

ПАСПОРТ НАБОРОВ ДАННЫХ И РЕЗУЛЬТАТОВ ИССЛЕДОВАНИЙ ДЛЯ ПУБЛИКАЦИИ В ОТКРЫТЫХ ИСТОЧНИКАХ

Борисов Р. С.¹, Ефименко А. А.²

Ключевые слова: открытые данные, структура набора данных, метаданные, паспорт набора данных, паспорт результата исследований, публикация наборов данных, конфиденциальность, классификационные признаки.

Аннотация.

Цель работы: разработка программно-функциональных моделей паспорта набора данных и результата исследования, содержащих классификационные признаки, позволяющие полностью или по частям относить наборы данных (результаты исследований) к определенным классам, а также методических рекомендаций по заполнению паспорта набора данных (результата исследования).

Методы: системное моделирование программной структуры и организационных приемов обработки паспортов наборов данных и результатов исследования, позволяющих обеспечить конфиденциальность информации при публикации в открытых источниках.

Результаты: определены требования к представлению паспортов наборов данных и результатов исследований, описаны процесс формирования и методические рекомендации по заполнению паспорта набора данных (результата исследования). Приведен формат разделов паспорта набора данных и паспорта результата исследования, которые обязательны к заполнению при представлении набора данных (результата исследования) для возможности их опубликования в открытом доступе.

DOI: 10.21681/1994-1404-2022-2-66-79

Введение

Соблюдение рационального баланса между возможностями открытого распространения привилегированной содержательной информации и законодательных ограничений на доступ к информации о деятельности государственных и муниципальных органов власти, предприятий промышленности, транспорта, связи, научных, образовательных, медицинских учреждений, общественных и других организаций является актуальной *организационно-технической проблемой*. В настоящее время решаются задачи разработки протоколов обработки наборов данных (результатов исследования) для их публикации в открытых источниках [1], классификаторов правовых актов для установления правового режима публикуемой информации [2], проводится постоянный мониторинг и анализ актуального федерального законодательства об ограничении публикации данных в открытых информационных ресурсах [3, 4].

Одним из ключевых элементов организационных протоколов и классификаторов наборов данных (результатов исследования) является использование унифицированного идентификатора или *паспорта набора данных* (ПНД) в форме, пригодной для обработки в автоматизированных системах [1, 12]. Наличие адекватной программно-функциональной модели ПНД (паспорта набора данных или результата исследования), содержащего классификационные признаки, позволяющие полностью или по частям относить наборы данных (результаты исследований) к определенным классам, а также обоснованных методических рекомендаций по заполнению ПНД позволит разработать программную реализацию паспортов в составе автоматизированной системы обработки открытых данных [6].

Паспорт набора данных (результата исследования) — совокупность свойств набора данных (результата исследования), позволяющая однозначно идентифицировать такой набор данных (результат исследования) и получить ключевые параметры, его характеризующие. ПНД как элементы информационного обеспечения *эргасистемы* [7] должны обеспечивать

¹ **Борисов Роман Сергеевич**, кандидат технических наук, доцент кафедры информационного права, информатики и математики Российского государственного университета правосудия, г. Москва, Российская Федерация.
E-mail: bestseller@bk.ru

² **Ефименко Алексей Анатольевич**, кандидат технических наук, доцент кафедры информационного права, информатики и математики Российского государственного университета правосудия, г. Москва, Российская Федерация.
E-mail: alex192@mail.ru

формализацию объективной и субъективной частей информационного ресурса *эргасистемы* с целью его рационального употребления и расходования для более полного и объективного обеспечения *информационной безопасности* в условиях *информационного соперничества* [8]. Каждый ПНД, полученный от Обладателя (источника) набора данных (результата исследования) в соответствии с установленным *протоколом обработки* [1], должен включать следующие разделы (поля):

- поле, в котором отмечено, являются ли данные этого набора данных (результата исследования) конфиденциальной информацией (персональные данные, коммерческая, служебная и профессиональная тайна и др.; одновременно можно выбрать несколько категорий);
- поле, в котором отмечена возможность опубликовать данные и, если определена возможность опубликования, то в каком виде публикуется: разрешена публикация в передаваемом виде; разрешена публикация в обезличенном (для персональных данных) или деперсонифицированном (для коммерческой тайны) виде; разрешена публикация в анонимизированном виде; запрещена публикация.

При формировании как паспортов набора данных (результата исследования), так и самих наборов данных (результатов исследований) следует учитывать правила:

- для каждого набора данных (результата исследования) может быть установлена периодичность обновления с учетом потребностей пользователей наборов данных;
- для каждого набора данных (результата исследования) в ПНД может быть указана дата его актуальности.

Периодичность (частота) обновления набора данных (результата исследования) зависит от существования данных.

Для *оперативных* данных: поступающих в интерактивном режиме сведений (например, о деятельности компаний, о текущей отчетности, о числе возгораний и природных пожаров в России (от МЧС России), о предварительных итогах голосования (от ЦИК России) и др.) — периодичность обновления может быть установлена следующими значениями: *чаще одного раза в день; ежедневно; еженедельно.*

Для *долговременных* данных (управленческие регламенты, например, должностные и рабочие инструкции, правила поведения персонала, законы, нормы, нормативы, стандарты, технологические карты производственных процессов; охраняемые результаты интеллектуальной деятельности [9] и др.) частота обновления может быть установлена следующими значениями: *ежемесячно; ежеквартально; каждые полгода; ежегодно; по мере изменения данных.*

Обладатель набора данных (результата исследования) может указать срок хранения набора данных (результата исследования), но для полноты предоставле-

ния услуг пользователям результата исследования необходимо стремиться обеспечить хранение всех версий наборов данных. ПНД формируются в электронном виде при передаче (актуализации) Обладателем набора данных (результата исследования) и подписываются ответственным лицом.

Структура ПНД включает поля, характеризующие ключевые параметры набора данных (результата исследования), сгруппированные в объекты ПНД, представляющие собой *информационную модель* набора данных (результата исследования) в виде структурированного графа. Для обеспечения совместимости ПНД с информационными системами и индексирования в поисковых системах разработку структуры объектов ПНД необходимо проводить в соответствии с базовыми международными и отечественными стандартами представления данных и метаданных.

Описание применения базовых стандартов к представлению паспортов наборов данных и результатов исследований

Требования к представлению ПНД базируются на следующих международных и национальных *стандартах*:

RDF (Resource Description Framework) — разработанная международным консорциумом сети Интернет (*World Wide Web Consortium — W3C*) модель для представления данных, в особенности — метаданных, пригодных для машинной обработки;

RDFa (Resource Description Framework in Attributes) — модель для представления данных в формате *RDF* через атрибуты программных сред *HTML (Hyper Text Markup Language — «язык разметки»), XHTML (extensible hypertext markup language — расширяемый язык гипертекстовой разметки) и XML (eXtensible Markup Language)*;

W3C Recommendation: RDFa Core 1.1 — Рекомендация от *W3C* по использованию *RDFa*³;

RFC2413 — международный стандарт использования метаданных *“Encoding Dublin Core Metadata in HTML”*; *ГОСТ Р 7.0.10-2019* — российский стандарт «Набор элементов метаданных «Дублинское ядро»;

W3C Recommendation: XML Schema Part 2: Datatypes — Рекомендация от *W3C* по использованию типов данных⁴.

Для представления ПНД используются следующие **словари *RDFa***:

foaf (Friend of a Friend) — <http://xmlns.com/foaf/0.1/> — для предоставления класса объекта;

dc (Dublin Core) — <http://purl.org/dc/terms/> — для предоставления свойств объекта (используется редакция *“Qualified Dublin Core”*, так как содержит расширенный перечень свойств).

³ *RDFa 1.1 Primer (3rd ed.) // W3C. Syntax and processing rules for embedding RDF through attributes 17 March 2015, Retrieved 2016-09-02.*

⁴ *XML Schema Part 2: Datatypes Second Edition // W3C Recommendation 28 October 2004.*

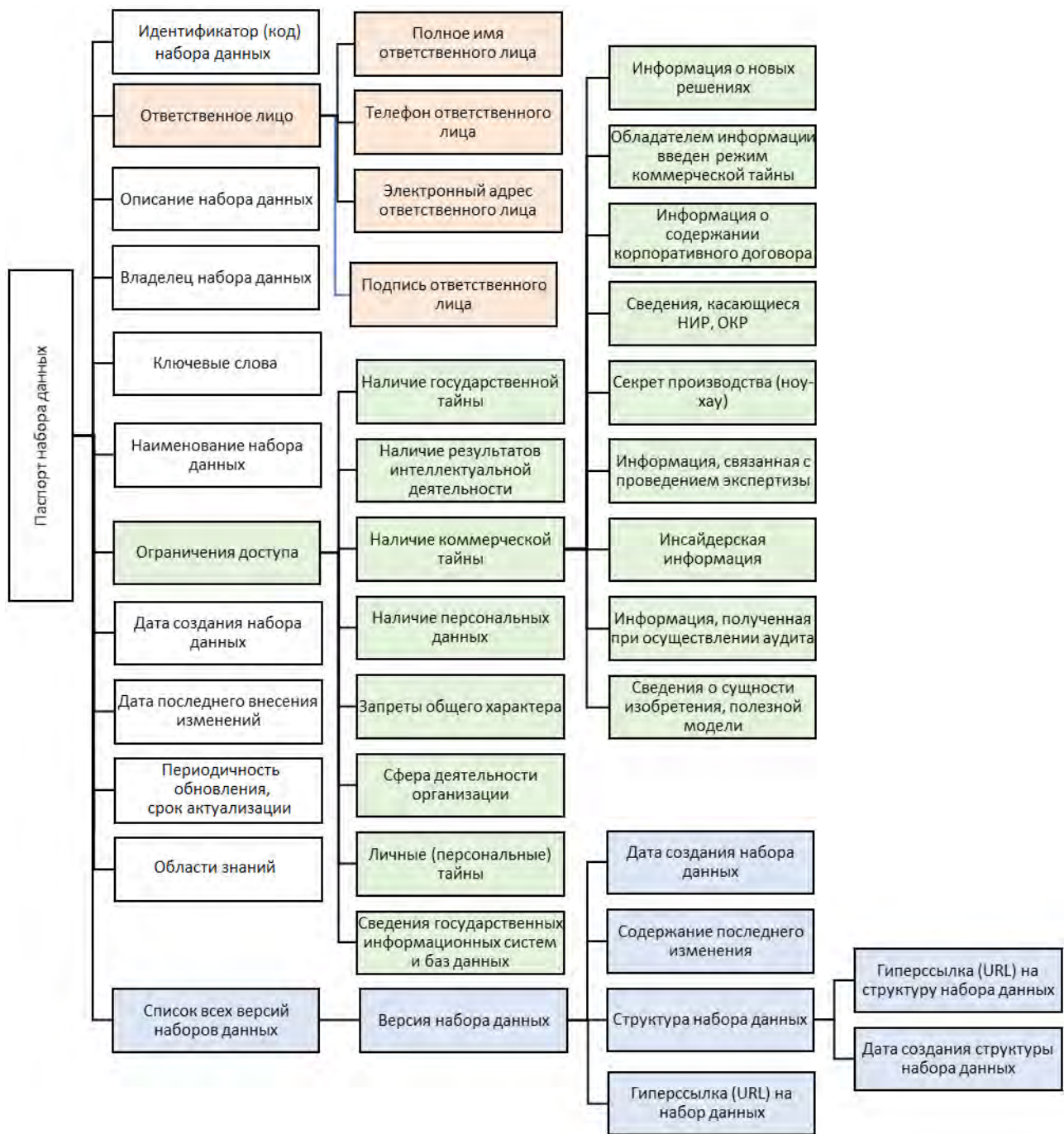


Рис. 1. Паспорт набора данных с иерархией RDFa объектов

RDFa предоставляет средства для построения информационных моделей наборов данных. Взятый в отдельности граф RDFa можно понимать как паспорт набора данных (результата исследования). Толкование значения основывается на способности пользователей RDF интерпретировать отдельные URI, строковые литералы и структуру графа и представлять остальные URI и семантику данных.

При использовании общепринятых стандартов, систем и справочников могут применяться средства связывания атрибутов наборов данных/результатов исследований между собой [5], которые представляют терминологический словарь синонимов/

соотношений в структуре набора открытых данных (например, для формата xsd — следующей структуры: `<xs: documentation xml: lang=»ru»>Дата первой публикации набора данных в формате ISO 8601.</xs:documentation>`). При использовании собственных справочников (например, ОКАТО или другого специализированного справочника) необходимо указывать ссылку на справочник или на средство, позволяющее связать атрибуты справочника в структуре набора данных/результата исследования. Используемые общепринятые стандарты, системы и справочники не должны иметь никаких ограничений в использовании.

Поля паспорта набора данных

№ п/п	Свойство	RDFa свойство	Формат значения	Значение
1	Набор данных	hasPart из словаря dc	xsd:anyURI	Ссылка на объект RDFa ПНД
2	Предыдущая страница	prev	xsd:anyURI	Ссылка на объект RDFa ПНД предыдущей страницы
3	Следующая страница	next	xsd:anyURI	Ссылка на объект RDFa ПНД следующей страницы

Описание структуры объектов паспортов наборов данных и результатов исследований

Свойства ПНД в терминологии *RDFa* представляют собой объекты из словарей *DCMI (Dublin Core Metadata Initiative)*: *dc*, *foaf*, а также пользовательских объектов. Структура ПНД описывается в терминологии объектов *xsd*-схемы с целью обеспечения совместимости паспортов с существующими информационными платформами и индексацией в глобальной сети Интернет. Требования к структурам метаданных, которыми, по сути, является ПНД, определены международными и российскими стандартами. Функциональное описание элементов паспорта рассмотрим в настоящем разделе.

Описание классификационных признаков ПНД представляется в терминологии языков *XML* и *XSD* и в соответствии с рекомендациями *W3C* и ГОСТ Р 7.0.10-2019. При этом используются следующие форматы данных:

xsd:string представляет собой символьную строку, включая символ перевода каретки; параметры строки — длина, нумерация, пробел и др. определяются атрибутами типа данных (*length*, *pattern*, *maxLength*, *minLength*, *enumeration*, *whiteSpace*);

xsd:dateTime представляет собой конкретное время и дату; шаблон для *dateTime* следующий — *CCYY-MM-DDThh:mm:ss*, где *CC* представляет столетие; *YY* — год; *MM* — месяц; *DD* — день; *T* — разделитель даты и времени; *hh* — число часов; *mm* — минут; *ss* — секунд (при необходимости можно указывать доли секунды, например, сотые доли в шаблоне: *ss.ss*);

xsd:anyURI представляет собой имя и адрес объекта (группы свойств паспорта набора данных) в сети; *URI* — *Uniform Resource Identifier* (унифицированный идентификатор ресурса) как определено в *RFC 2396* (значение *anyURI* может быть абсолютно или относительно и может иметь необязательный идентификатор фрагмента);

xsd:anyURL представляет собой адрес ресурса (версии набора данных) в сети, определяет местонахождение и способ обращения к нему; *URL* — *Uniform Resource Locator* (унифицированный определитель местонахождения ресурса); *URL* включает в себя протокол, который используется для доступа к ресурсу — *http*, *https*, *ftp*; расположение сервера с использованием *IP*-адреса или имени домена; номер порта на сервере; точное местоположение в структуре каталогов сервера.

Разбиение ПНД на объекты соответствует международным стандартам *W3C* и требованиям отечественно-

го ГОСТ Р 7.0.10-2019, что позволяет представить ПНД в форме, пригодной для машинной обработки и для обмена данными (выполнить так называемую *сериализацию*).

На *рис. 1* представлен паспорт набора данных в виде структурированного графа с иерархией *RDFa* объектов. Так, при выборе поля «Ответственное лицо» раскрываются для заполнения поля следующего уровня иерархии: «Полное имя ответственного лица», «Телефон ответственного лица» и др.

ПНД должен содержать объект *RDFa*, к которому предъявляются следующие обязательные требования:

- объект *RDFa* для обеспечения уникальности атрибутов должен быть элементом текущей страницы паспорта;
- объект должен иметь класс «*Collection*» из словаря «*dc*».

Соблюдение этих требований позволяет сопоставить данный объект *RDFa* с текущей страницей ПНД и определяет данный объект *RDFa*. Объект *RDFa* ПНД содержит значения, представленные в *табл. 1*.

Свойство «Набор данных» объекта *RDFa* ПНД повторяется столько раз, сколько существует передаваемых наборов данных.

Свойства «Предыдущая страница» и «Следующая страница» объекта *RDFa* ПНД указываются обязательно, если данные значения существуют.

Каждая страница ПНД должна содержать объект *RDFa*, к которому предъявляются следующие обязательные требования:

- объект должен быть субъектом текущей страницы паспорта;
- объект должен иметь класс *Document* из словаря *foaf*.

Соблюдение этих требований однозначно сопоставляет данный объект *RDFa* с текущей страницей паспорта набора данных.

Описание классификационных признаков, позволяющих полностью или по частям отнести набор данных к определенным классам

Описание классификационных признаков в ПНД представлено в терминологии языков *XML* и *XSD* и в соответствии с рекомендациями *W3C* и ГОСТ Р 7.0.10-2019. Классификационные признаки в паспортах наборов данных содержат значения, представленные в *табл. 2*.

Классификационные признаки в паспортах наборов данных

№ п/п	Свойство	RDFa свойство	Формат значения	Значение
1.	Идентификатор (код) набора данных	<i>identifier</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение идентификатора (кода) набора данных
2.	Наименование набора данных	<i>title</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение наименования набора данных
3.	Описание набора данных	<i>description</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение описания набора данных
4.	Обладатель набора данных	<i>creator</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение обладателя набора данных
5.	Ответственное лицо	<i>publisher</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект <i>RDFa</i> ответственного лица набора данных
6.	Ограничения доступа	<i>rights</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект <i>RDFa</i> коллекции режимов доступа набора данных
7.	Дата создания набора данных	<i>created</i> из словаря <i>dc</i>	<i>xsd:dateTime</i>	Пользовательское значение даты создания набора данных
8.	Дата последнего внесения изменений	<i>modified</i> из словаря <i>dc</i>	<i>xsd:dateTime</i>	Пользовательское значение даты последнего изменения набора данных
9.	Периодичность обновления, срок актуализации набора данных	<i>valid</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение периодичности обновления, срока актуализации набора данных
10.	Ключевые слова, соответствующие содержанию набора данных	<i>subject</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение ключевых слов набора данных
11.	Область знаний	<i>description</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение областей знаний, к которым относится набор данных
12.	Список всех версий наборов данных	<i>source</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект <i>RDFa</i> коллекции всех версий наборов данных

ПНД может содержать и другие *RDFa*-свойства из словаря *dc*, содержащие классификационные признаки, позволяющие полностью или по частям отнести набор данных к определенным классам.

Свойство «Ответственное лицо» объекта *RDFa* ПНД ссылается на объект *RDFa* ответственного лица

набора данных. К объекту *RDFa* ответственного лица ПНД предъявляется следующее обязательное требование: объект должен иметь класс *Person* из словаря *foaf*.

Свойство «Ответственное лицо» набора данных содержит значения, представленные в табл. 3.

Таблица 3

Описание значений «Ответственное лицо»

№ п/п	Свойство	RDFa свойство	Формат значения	Значение
1.	Полное имя ответственного лица	<i>name</i> из словаря <i>foaf</i>	<i>xsd:string</i>	Пользовательское значение идентификационного номера (кода) набора данных
2.	Телефон ответственного лица	<i>phone</i> из словаря <i>foaf</i>	<i>xsd:string</i>	Пользовательское значение телефона ответственного лица
3.	Электронный адрес ответственного лица	<i>mbox</i> из словаря <i>foaf</i>	<i>xsd:string</i>	Пользовательское значение электронного адреса ответственного лица
4.	Подпись ответственного лица	<i>sign</i> из словаря <i>foaf</i>	<i>xsd:string</i>	Пользовательское значение подписи ответственного лица

Свойство «Коллекция (список) всех версий паспорта наборов данных» объекта *RDFa* ПНД ссылается на объект *RDFa* коллекции всех версий наборов данных. К объекту *RDFa* коллекции всех версий наборов данных

предъявляется следующее обязательное *требование*: объект должен иметь класс *Collection* из словаря *dc*. Коллекции всех версий наборов данных содержат следующие значения, представленные в табл. 4.

Таблица 4

Коллекции всех версий в паспорте наборов данных

Свойство	<i>RDFa</i> свойство	Формат значения	Значение
Версия набора данных	<i>hasPart</i> из словаря <i>dc</i>	<i>xsd:anyURL</i>	Ссылка на объект <i>RDFa</i> набора данных

Свойство «Версия набора данных» объекта *RDFa* коллекции всех версий в ПНД повторяется ровно столько раз, сколько существует публикуемых версий текущего набора открытых данных. Данное свойство ссылается на объект *RDFa* набора данных. К объ-

екту *RDFa* набора данных предъявляется следующее обязательное *требование*: объект должен иметь класс *Document* из словаря *foaf*. В табл. 5 представлены поля свойства «Версия набора данных» объекта *RDFa*.

Таблица 5

Версия набора данных

№ п/п	Свойство	<i>RDFa</i> свойство	Формат значения	Значение
1.	Гиперссылка (<i>URL</i>) на набор данных	<i>source</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на файл набора данных
2.	Дата создания набора данных	<i>created</i> из словаря <i>dc</i>	<i>xsd:dateTime</i>	Пользовательское значение даты создания набора данных
3.	Содержание последнего изменения	<i>provenance</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение содержания последнего изменения набора данных
4.	Структура набора данных	<i>conformsTo</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект <i>RDFa</i> структуры набора данных

Свойство «Структура набора данных» объекта *RDFa* ПНД ссылается на объект *RDFa* структуры набора данных. К объекту *RDFa* структуры набора данных предъявляется следующее обязательное *требование*: объект должен иметь класс *Document* из словаря *foaf*. Структура набора данных содержит значения, представленные в табл. 6.

Таблица 6

Структура набора данных

№ п/п	Свойство	<i>RDFa</i> свойство	Формат значения	Значение
1.	Гиперссылка (<i>URL</i>) на структуру набора данных	<i>source</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на файл структуры набора данных
2.	Дата создания структуры набора данных	<i>created</i> из словаря <i>dc</i>	<i>xsd:dateTime</i>	Пользовательское значение даты создания структуры набора данных

Свойство «Ограничения доступа» объекта *RDFa* ПНД ссылается на объект *RDFa* коллекции режимов доступа [10] набора данных. К объекту *RDFa* коллекции режимов доступа набора данных предъявляется следующее

обязательное *требование*: объект должен иметь класс *Document* из словаря *foaf*. В табл. 7. представлены коллекции режимов доступа набора данных.

Коллекции режимов доступа набора данных

№ п/п	Свойство	RDFa свойство	Формат значения	Значение
1.	Наличие государственной тайны	<i>state</i> из пользовательского словаря <i>security</i>	<i>xsd:string</i>	Пользовательское значение наличия государственной тайны
2.	Наличие результатов интеллектуальной деятельности	<i>intellectual</i> из пользовательского словаря <i>security</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa структуры Наличие результатов интеллектуальной деятельности
3.	Наличие коммерческой тайны	<i>commercial</i> из пользовательского словаря <i>security</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa структуры Наличие коммерческой тайны
4.	Наличие персональных данных	<i>personal</i> из пользовательского словаря <i>security</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa структуры Наличие персональных данных
5.	Запреты общего характера	<i>general</i> из пользовательского словаря <i>security</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa структуры Запреты общего характера
6.	Сфера деятельности организации	<i>field</i> из пользовательского словаря <i>security</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa структуры Сфера деятельности организации
7.	Личные (персональные) тайны	<i>secret</i> из пользовательского словаря <i>security</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa структуры Личные (персональные) тайны
8.	Сведения государственных информационных систем и баз данных	<i>systems</i> из пользовательского словаря <i>security</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa структуры Сведения государственных информационных систем и баз данных

Объекты RDFa свойств коллекции режимов доступа ПНД содержат квалификационные признаки ограничения доступа к набору данных в соответствии с классификатором наборов данных. Например, свойство «Наличие коммерческой тайны» объекта RDFa ПНД

ссылается на объект RDFa «Наличие коммерческой тайны». К объекту RDFa «Наличие коммерческой тайны» предъявляется следующее обязательное требование (табл. 8): объект должен иметь класс *Document* из словаря *foaf*.

Таблица 8

Наличие коммерческой тайны

№	Свойство	RDFa свойство	Формат значения	Значение
1.	Информация о новых решениях и технических знаниях, полученных сторонами по договору подряда	<i>technology</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Информация о новых решениях и технических знаниях, полученных сторонами по договору подряда
2.	Обладателем информации введен режим коммерческой тайны	<i>arbitr</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Обладателем информации введен режим коммерческой тайны
3.	Информация о содержании корпоративного договора, заключенного участниками непубличного общества	<i>contract</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Информация о содержании корпоративного договора, заключенного участниками непубличного общества

4.	Сведения, касающиеся предмета договоров на выполнение научно-исследовательских работ, опытно-конструкторских и технологических работ, хода их исполнения и полученных результатов, если иное не предусмотрено договорами	<i>design</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Сведения, касающиеся предмета договоров на выполнение научно-исследовательских работ, опытно-конструкторских и технологических работ, хода их исполнения и полученных результатов, если иное не предусмотрено договорами
5.	Секрет производства (ноу-хау)	<i>knowhow</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Секрет производства (ноу-хау)
6.	Информация, связанная с проведением экспертизы	<i>expert</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Информация, связанная с проведением экспертизы
7.	Инсайдерская информация	<i>insider</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Инсайдерская информация
8.	Информация, полученная при осуществлении своих полномочий службой внутреннего аудита публично-правовой компании	<i>audit</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Информация, полученная при осуществлении своих полномочий службой внутреннего аудита публично-правовой компании
9.	Сведения о сущности изобретения, полезной модели или промышленного образца до их официальной публикации	<i>invention</i> из пользовательского словаря <i>commercial</i>	<i>xsd:string</i>	Пользовательское значение свойства Сведения о сущности изобретения, полезной модели или промышленного образца до их официальной публикации

Разработка методических рекомендаций по заполнению паспорта набора данных

Заполнение ПНД выполняется должностным лицом, ответственным за предоставление наборов данных.

ПНД относится только к одной версии набора данных. При актуализации набора данных и формировании новой версии заполняются соответствующие свойства. Правила заполнения свойств паспортов наборов данных представлены в табл. 9.

Таблица 9

Правила заполнения свойств паспортов наборов данных

№ п/п	Свойство	Формат значения	Требования к заполнению	Пример записи
1.	Идентификатор (код) набора данных	Строка символов	Правила формирования идентификационных кодов наборов данных или Приложение 1 к Регламенту публикации набора данных (результата исследования)	7710168360-CheckSROA или Д21202.000123.1
2.	Наименование набора данных	Строка символов	ГОСТ 7.80—2000	План проведения плановых проверок саморегулируемых организаций аудиторов Министерством финансов Российской Федерации.
3.	Описание набора данных	Строка символов	ГОСТ 7.80—2000	Содержит сведения о плановых проверках саморегулируемых организаций аудиторов, осуществляемых Министерством финансов Российской Федерации в текущем году
4.	Обладатель набора данных	Строка символов	ГОСТ 7.80—2000 или Приложение 2 к Регламенту публикации набора данных (результата исследования)	Министерство финансов Российской Федерации или 0412

5.	Ответственное лицо	Ссылка на объект ответственного лица набора данных	ГОСТ 7.80—2000 ГОСТ Р 7.0.64-2018 RFC 5322	Полное имя ответственного лица. Телефон ответственного лица. Электронный адрес ответственного лица.
6.	Дата создания набора данных	Полная календарная дата	ГОСТ Р 7.0.64-2018	2013-06-19
7.	Дата последнего внесения изменений	Полная календарная дата	ГОСТ Р 7.0.64-2018	2020-10-21
8.	Периодичность актуализации набора данных	Строка символов	ГОСТ Р 7.0.64-2018	месяц
9.	Ключевые слова, соответствующие содержанию набора данных	Строка символов	ГОСТ Р 7.0.66-2010 СИБИД	Некоммерческое партнерство
10	Области знаний, к которым относится набор данных	Строка символов	Приложение 3 к Регламенту публикации набора данных (результата исследования)	0000080001/ Инфекционные заболевания/ Паразитология/Covid19
11.	Список всех версий наборов данных	Ссылка на объект коллекции всех версий наборов данных	ГОСТ Р 7.0.5-2008 RFC 3986	http://minfin.ru/OpenData/7710168360-CheckSROA/data-20191230-structure-20151111.csv

Таблица 10

Паспорт результата исследования как объект RDFa

№ п/п	Свойство	RDFa свойство	Формат значения	Значение
1.	Идентификатор (код) результата исследования	<i>identifier</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение идентификационного номера (кода) исследования
2.	Наименование результата исследования	<i>title</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение наименования результата исследования
3.	Сведения о результатах исследования	<i>description</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa описания результатов исследования
4.	Автор исследования	<i>creator</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение автора исследования
5.	Ответственное лицо	<i>publisher</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa ответственного лица, проводящего исследования
6.	Ограничения доступа	<i>rights</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект RDFa коллекции режимов доступа исследования
7.	Дата начала проведения исследования	<i>created</i> из словаря <i>dc</i>	<i>xsd:dateTime</i>	Пользовательское значение даты проведения исследования
8.	Дата последнего внесения изменений	<i>modified</i> из словаря <i>dc</i>	<i>xsd:dateTime</i>	Пользовательское значение даты последнего изменения

9.	Периодичность обновления, срок актуализации	<i>valid</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение периодичности обновления, срока актуализации исследований
10.	Ключевые слова исследований	<i>subject</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение ключевых слов исследований
11.	Области знаний	<i>description</i> из словаря <i>dc</i>	<i>xsd:string</i>	Пользовательское значение областей знаний, к которым относится результат исследования
12.	Список всех использованных в исследовании наборов данных	<i>source</i> из словаря <i>dc</i>	<i>xsd:anyURI</i>	Ссылка на объект <i>RDFa</i> коллекции всех использованных в исследовании наборов данных

Описание классификационных признаков, позволяющих полностью или по частям отнести результаты исследований к определенным классам

Оформление результатов исследования наборов данных предполагает составление паспорта результата исследования как обобщающего документа. паспорт результата исследования представляет собой описание сущности исследования в форме, позволяющей классифицировать результаты исследования. Формы представления результатов исследования могут быть следующие: текст как обобщение в виде научного доклада на конференции, семинаре; тезисы для сообщения на конференции, семинаре или публикации в сборнике научных трудов; научная статья, подготовленная для издания в научном журнале, сборнике; элементы НИР, ОКР; графические материалы; мультимедиа контент; другие формы, приведенные в табл. 10.

Паспорт результата исследования с иерархией объектов *RDFa* представлен на рис. 2.

Описание классификационных признаков паспорта результата исследования представлено в терминологии языков *XML* и *XSD* и в соответствии с рекомендациями *W3C* и ГОСТ Р 7.0.10-2019. Значения свойств объектов *RDFa* паспорта результата исследования представлены в табл. 10.

Расширенные свойства «Ответственное лицо» и «Список всех версий использованных в исследовании наборов данных» *RDFa* паспорта исследования оформляются аналогично соответствующим свойствам *RDFa* паспортов наборов данных.

Свойство «Сведения о результатах исследования» *RDFa* паспорта исследования ссылается на объект *RDFa*, содержащий квалификационные признаки, позволяющие отнести результаты исследования к определенному классу; представлено в табл. 11.

Таблица 11

Описания результатов исследования

№ п/п	Свойство	<i>RDFa</i> свойство	Формат значения	Значение
1.	Тема исследования	<i>topic</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение темы исследования
2.	Научное направление	<i>scientific</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение области научных знаний, к которой относится данное исследование
3.	Научная проблема	<i>problem</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение вопроса, противоречащего знаниям и требующего разрешения
4.	Объект исследования	<i>object</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение носителя проблемной ситуации, сферы деятельности, на которую направлена познавательная деятельность
5.	Предмет исследования	<i>subject</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение стороны, свойства, характеристики объекта, подлежащие непосредственному изучению
6.	Цель исследования и основные задачи	<i>purpose</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение предполагаемого результата исследования

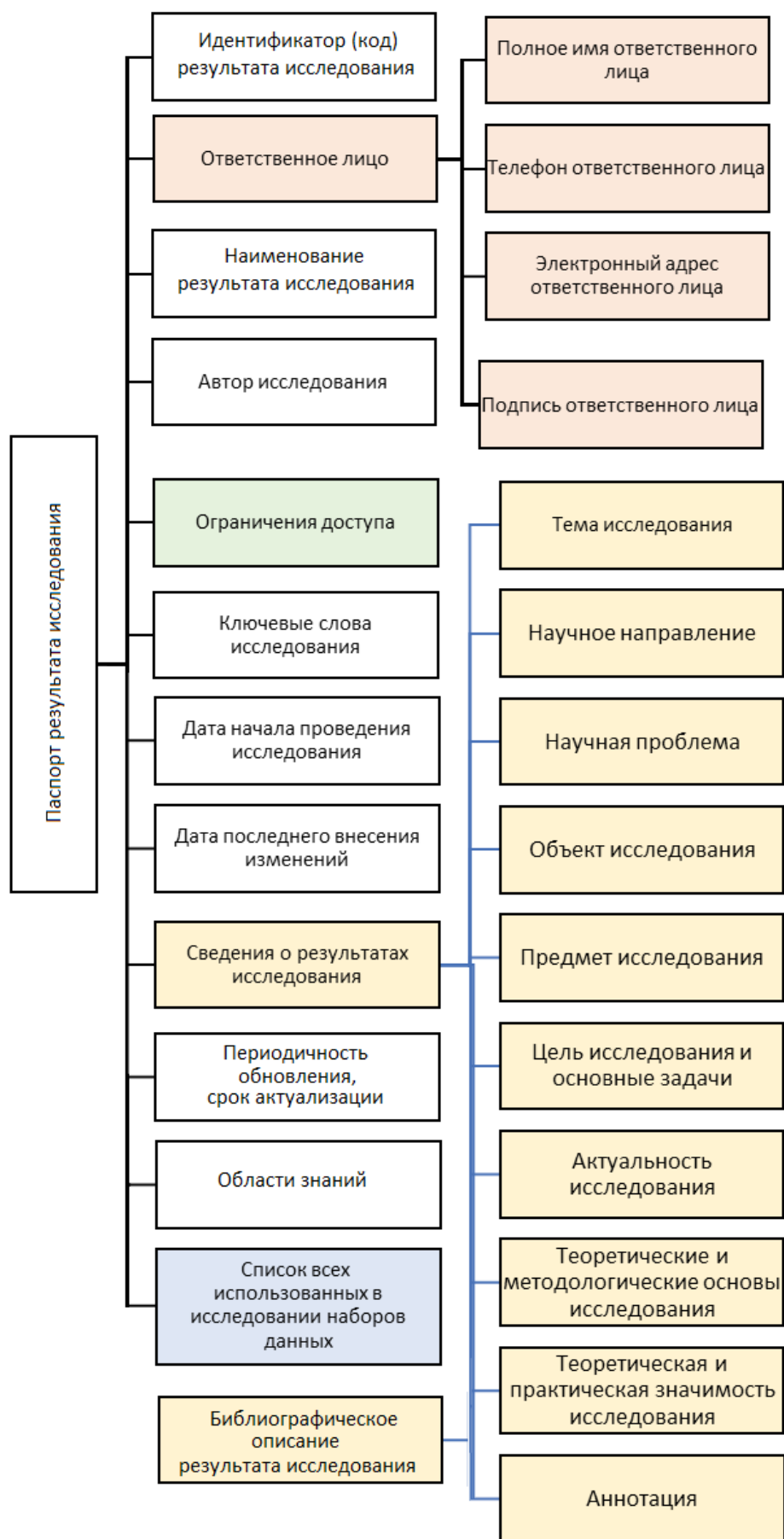


Рис. 2. Паспорт результата исследования с иерархией объектов RDFS

7.	Актуальность исследования	<i>relevance</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение потребности, причины, по которым эту тему необходимо изучать в настоящее время
8.	Теоретические и методологические основы исследования	<i>bases</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение действующие нормативные акты, научные подходы и используемые научные и специальные методы исследования
9.	Теоретическая и практическая значимость исследования	<i>practice</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение уточнение понятий, изменение структуры, развитие методов; использование результатов в различных областях жизнедеятельности человека
10.	Аннотация	<i>annotation</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение на русском языке
11.	Библиографическое описание работы	<i>sources</i> из пользовательского словаря <i>research</i>	<i>xsd:string</i>	Пользовательское значение списка библиографических источников

Заключение

Разработанные программно-функциональные модели паспорта набора данных и паспорта результата исследования содержат ключевые параметры, характеризующие набор данных и результат исследования соответственно, что позволяет их однозначно идентифицировать и обеспечивает возможность классифицировать набор данных и результат исследования. Особенностью предлагаемых моделей паспорта набора данных и паспорта результата исследования является использование для описания ключевых свойств унифицированных объектов, обеспечивающих необходимую гибкость структуры и возможность создания пользовательских классификаторов.

Структура и содержание паспорта набора данных и паспорта результата исследования соответствуют международным и отечественным стандартам представления метаданных, что позволяет выполнять логическую обработку и обмен данными в автоматизированных системах и открытых ресурсах [11].

Реализация паспорта набора данных и паспорта результата исследования рекомендуется в составе информационно-программного обеспечения существующих и перспективных информационных платформ организаций, принимающих, обрабатывающих и публикующих наборы данных с целью повышения уровня публичной открытости и доступности наборов данных и соблюдения законодательных ограничений на доступ к ним и обеспечения их информационной безопасности.

Рецензент: **Федосеев Сергей Витальевич**, кандидат технических наук, доцент, профессор кафедры информационного права, информатики и математики Российского государственного университета правосудия, г. Москва, Российская Федерация.

E-mail: fedsergvit@mail.ru

Литература

1. Борисов Р.С., Ефименко А.А. Протокол обработки наборов данных для их публикации в открытых источниках // Правовая информатика. 2021. № 2. С. 59—70. DOI: 10.21681/1994-1404-2021-2-59-70
2. Борисов Р.С., Ефименко А.А. Классификатор правовых актов для установления правового режима публикуемой информации // Правовая информатика. 2021. № 4. С. 31—45. DOI: 10.21681/1994-1404-2021-4-31-45
3. Борисов Р.С., Ефименко А.А. Анализ федерального законодательства об ограничении публикации открытых данных // Государственная власть и местное самоуправление. 2022. № 2. С. 42—47. DOI: 10.18572/1813-1247-2022-2-42-47
4. Ващекин А.Н., Дзедзинский А.В. Правовое регулирование отношений в цифровом пространстве // Правосудие. 2020. Т. 2. № 2. С. 126—147. DOI: 10.37399/issn2686-9241.2020.2-126-147
5. Информационная безопасность цифрового пространства / под ред. Е.В. Стельмашонок, И.Н. Васильевой. СПб. : Изд-во СПбГЭУ, 2019. 155 с.
6. Ловцов Д.А. Информационная теория эргасистем : монография. М. : РГУП, 2021. 314 с. ISBN 978-5-93916-887-8.

7. Ловцов Д.А. Системный анализ. Часть 1. Теоретические основы. М. : РГУП, 2018. 224 с. ISBN 978-5-93916-701-7.
8. Ловцов Д.А. Системология правового регулирования информационных отношений в инфосфере : монография. М. : РГУП, 2016. 316 с. ISBN 978-5-93916-505-1.
9. Ловцов Д.А., Богданова М.В. Экономико-правовое регулирование оборота результатов интеллектуальной деятельности предприятий промышленности России // Экономика, статистика и информатика. Вестник УМО. 2013. № 1. С. 53—56.
10. Ловцов Д.А., Федичев А.В. Архитектура национального классификатора правовых режимов информации ограниченного доступа // Правовая информатика. 2017. № 2. С. 35—54. DOI: 10.21681/1994-1404-2017-2-35-54
11. Марков А.С. Правоприменение открытых данных с учетом требований по информационной безопасности // Мониторинг правоприменения. 2017. № 3 (24). С. 86—96. DOI: 10.21681/2412-8163-2017-3-86-96
12. Федосеев С.В. Применение современных технологий больших данных в правовой сфере // Правовая информатика. 2018. № 4. С. 50—58. DOI: 10.21681/1994-1404-2018-4-5-58

PASSPORTS FOR DATA SETS AND RESEARCH RESULTS FOR THEIR PUBLICATION IN OPEN SOURCES

Roman Borisov⁵, Aleksei Efimenko⁶

Keywords: *open data, data set structure, metadata, data set passport, research results passport, data set publication, confidentiality, classification attributes.*

Abstract.

Purpose of the work: developing software and functional models of passports for data sets and research results containing classification attributes allowing to assign, fully or partly, data sets or research results to certain classes, as well as methodological recommendations for filling in a data set (research result) passport.

Methods used: system modelling of the program structure and organisational techniques for processing passports for data sets and research results, which would allow to secure the confidentiality of information when published in open sources.

Results obtained. Requirements for presenting passports for data sets and research results are identified, the process of forming and methodological recommendations for filling in a data set (research result) passport are described. The format is presented for the sections of the Data Set Passport and the Research Result Passport which are required to be filled in when submitting a data set (research result) for the possibility of their open access publication.

References

1. Borisov R.S., Efimenko A.A. Protokol obrabotki naborov dannykh dlia ikh publikatsii v otkrytykh istochnikakh. Pravovaia informatika, 2021, No. 2, pp. 59–70. DOI: 10.21681/1994-1404-2021-2-59-70
2. Borisov R.S., Efimenko A.A. Klassifikator pravovykh aktov dlia ustanovleniia pravovogo rezhima publikuemoi informatsii. Pravovaia informatika, 2021, No. 4, pp. 31–45. DOI: 10.21681/1994-1404-2021-4-31-45
3. Borisov R.S., Efimenko A.A. Analiz federal'nogo zakonodatel'stva ob ogranichenii publikatsii otkrytykh dannykh. Gosudarstvennaia vlast' i mestnoe samoupravlenie, 2022, No. 2, pp. 42–47. DOI: 10.18572/1813-1247-2022-2-42-47
4. Vashchekin A.N., Dzedzinskii A.V. Pravovoe regulirovanie otnoshenii v tsifrovom prostranstve. Pravosudie, 2020. T. 2, No. 2, pp. 126–147. DOI: 10.37399/issn2686-9241.2020.2-126-147
5. Informatsionnaia bezopasnost' tsifrovogo prostranstva. Pod red. E.V. Stel'mashonok, I.N. Vasil'evoi. SPb. : Izd-vo SPbGEU, 2019. 155 s.
6. Lovtsov D.A. Informatsionnaia teoriia ergasistem : monografiia. M. : RGUP, 2021. 314 s. ISBN 978-5-93916-887-8.
7. Lovtsov D.A. Sistemnyi analiz. Chast' 1. Teoreticheskie osnovy. M. : RGUP, 2018. 224 s. ISBN 978-5-93916-701-7.
8. Lovtsov D.A. Sistemologiya pravovogo regulirovaniia informatsionnykh otnoshenii v infosfere : monografiia. M. : RGUP, 2016. 316 s. ISBN 978-5-93916-505-1.

⁵ **Roman Borisov**, Ph.D. (Technology), Associate Professor at the Department of Information Technology Law, Informatics and Mathematics of the Russian State University of Justice, Moscow, Russian Federation.
E-mail: bestseller@bk.ru

⁶ **Aleksei Efimenko**, Ph.D. (Technology), Associate Professor at the Department of Information Technology Law, Informatics and Mathematics of the Russian State University of Justice, Moscow, Russian Federation.
E-mail: alex192@mail.ru

9. Lovtsov D.A., Bogdanova M.V. Ekonomiko-pravovoe regulirovanie oborota rezul'tatov intellektual'noi deiatel'nosti predpriatii promyshlennosti Rossii. *Ekonomika, statistika i informatika. Vestnik UMO*, 2013, No. 1, pp. 53–56.
10. Lovtsov D.A., Fedichev A.V. Arkhitektura natsional'nogo klassifikatora pravovykh rezhimov informatsii ogranichennogo dostupa. *Pravovaia informatika*, 2017, No. 2, pp. 35–54. DOI: 10.21681/1994-1404-2017-2-35-54
11. Markov A.S. Pravoprimerenie otkrytykh dannykh s uchetom trebovaniy po informatsionnoi bezopasnosti. *Monitoring pravoprimereniia*, 2017, No. 3 (24), pp. 86–96. DOI: 10.21681/2412-8163-2017-3-86-96
12. Fedoseev S.V. Primenenie sovremennykh tekhnologii bol'shikh dannykh v pravovoi sfere. *Pravovaia informatika*, 2018, No. 4, pp. 50–58. DOI: 10.21681/1994-1404-2018-4-5-58