НАУКОМЕТРИЯ. БИБЛИОМЕТРИЯ

УДК [001.83:01]-047.44+004:002.1-021.341
https://doi.org/10.33186/1027-3689-2022-12-15-34

Оценка качества открытых данных Роспатента в контексте интеграции с отечественными информационными системами текущих исследований

В. А. Зелепухина

Астраханский государственный университет им. В. Н. Татищева, Астрахань, Российская Федерация, v.zelepukhina@asu.edu.ru, https://orcid.org/0000-0001-9339-1589

Аннотация. Информационные системы текущих исследований (Current Research Information Systems, CRIS) агрегируют сведения о научно-исследовательских проектах организации и их финансировании, о публикациях сотрудников и объектах интеллектуальной собственности. На основе данных, представленных в CRIS, проводится наукометрический анализ, оцениваются результативность научной деятельности и инновационный потенциал организации, принимаются управленческие решения. Поэтому своевременная загрузка качественной и достоверной информации – важная задача для CRIS. Потенциальным источником сведений об объектах интеллектуальной собственности (патентах и свидетельствах о государственной регистрации) для отечественных CRIS являются открытые данные (ОД) Роспатента, которые, согласно концепции ОД, допускают автоматизированную обработку на условиях свободной лицензии (бесплатно). Исследования показывают, что, несмотря на публикацию ОД в машиночитаемых форматах, их практическое применение осложняется наличием некорректных, неполных и несогласованных записей. Поэтому перед загрузкой ОД Роспатента в CRIS требуются предварительная оценка качества данных и их улучшение, если это возможно. К настоящему времени качество ОД Роспатента исследовано по нескольким критериям: доступность, заполненность метаданных, наличие обратной связи. Оценка качества данных на уровне содержимого наборов не проводилась. Цель настоящей работы – оценить внутреннее качество наборов ОД Роспатента, включающих сведения об изобретениях, полезных моделях, промышленных образцах, программах для ЭВМ, базах данных, топологиях интегральных микросхем, в контексте интеграции этих данных с системами CRIS. Качество измерялось по следующим характери-

стикам: полнота, точность, согласованность, своевременность и актуальность. В результате исследования выявлены неполные, неточные и несогласованные записи.

Ключевые слова: CRIS, информационные системы текущих исследований, качество данных, измерение качества данных, оценка качества данных, открытые данные, Роспатент, объекты интеллектуальной собственности, патенты

Для цитирования: Зелепухина В. А. Оценка качества открытых данных Роспатента в контексте интеграции с отечественными информационными системами текущих исследований / В. А. Зелепухина // Научные и технические библиотеки. 2022. № 12. С. 15–34. https://doi.org/10.33186/1027-3689-2022-12-15-34

SCIENTOMETRICS. BIBLIOMETRICS

UDC [001.83:01]-047.44+004:002.1-021.341
https://doi.org/10.33186/1027-3689-2022-12-15-34

Assesment of Rospatent open data quality within the context of integration with national current research information systems

Victoria A. Zelepukhina

V. N. Tatishchev Astrakhan State University, Astrakhan, Russian Federation, v.zelepukhina@asu.edu.ru, https://orcid.org/0000-0001-9339-1589

Abstract. Current Research Information Systems (CRIS) are to aggregate data on organizational research projects and their funding, on employers' publications and intellectual property subject matters. The scientometric analysis is based on CRIS data to assess research output and innovative potential of organizations, and to make management solutions. The early loading of quality and reliable information is the most important task for CRIS. The Rospatent open data that allow for automated processing based on free (free of charge) licensing makes the potential source for data on intellectual property subject matters (patent and government

registration certificates) for national CRIS. The studies evidence that despite publishing OD in machine-readable formats, their practical application is impeded by incorrect, incomplete and uncoordinated entries. Therefore, before loading. Rospatent OD have to be assessed for quality and to be improved, if possible. As for today, Rospatent OD quality is assessed by several criteria: accessibility, metadata completeness, feedback. However, at the content level, the open data have not been assessed. The purpose of the article is to evaluate the internal quality of Rospatent OD sets including information on inventions, utility models, industrial designs, computer programs, databases, circuit layouts, within the context of OD integration in CRIS systems. The quality is assessed in several characteristics: completeness, accuracy, consistency, expedience, and relevancy. The study has revealed incomplete, inaccurate and uncoordinated entries.

Keywords: CRIS, current research information system, data quality, data quality measurements, data quality assessment, open data, Rospatent, intellectual property subject matters, patents

Cite: Zelepukhina V. A. Assesment of Rospatent open data quality within the context of integration with national current research information systems / V. A. Zelepukhina // Scientific and technical libraries. 2022. No. 12. P. 15–34. https://doi.org/10.33186/1027-3689-2022-12-15-34

Введение

Информационные системы текущих исследований (Current Research Information Systems, CRIS) предназначены для сбора и хранения информации о публикациях, патентах и других объектах интеллектуальной собственности (ОИС), грантах, участии в мероприятиях. В России разработаны такие CRIS, как «ИСТИНА» [**1**], Sciact [**2**], ИАС РНД АГУ [**3**].

CRIS используются при аттестации научно-педагогических работников, определении стимулирующих выплат, для оценки результативности научно-исследовательской деятельности и научно-инновационного потенциала организации. Поэтому данные, представленные в CRIS, должны быть высокого качества [**4**, **5**]. Качество данных можно понимать как «степень, с которой набор характеристик, присущих данным, отвечает конкретным требованиям с точки зрения их применения» [**6**].

Качество данных в CRIS повышает интеграция с внешними источниками информации. С одной стороны, это позволяет своевременно загрузить в CRIS новые данные, а с другой – верифицировать и актуализировать уже имеющиеся [**7**].

Так как не у всех научно-образовательных организаций имеется коммерческая подписка к внешним базам данных (БД) и сервисам, то поиск бесплатных и достоверных источников информации является одной из важных задач при разработке и поддержке таких систем.

Для автоматизации процессов управления информацией об ОИС, зарегистрированных в Федеральной службе по интеллектуальной собственности (Роспатент), особый интерес представляют открытые данные (ОД) Роспатента [**8**], содержащие сведения из открытых реестров (ОР) [**9**] изобретений (ИЗ), полезных моделей (ПМ), промышленных образцов (ПО), программ для ЭВМ (ПрЭВМ), БД и топологий интегральных микросхем (ТИМС). Владельцем наборов данных выступает Федеральный институт промышленной собственности (ФИПС), являющийся подведомственным учреждением Роспатента.

Так как ОД Роспатента доступны для автоматизированной обработки бесплатно [**10**], их использование в качестве внешнего источника данных может снизить временные и финансовые затраты, необходимые на загрузку и обновление информации в CRIS.

Однако на практике наборы ОД зачастую оказываются малосодержательными, плохо структурированными, неточными и несвоевременными [**11**–**15**]. Подобные проблемы выявляются измерением и оценкой качества данных.

В рейтинге открытости органов исполнительной власти, составленном по результатам исследования, проведённого в 2021 г. Счётной палатой РФ совместно с АНО «Информационная культура» и Центром перспективных управленческих решений [**16**], ОД Роспатента получили высокую оценку. Однако анализ содержимого ОД не проводился. Целью настоящей работы являются измерение и оценка внутреннего качества наборов ОД Роспатента в контексте интеграции этих данных с системами CRIS.

Статья организована следующим образом. В первой части рассмотрены общедоступные источники информации об ОИС, зарегистрированных в РФ, во второй – структура и форматы наборов ОД Роспатента,

в третьей сформулированы требования к качеству ОД Роспатента, названы метрики и подходы, использованные в процессе измерения качества. В четвёртой части приведены результаты измерения качества ОД Роспатента и описаны обнаруженные в наборах проблемы.

Интернет-источники сведений об ОИС, зарегистрированных в РФ

Нами был проведён сравнительный анализ общедоступных и авторитетных источников информации о зарегистрированных в РФ ОИС (табл. 1).

Обнаружено, что возможность выгрузки данных в машиночитаемых форматах реализована в Google Patents, Espacenet, eLIBRARY, Designview и ОД Роспатента. При этом eLIBRARY предоставляет такую возможность исключительно на коммерческой основе.

Среди интернет-источников, бесплатно предоставляющих данные в машиночитаемых форматах, лишь ОД Роспатента имеют наибольший охват по видам ОИС. Кроме того, на момент наблюдения этот источник предоставлял самые актуальные данные. Таким образом, ОД Роспатента можно рассматривать как авторитетный, бесплатный и доступный источник информации об ОИС, зарегистрированных в РФ.

Таблица 1

Сравнительные характеристики интернет-источников сведений об ОИС, зарегистрированных в РФ (дата обращения: 28.01.2022 г.)

Охват по видам ОИС	Дата актуальности	Возможность выгрузки	Периодичность обновления
1. Поисковая система Espacenet [**17**]			
ИЗ, ПМ	20.12.2021 г.	CSV, XLS; бесплатно	Не указана
2. Поисковая система Designview [**18**]			
ПО	27.02.2022 г.	PDF, XLS, DOC; бесплатно	Ежедневно

Продолжение таблицы 1

Охват по видам ОИС	**Дата актуальности**	**Возможность выгрузки**	**Периодичность обновления**
3. Поисковая система Google Patents [**19**]			
ИЗ, ПМ	27.12.2021 г.	CSV, XLSX; бесплатно	Не указана
ПО	16.04.2009 г.		
4. Поисковая система «Яндекс.Патенты» [**20**]			
ИЗ, ПМ	10.11.2021 г.	Нет	Не указана
5. Электронная библиотека eLIBRARY [**21**]			
ИЗ, ПМ, ПО	27.12.2021 г.	API; платно	Не указана
ПрЭВМ, БД	20.12.2021 г.		
ТИМС	17.12.2021 г.		
6. Открытые реестры ФИПС [**9**]			
ИЗ, ПМ	27.01.2022 г.	Нет	Ежедневно
ПрЭВМ, БД	27.01.2022 г.		
ТИМС	29.12.2021 г.		
7. Официальные публикации Роспатента [**22**]			
ИЗ, ПМ, ПО, ПрЭВМ, БД	28.01.2022 г.	Нет	Непрерывно
ТИМС	25.01.2022 г.		
8. Информационно-поисковая система ФИПС [**23**]			
ИЗ, ПМ, ПО	27.01.2022 г.	Нет	Ежедневно
ПрЭВМ, БД	27.01.2022 г.		
ТИМС	25.01.2022 г.		

Окончание таблицы 1

Охват по видам ОИС	Дата актуальности	Возможность выгрузки	Периодичность обновления
9. Наборы ОД Роспатента [**8**]			
ИЗ, ПМ, ПрЭВМ	30.12.2021 г.	CSV; бесплатно	Ежемесячно
ПО	28.12.2021 г.		
БД, ТИМС	29.12.2021 г.		

Структура наборов ОД Роспатента и форматы значений

Каждый вид ОИС представлен в ОД Роспатента отдельным набором данных. Данные публикуются в формате CSV и поэтому имеют табличную структуру: каждый ОИС представлен в отдельной строке (далее – запись), а его характеристики – в столбцах (далее – атрибуты).

Обновление наборов ОД происходит путём публикации нового CSV-файла, который содержит записи из предыдущей версии, сведения о новых ОИС и изменения в ранее зарегистрированных.

Наибольшее количество записей содержится в наборе ИЗ, наименьшее – в наборе ТИМС (табл. 2).

Таблица 2

Количественные характеристики наборов ОД Роспатента (дата актуальности: 01.01.2022 г.)

Характеристика	ИЗ	ПМ	ПО	ПрЭВМ	БД	ТИМС
Количество записей	759 006	205 414	89 133	125 022	16 662	1 761
Количество атрибутов	46	38	34	14	20	17

В каждом наборе данных присутствует минимально необходимый для CRIS набор атрибутов:

название ОИС *(title)*;

сведения об авторах *(authors)*;

сведения о патентообладателях и правообладателях *(holders)*;

регистрационный номер (номер свидетельства или номер патента – *reg_number*);

дата регистрации *(reg_date)*;

номер заявки *(app_number)*;

дата подачи заявки *(app_date)*;

статус действия правовой охраны для патентов *(actual)*;

адрес соответствующего документа в ОР на сайте ФИПС *(url)*, позволяющего получить полную информацию об ОИС.

Предварительный просмотр случайно отобранных записей показал, что в значениях атрибутов *reg_number*, *reg_date*, *app_number*, *app_date* содержатся только цифры.

В атрибутах *authors* и *holders* имена разделяются комбинацией символов \r\n, после каждого имени автора и патентообладателя (правообладателя), составленного с использованием символов кириллицы, в круглых скобках указывается двухбуквенный код страны места его нахождения или жительства. Наличие такого формата позволит с помощью регулярных выражений извлекать имя каждого автора и патентообладателя (правообладателя) по отдельности и выполнять связывание с профилями сотрудников и организаций, зарегистрированных в CRIS.

Методы

В настоящем исследовании оценка качества данных проводилась в отношении версий наборов ИЗ, ПМ, ПО, ПрЭВМ, БД и ТИМС, опубликованных 26.01.2022 г. (дата актуальности: 01.01.2022 г.).

Качество ОД оценивалось по полученным в процессе измерения качества данных результатам [**24**–**32**], по характеристикам, актуальным для CRIS [**4, 5**]. Среди этих характеристик: *полнота*, *точность* (правильность, корректность), *согласованность* (непротиворечивость, целостность, уникальность), *своевременность* (актуальность).

Для исследуемых наборов ОД были сформулированы следующие требования:

1. В наборах ОД должны присутствовать записи обо всех ОИС из ОР на сайте ФИПС.

2. В каждой записи должны быть заполнены все обязательные атрибуты.

3. Значения исследуемых атрибутов должны совпадать с характеристиками ОИС, опубликованными в ОР на сайте ФИПС.

4. Значения атрибутов, содержащих сведения о дате регистрации и дате подачи заявки, не должны превышать дату актуальности версии набора.

5. Атрибуты не должны содержать фиктивных значений, представляющих собой многократные повторы одного и того же символа (например, 00000000, #####).

6. Значения атрибутов должны соответствовать формату, определённому на этапе предварительного просмотра данных (см. п. 2).

7. В наборах не должно быть записей, связанных с ОИС, отсутствующими в ОР на сайте ФИПС.

8. В атрибутах, содержащих регистрационный номер ОИС и номер заявки, не должно быть повторяющихся значений.

9. Значения атрибутов, содержащих дату регистрации и дату заявки, должны быть согласованы.

10. Все заявленные версии наборов должны быть опубликованы.

11. Все опубликованные версии наборов должны быть актуальны на дату публикации.

На основе сформулированных требований определены метрики (табл. 3, табл. 4, табл. 5, табл. 6, представляющие собой соотношения вида [**26**]:

$$X = \frac{A - B}{A},$$

где B – количество отбракованных данных (ячеек, записей, атрибутов), A – количество проверенных данных (ячеек, записей, атрибутов). При этом $X = 1$ означает наилучшее качество данных, а $X = 0$ – наихудшее.

Таблица 3

Метрики для измерения полноты

Требование	Метрика
1	M1. Доля документов из ОР ФИПС, которые включены в набор ОД.
2	M2. Доля записей, у которых атрибуты *reg_number*, *reg_date*, *app_number*, *app_date*, *title*, *holders*, *actual* (для патентов) содержат непустую строку.

Таблица 4

Метрики для измерения точности

Требование	Метрика
3	M3. Доля записей, у которых значения атрибутов *reg_number*, *reg_date*, *app_number*, *app_date*, *title*, *authors*, *holders*, *actual* (для патентов) совпадают с соответствующими характеристиками в ОР на сайте ФИПС.
4	M4. Доля записей, у которых непустые значения атрибутов *reg_date* и *app_date* не превышают дату актуальности набора
5	M5. Доля записей, в атрибутах которых отсутствуют фиктивные значения (см. требование T5).
6	M6. Доля записей, у которых значения атрибутов *reg_number*, *reg_date*, *app_number*, *app_date*, *authors*, *holders* соответствуют форматам, определённым на этапе предварительного просмотра данных (см. п. 2).

Таблица 5

Метрики для измерения согласованности

Требование	Метрика
7	M7. Доля записей, у которых значение атрибута *url* указывает на документ, действительно опубликованный в ОР на сайте ФИПС.
8	M8. Доля записей, в которых значения *app_number* (кроме патентов) и *reg_number* не встречаются в других записях набора.
9	M9. Доля записей, у которых значение *app_date* не превышает значение *reg_date*.

Таблица 6

Метрики для измерения своевременности

Требование	Метрика
10	M10. Отношение количества фактически опубликованных версий наборов ОД к запланированному количеству (на дату измерения).
T	M11. Доля опубликованных версий наборов ОД, у которых дата публикации совпадает с датой актуальности данных.

Для вычисления метрик M1 и M7 предварительно были извлечены адреса документов из соответствующих ОР на сайте ФИПС. На основе этих данных был сформирован новый набор (далее – индекс ОР ФИПС). Если адрес документа из индекса ОР ФИПС отсутствовал в наборе ОД, то этот документ считался отсутствующим в ОД. И, наоборот, если значение атрибута *url* набора ОД отсутствовало в индексе ОР ФИПС, то запись из ОД считалась несогласованной.

При вычислении метрики M2 атрибут *authors* определён как необязательный, так как при регистрации ПрЭВМ, БД, ТИМС и ПО авторы могут отказаться быть упомянутыми в качестве таковых. А в случае ИЗ и ПМ авторы могут отказаться быть упомянутыми при официальной публикации сведений о выдаче патента.

Документы, опубликованные в ОР на сайте ФИПС, не предназначены для массовой загрузки. Поэтому проверка соответствия значений и вычисление метрики M3 проводились с использованием простых случайных выборок [**27**, **32**]. Для этого из каждого набора ОД было извлечено 800 записей. При этом значения атрибутов, для которых соответствующая характеристика в связанном документе из ОР ФИПС отсутствовала, игнорировались.

Перед сопоставлением значений атрибутов записей с характеристиками в ОР на сайте ФИПС была выполнена очистка данных: приведение символов к единому регистру, значений дат к единому формату, удаление знаков препинания и повторяющихся пробелов, замена буквы ё на е и др.

При вычислении метрик М4, М5, М6, М8 просматривались записи, в которых хотя бы один из атрибутов, исследуемых в метрике, содержит непустую строку.

При вычислении метрики М8 фиктивные значения (см. требование 5) не просматривались. Исключение атрибута *app_number* для наборов ИЗ, ПМ и ПО при вычислении метрики М8 связано с тем, что патенты могут быть аннулированы в связи с признанием их недействительными частично. В такой ситуации выдаётся новый патент, при этом номер заявки остаётся прежним.

При вычислении метрики М9 просматривались записи, в которых оба исследуемых атрибута содержат непустые строки.

Вычисление метрики М10 проводилось на основе сведений о фактически опубликованных версиях наборов ОД и с учётом того, что Роспатентом заявлена ежемесячная периодичность обновления данных.

Для вычисления метрики М11 были использованы результаты собственных наблюдений, проведённых в 2022 г., и данные сервиса https://web.archive.org, в котором сохранены архивные копии вебстраниц ОД Роспатента за 2020–2021 гг.

Результаты

Полнота

Ни в одном наборе ОД не представлено всех записей из соответствующих ОР ФИПС (табл. 7, метрика М1). Патенты, сведения о которых отсутствуют в наборах ОД, выданы не позже 2010 г. Пропущенные сведения о ПрЭВМ, БД, ТИМС относятся к ОИС, зарегистрированным до 2015 г.

У каждой записи набора ТИМС указаны все обязательные атрибуты, у остальных наборов имеются пропуски (табл. 7, метрика М2).

Таблица 7

Результаты измерения полноты

Метрика	ИЗ	ПМ	ПО	ПрЭВМ	БД	ТИМС
М1	0,335	0,984	0,948	0,856	0,856	0,923
М2	0,996	0,994	0,999	0,999	0,999	1,0

Точность

Наиболее точными в сравнении с ОР ФИПС оказались наборы ПрЭВМ, БД, ТИМС, наименее – ИЗ и ПМ (табл. 8, метрика М3).

Фиктивные значения (табл. 8, метрика М5) и несогласованные даты подачи заявки и регистрации (табл. 8, метрика М4) обнаружены в ОИС, зарегистрированных до 2013 г.

Наименьшую точность с точки зрения соответствия форматам показал набор ИЗ (табл. 8, метрика М6), наибольшую – набор ПО.

Таблица 8

Результаты измерения точности

Метрика	ИЗ	ПМ	ПО	ПрЭВМ	БД	ТИМС
М3	0,774	0,764	0,839	0,995	0,997	0,981
М4	0,999	1,0	0,999	0,999	1,0	1,0
М5	0,999	1,0	1,0	0,999	0,999	1,0
М6	0,716	0,879	0,978	0,945	0,938	0,955

Наибольшая доля значений, не соответствующих характеристикам, указанным в документах ОР на сайте ФИПС (метрика М3), или форматам (метрика М6), обнаружена в атрибутах *authors* и *holders*.

Несоответствие значений атрибутов *authors*, *holders* и *name* характеристикам в ОР на сайте ФИПС связано с графическими расхождениями: сокращениями, визуально неоднозначными символами, неполной или, наоборот, избыточной информацией.

Случаи семантического расхождения в атрибутах *authors* и *holders* выявлены в сведениях о патентах, выданных не позже 1998 г., и ТИМС, зарегистрированной в 2020 г. Кроме того, в наборах ИЗ, ПМ и ПО обнаружены записи со статусом правовой охраны, не соответствующим действительности.

Нарушение форматов значений атрибутов *authors* и *holders* связано с присутствием символов латиницы, отсутствием кода страны или, наоборот, его дублированием, использованием другого разделителя имён. В отбракованных значениях атрибута *reg_number* обнаружен маркер последовательности байтов (англ. Byte Order Mark, BOM), а в атрибуте *app_number* набора ИЗ присутствуют лишние символы (например, 2011102118/02).

Согласованность

Наилучшая согласованность с индексом ОР ФИПС наблюдается в наборах ПО и ТИМС (табл. 9, метрика М7). Повторяющиеся значения обнаружены в наборах ПрЭВМ и БД (табл. 9, метрика М8). Значения атрибутов согласованы у всех записей в наборах ТИМС и БД (табл. 9, метрика М9).

Таблица 9

Результаты измерения согласованности

Метрика	ИЗ	ПМ	ПО	ПрЭВМ	БД	ТИМС
М7	0,998	0,999	1,0	0,996	0,999	1,0
М8	–	–	–	0,999	0,999	1,0
М9	0,999	0,999	0,999	0,999	1,0	1,0

ПрЭВМ и БД, отсутствующие в ОР на сайте ФИПС, зарегистрированы до 2013 г., а ИЗ и ПМ – до 2021 г. включительно.

Повторяющиеся значения в атрибуте *app_number* обнаружены у ПрЭВМ и БД, зарегистрированных, преимущественно, не позже 2012 г. У ОИС, зарегистрированных в 2021 г., такие значения связаны с «ошибочной» регистрацией, после которой произведена повторная «успешная» регистрация ОИС.

Дата подачи заявки превышает дату регистрации в ОИС, зарегистрированных не позже 2012 г.

Своевременность

Не все ожидаемые версии наборов ОД опубликованы, у большинства опубликованных версий дата актуальности не совпадает с датой публикации (табл. 10).

Таблица 10

Результаты измерения своевременности

Метрика	ИЗ	ПМ	ПО	ПрЭВМ	БД	ТИМС
М10	0,73					
М11	0,27					

Дата актуальности всех версий приходится на первое число месяца. Следовательно, 73% версий ОД Роспатента опубликованы с задержкой. При этом максимальное превышение в публикации очередной версии составляет 25 дней.

Выводы

Исследование показало, что в схемах наборов ОД Роспатента присутствуют все обязательные для интеграции с системами CRIS атрибуты. Однако наборы ОД Роспатента не подойдут для оперативного или непрерывного получения актуальной информации об ОИС, так как публикация новых версий происходит с задержкой. В части записей наборов наблюдаются неточность и несогласованность, присутствует информация не обо всех ОИС, опубликованных в ОР на сайте ФИПС.

В большинстве случаев анализ результатов научной деятельности ограничивается пятилетней ретроспективой. И, как показало исследование, в записях об ОИС, зарегистрированных в период 2017–2021 гг., проблем с качеством в критически значимых атрибутах (даты регистрации и подачи заявки, регистрационный номер ОИС и номер заявки) нет. Исключением являются сведения о статусе действия правовой охраны для патентов: данную информацию перед формированием отчётов в CRIS необходимо проверять дополнительно.

Таким образом, перед загрузкой записей ОД Роспатента в CRIS нам представляются необходимыми предварительное измерение их качества и последующая очистка. Так, если наблюдается несогласованность в значениях атрибутов или пропуски, то нужно или дополнительно проверить характеристики ОИС (например, в ОР на сайте ФИПС), или не загружать данную запись в CRIS. Для связывания сведений об авторах и патентообладателях (правообладателях) с сотрудниками и организациями, зарегистрированными в CRIS, необходимо учитывать обнаруженные в настоящем исследовании проблемы несоответствия форматов.

Список источников

1. **Интеллектуальная** Система Тематического Исследования Наукометрических данных. URL: https://istina.msu.ru (дата обращения: 18.03.2022).

2. **SciAct** – информационно-аналитическая система мониторинга и учёта научной деятельности. URL: https://sciact.ru (дата обращения: 18.03.2022).

3. **Информационно-аналитическая** система «Результаты научной деятельности». URL: https://science.asu.edu.ru (дата обращения: 18.03.2022).

4. **Azeroual O., Saake G., Abuosba M., Schöpfel J.** Quality of Research Information in RIS Databases: A Multidimensional Approach // Business Information Systems. BIS 2019. Lecture Notes in Business Information Processing. 2019. Vol. 353. P. 337–349. URL: https://doi.org/10.1007/978-3-030-20485-3_26.

5. **Azeroual O., Schöpfel J.** Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries // Publications. 2019. URL: https://doi.org/10.3390/publications7010014 (дата обращения: 02.11.2022).

6. **ГОСТ Р ИСО 8000-2-2019.** Качество данных. Часть 2. Словарь. Москва : Стандартинформ, 2019. 12 с.

7. **Васенин В. А., Афонин С. А., Зензинов А. А., Лунев К. В., Шачнев Д. А.** Механизмы системы «ИСТИНА» для интеллектуального анализа состояния и стимулирования хода выполнения проектов в сфере науки и высшего образования // Научный сервис в сети Интернет. 2019. № 21. С. 210–221. URL: http://doi.org/10.20948/abrau-2019-48.

8. **Открытые** данные Роспатента. URL: https://rospatent.gov.ru/opendata (дата обращения: 18.03.2022).

9. **Открытые** реестры. URL: https://new.fips.ru/registers-web/ (дата обращения: 18.03.2022).

10. **Условия** использования открытых данных Роспатента / Открытая лицензия. URL: https://rospatent.gov.ru/content/uploadfiles/opendata-terms-of-use.docx (дата обращения: 18.03.2022).

11. **Чесноков М. Ю.** Поиск аномалий в задаче повышения качества открытых данных // Проблемы управления. 2019. № 3. С. 53–62. URL: https://doi.org/10.25728/pu.2019.3.6.

12. **Sadiq S., Indulska M.** Open data: Quality over quantity // International Journal of Information Management. 2017. Vol. 37 (3). P. 150–154. URL: https://doi.org/10.1016/j.ijinfomgt.2017.01.003.

13. **Torchiano M., Vetrò A., Iuliano F.** Preserving the benefits of Open Government Data by measuring and improving their quality: an empirical study // 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). 2017. Vol. 1. P. 144–153. URL: https://doi.org/10.1109/COMPSAC.2017.192.

14. **Vetrò A., Canova L., Torchiano M., Minotas C. O., Iemma R., Morando F.** Open data quality measurement framework: Definition and application to Open Government Data // Government Information Quarterly. 2016. Vol. 33 (2). P. 325–337. URL: http://doi.org/10.1016/j.giq.2016.02.001.

15. **Rula A., Maurino A., Batini C.** Data quality issues in linked open data // Data and information quality. 2016. P. 87–112. URL: https://doi.org/10.1007/978-3-319-24106-7_4.

16. **Открытость** государства в России – 2021.
URL: https://ach.gov.ru/upload/pdf/Otkrytost-2021.pdf (дата обращения: 18.03.2022).

17. **Российский** сервер Espacenet. URL: https://ru.espacenet.com (дата обращения: 18.03.2022).

18. **Поисковая** система Designview. URL: https://www.tmdn.org/tmdsview-web/welcome#/dsview (дата обращения: 18.03.2022).

19. **Поисковая** система Google Patents. URL: https://patents.google.com (дата обращения: 18.03.2022).

20. **Яндекс.Патенты** – поиск по патентным документам. URL: https://yandex.ru/patents (дата обращения: 18.03.2022).

21. **eLIBRARY.RU.** Поиск патентов. URL: https://elibrary.ru/patents.asp (дата обращения: 18.03.2022).

22. **Официальные** публикации ФИПС. URL: https://www.fips.ru/publication-web/ (дата обращения: 18.03.2022).

23. **Информационно-поисковая** система ФИПС. URL: https://www.fips.ru/iiss/ (дата обращения: 18.03.2022).

24. **Batini C.** Data Quality Assessment // Encyclopedia of Database Systems. Boston: Springer, 2009. P. 608–612. URL: https://doi.org/10.1007/978-0-387-39940-9_107.

25. **DAMA-DMBOK:** Свод знаний по управлению данными. Второе издание / Dama International [пер. с англ. Г. Агафонова]. Москва : Олимп-Бизнес, 2020. 828 с.

26. **Mahanti R.** Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. Quality Press, 2019. 526 p.

27. **Lee Y. W., Pipino L. L., Funk J. D., Wang R. Y.** Journey to data quality. The MIT Press, 2006. 240 p.

28. **Sattler Ku.** Data Quality Dimensions // Encyclopedia of Database Systems. Boston: Springer, 2009. P. 612–615. URL: https://doi.org/10.1007/978-0-387-39940-9.

29. **Gualo F., Rodriguez M., Verdugo J., Caballero I., Piattini M.** Data quality certification using ISO/IEC 25012: Industrial experiences // Journal of Systems and Software. 2021. Vol. 176. P. 110938. URL: https://doi.org/10.1016/j.jss.2021.110938.

30. **Zhao Y., Gong J., Hu Y., Liu Z., Cai L.** Analysis of quality evaluation based on ISO/IEC SQuaRE series standards and its considerations // 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). 2017. P. 245–250. URL: https://doi.org/10.1109/ICIS.2017.7960001.

31. **Behkamal B., Kahani M., Bagheri E., Jeremic Z.** A metrics-driven approach for quality assessment of linked open data // Journal of theoretical and applied electronic commerce research. 2014. Vol. 9 (2). P. 64–79. URL: https://doi.org/10.4067/S0718-18762014000200006.

32. **Liu H., Sang Z., Karali S.** Approximate quality assessment with sampling approaches // 2019 International Conference on Computational Science and Computational Intelligence (CSCI). 2019. P. 1306–1311. URL: https://doi.org/10.1109/CSCI49370.2019.00244.

References

1. **Intellektual`naia** Sistema Tematicheskogo Issledovaniia Naukometricheskikh danny`kh. URL: https://istina.msu.ru (data obrashcheniia: 18.03.2022).

2. **SciAct** – informatcionno-analiticheskaia sistema monitoringa i uchyota nauchnoi` deiatel`nosti. URL: https://sciact.ru (data obrashcheniia: 18.03.2022).

3. **Informatcionno-analiticheskaia** sistema «Rezul`taty` nauchnoi` deiatel`nosti». URL: https://science.asu.edu.ru (data obrashcheniia: 18.03.2022).

4. **Azeroual O., Saake G., Abuosba M., Schöpfel J.** Quality of Research Information in RIS Databases: A Multidimensional Approach // Business Information Systems. BIS 2019. Lecture Notes in Business Information Processing. 2019. Vol. 353. P. 337–349. URL: https://doi.org/10.1007/978-3-030-20485-3_26.

5. **Azeroual O., Schöpfel J.** Quality Issues of CRIS Data: An Exploratory Investigation with Universities from Twelve Countries // Publications. 2019. URL: https://doi.org/10.3390/publications7010014 (data obrashcheniia: 02.11.2022).

6. **GOST R ISO 8000-2-2019.** Kachestvo danny`kh. Chast` 2. Slovar`. Moskva : Standartinform, 2019. 12 c.

7. **Vasenin V. A., Afonin S. A., Zenzinov A. A., Lunev K. V., Shachnev D. A.** Mehanizmy` sistemy` «ISTINA» dlia intellektual`nogo analiza sostoianiia i stimulirovaniia hoda vy`polneniia proektov v sfere nauki i vy`sshego obrazovaniia // Nauchny`i` servis v seti Internet. 2019. № 21. S. 210–221. URL: http://doi.org/10.20948/abrau-2019-48.

8. **Otkry`ty`e** danny`e Rospatenta. URL: https://rospatent.gov.ru/opendata (data obrashcheniia: 18.03.2022).

9. **Otkry`ty`e reestry`.** URL: https://new.fips.ru/registers-web/ (data obrashcheniia: 18.03.2022).

10. **Usloviia** ispol`zovaniia otkry`ty`kh danny`kh Rospatenta / Otkry`taia licenziia. URL: https://rospatent.gov.ru/content/uploadfiles/opendata-terms-of-use.docx (data obrashcheniia: 18.03.2022).

11. **Chesnokov M. Iu.** Poisk anomalii` v zadache povy`sheniia kachestva otkry`ty`kh danny`kh // Problemy` upravleniia. 2019. № 3. S. 53–62. URL: https://doi.org/10.25728/pu.2019.3.6.

12. **Sadiq S., Indulska M.** Open data: Quality over quantity // International Journal of Information Management. 2017. Vol. 37 (3). P. 150–154. URL: https://doi.org/10.1016/j.ijinfomgt.2017.01.003.

13. **Torchiano M., Vetrò A., Iuliano F.** Preserving the benefits of Open Government Data by measuring and improving their quality: an empirical study // 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). 2017. Vol. 1. P. 144–153. URL: https://doi.org/10.1109/COMPSAC.2017.192.

14. **Vetrò A., Canova L., Torchiano M., Minotas C. O., Iemma R., Morando F.** Open data quality measurement framework: Definition and application to Open Government Data // Government Information Quarterly. 2016. Vol. 33 (2). P. 325–337. URL: http://doi.org/10.1016/j.giq.2016.02.001.

15. **Rula A., Maurino A., Batini C.** Data quality issues in linked open data // Data and information quality. 2016. P. 87–112. URL: https://doi.org/10.1007/978-3-319-24106-7_4.

16. **Otkry`tost`** gosudarstva v Rossii – 2021. URL: https://ach.gov.ru/upload/pdf/Otkrytost-2021.pdf (data obrashcheniia: 18.03.2022).

17. **Rossii`skii`** server Espacenet. URL: https://ru.espacenet.com (data obrashcheniia: 18.03.2022).

18. **Poiskovaia** sistema Designview. URL: https://www.tmdn.org/tmdsview-web/welcome#/dsview (data obrashcheniia: 18.03.2022).

19. **Poiskovaia** sistema Google Patents. URL: https://patents.google.com (data obrashcheniia: 18.03.2022).

20. **Yandex.Patenty`** – poisk po patentny`m dokumentam. URL: https://yandex.ru/patents (data obrashcheniia: 18.03.2022).

21. **eLIBRARY.RU.** Poisk patentov. URL: https://elibrary.ru/patents.asp (data obrashcheniia: 18.03.2022).

22. **Ofitcial`ny`e** publikatcii FIPS. URL: https://www.fips.ru/publication-web/ (data obrashcheniia: 18.03.2022).

23. **Informatcionno-poiskovaia** sistema FIPS. URL: https://www.fips.ru/iiss/ (data obrashcheniia: 18.03.2022).

24. **Batini C.** Data Quality Assessment // Encyclopedia of Database Systems. Boston: Springer, 2009. P. 608–612. URL: https://doi.org/10.1007/978-0-387-39940-9_107.

25. **DAMA-DMBOK:** Svod znanii` po upravleniiu danny`mi. Vtoroe izdanie / Dama International [per. s angl. G. Agafonova]. Moskva : Olimp-Biznes, 2020. 828 s.

26. **Mahanti R.** Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. Quality Press, 2019. 526 p.

27. **Lee Y. W., Pipino L. L., Funk J. D., Wang R. Y.** Journey to data quality. The MIT Press, 2006. 240 p.

28. **Sattler Ku.** Data Quality Dimensions // Encyclopedia of Database Systems. Boston: Springer, 2009. P. 612–615. URL: https://doi.org/10.1007/978-0-387-39940-9.

29. **Gualo F., Rodriguez M., Verdugo J., Caballero I., Piattini M.** Data quality certification using ISO/IEC 25012: Industrial experiences // Journal of Systems and Software. 2021. Vol. 176. P. 110938. URL: https://doi.org/10.1016/j.jss.2021.110938.

30. **Zhao Y., Gong J., Hu Y., Liu Z., Cai L.** Analysis of quality evaluation based on ISO/IEC SQuaRE series standards and its considerations // 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). 2017. P. 245–250. URL: https://doi.org/10.1109/ICIS.2017.7960001.

31. **Behkamal B., Kahani M., Bagheri E., Jeremic Z.** A metrics-driven approach for quality assessment of linked open data // Journal of theoretical and applied electronic commerce research. 2014. Vol. 9 (2). P. 64–79. URL: https://doi.org/10.4067/S0718-18762014000200006.

32. **Liu H., Sang Z., Karali S.** Approximate quality assessment with sampling approaches // 2019 International Conference on Computational Science and Computational Intelligence (CSCI). 2019. P. 1306–1311. URL: https://doi.org/10.1109/CSCI49370.2019.00244.

Информация об авторе / Information about the author

Зелепухина Виктория Андреевна – канд. техн. наук, старший научный сотрудник Астраханского государственного университета им. В. Н. Татищева, Астрахань, Российская Федерация

v.zelepukhina@asu.edu.ru

Victoria A. Zelepukhina – Cand. Sc. (Engineering), Senior Researcher, V. N. Tatishchev Astrakhan State University, Astrakhan, Russian Federation

v.zelepukhina@asu.edu.ru