в этих системах;
- идентификатор публикации DOI;
- ссылка URL на страницу с полным текстом публикации;
- ISBN и тираж для монографий;
- номера программ и грантов, по которым финансируются исследования;
- тематические коды ГРНТИ, УДК, All Science Journal Classification из Scopus, по которым возможен тематический поиск в базе данных;
- число источников в пристатейной библиографии;
- номер внутренней экспертизы по экспортному контролю.

2. Модуль индексации авторов насчитывает более 29 тыс. записей, из них более 1000 уникальных авторских профилей сотрудников института за все годы, где в каждом случае учтены все возможные способы передачи имени автора. По каждому автору в системе содержатся следующие сведения:

- аффилированность с институтом (по умолчанию в базе данных учитываются все публикации сотрудника, независимо от указанной аффилиации);
- число аффилиаций, указанное в публикации каждым из сотрудников организации;
- сведения о наличии зарубежных соавторов / аффилиаций;
- ссылка на страницу сотрудника на сайте организации;
- ссылка на список публикаций автора;
- идентификаторы автора во внешних системах РИНЦ (включая SPIN-код), Scopus, Web of Science (ResearcherID), ORCID, Google Scholar, ResearchGate и др., снабженные гиперссылками для перехода в соответствующие внешние профили.

3. Модуль описания источников содержит информацию преимущественно о научных журналах, в которых опубликованы статьи сотрудников института. В настоящее время база данных насчитывает около 500 источников с указанием следующих характеристик:

- страна издания;
- коды печатного и электронного ISSN;
- сведения об индексации в ВАК и Russian Science Citation Index, обновляемые ежегодно;
- квартили журнала в Journal Citation Reports и SciMago Journal Rank, обновляемые ежегодно.

Процессы формирования баз данных включают поиск информации о публикациях и авторах в различных источниках и ее загрузку либо ручной ввод во внутреннюю систему. Поиск публикаций во многом основан на постоянном мониторинге РИНЦ, РЖ ВИНТИ, Scopus и Web of Science. Информация от самих сотрудников занимает незначительную долю и затрагивает лишь публикации, отсутствующие в каких-либо индексирующих системах. В Scopus используется комплексный запрос, включающий возможные вариации в написании названия аффилиации, а также запрос по всем сотрудникам организации по их внутренним идентификаторам AuthorID. В Web of Science в связи с введенными в 2022 г. ограничениями к полному модулю используется

бесплатная версия, где поиск может осуществляться по авторским профилям зарегистрированных в системе сотрудников института.

Метаданные в систему вводятся преимущественно в полуавтоматическом режиме: предварительно выгруженные из внешних систем, описания публикаций включаются во внутреннюю базу данных и дорабатываются с учетом периодически обновляемых требований. Важным этапом в работе с базой данных является ее дополнительное представление на сайте организации, где используется упрощенный и более привычный современным пользователям интерфейс и навигация. Данная web-реплика, реализованная отделом информационных технологий ИНГГ СО РАН, повышает видимость записей базы данных в сети Интернет, популяризируя научные достижения организации, а также снижает информационные риски утраты данных, по сути представляя собой копию основной базы данных в другом формате.

Описанный технологический комплекс, реализуемый в ИНГГ СО РАН, как и другие подобные технологические подходы российских организаций, требуют постоянного внимания и доработки в ответ на все новые информационные потребности ученых и запросы проверяющих инстанций. Однако в результате они отвечают всем современным задачам по учету и представлению публикаций сотрудников научных и образовательных учреждений, среди которых:

- полнофункциональный поиск информации по широкому набору критериев;
- получение большого спектра библиометрических индикаторов для отчетных, поисковых, исследовательских и иных целей;
- возможность экспорта данных в любом заданном формате;
- интеграция с описаниями публикаций и авторскими профилями во внешних системах;
- удаленный доступ к базам данных и полнотекстовым репозиториям (при наличии) для удобства пользователей.

## ЛИТЕРАТУРА

- [1] Мазов Н.А., Гуреев В.Н. Базы данных публикаций научной организации как основа информационных исследований // Научно-техническая информация. Серия 1: Организация и методика информационной работы. 2022. № 5. С. 8-18.
- [2] Власова С.А., Каленов Н.Е. Многофункциональная веб-система регистрации и учета результатов интеллектуальной деятельности ученых // Научные и технические библиотеки. 2022. № 2. С. 29-48.
- [3] Власова С.А. Автоматизированная система поддержки базы данных научных трудов сотрудников академических учреждений // Информационные ресурсы России. 2020. № 5. С. 29-31.
- [4] Ковязина Е.В. БД трудов сотрудников как средство учета и продвижения научных публикаций // Труды ГПНТБ СО РАН. 2017. № 12-2. С. 336-343.
- [5] Альперин Б.Л., Ведягин А.А., Зибарева И.В. SciAct – информационно-аналитическая система Института катализа СО РАН для мониторинга и стимулирования научной деятельности // Труды ГПНТБ СО РАН. 2015. № 9. С. 95-102.
- [6] Альперин Б.Л., Зибарева И.В., Ведягин А.А. Анализ скорости публикации научных статей с использованием CRIS-системы SciAct // Библиосфера. 2020. № 1. С. 83-92.
- [7] Захарова С.С., Гуреева Ю.А. Научные публикации: от картотеки трудов до библиографических профилей // Библиосфера. 2017. № 2. С. 85-89.
- [8] Баженов С.Р., Rogoznikova O.A. Научная публикация как специфический объект описания и требования к базе данных научных публикаций // Материалы Третьего международного профессионального форума «Книга. Культура. Образование. Инновации». Москва: Изд-во ГПНТБ России, 2017. С. 277-280.
- [9] Баженов С.Р., Данилин М.В., Rogoznikova O.A. Интеграция базы данных публикаций организации с индексами научного цитирования: реализация средствами САБ ИРБИС64 // Труды 22-й Международной конференции «Библиотеки и информационные ресурсы в

- современном мире науки, культуры, образования и бизнеса». Москва: Изд-во ГПНТБ России, 2015. С. 1-4.
- [10] *Мазов Н.А., Гуреев В.Н.* IPGGTR Труды сотрудников ИНГГ СО РАН (реферативно-полнотекстовая библиография): Свидетельство о государственной регистрации программы для ЭВМ // Свид-во о прогр. 2020621025; RU; № 2020620872, заявл. 10.06.2020, опубл. 19.06.2020, ИНГГ СО РАН.
- [11] *Мазов Н.А., Гуреев В.Н.* IPGGAU Авторские идентификационные профили: Свидетельство о государственной регистрации программы для ЭВМ // Свид-во о прогр. 2020621128; RU; № 2020620879, заявл. 10.06.2020, опубл. 02.07.2020, ИНГГ СО РАН.