

Цифровое пространстве науки, постоянные идентификаторы и библиотеки

Соколова Наталия Викторовна, директор центра информационно-библиотечных систем Санкт-Петербургский политехнический университет Петра Великого

Парадигмы науки

• Тысячи лет назад:

• наука эмпирическая: описание наблюдаемых явлений

• Сотни лет назад:

• науки теоретическая: модели, обобщения

• Десятки лет назад:

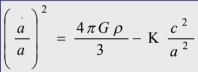
• вычислительные модели

• Сегодня:

• eScience – объединение теории, экспериментов и цифрового моделирования

Jim Gray, eScience Group, Microsoft Research









Эволюция «открытости»

Открытый доступ

Открытые данные Открытая наука Initiative for Open Citations I4OC
Initiative for Open Abstracts I4OA
Постоянные идентификаторы (ПИД, PID)
The Principles of Open Scholarly
Infrastructure





Функциональные требования к постоянным идентификаторам (PID, ПИД)

- Это **имя**, а не адрес
- Уникальность на глобальном уровне
- Постоянство, ориентированное на время, более продолжительное, чем время жизни отдельной системы или организации
- Управление на уровне уполномоченной организации, ответственной за организацию процесса
- Сопровождаться **метаданными**, которые раскрывают наиболее важные свойства, включая минимально необходимый набор элементов
- Обладать свойством связывания с другими ПИД
- Обладать свойством **интероперабельности** (interoperable) с другими идентификаторами через элементы метаданных, описывающие соответствующие связи
- Обеспечено **индексирование и поиск** по метаданным и по другим ПИД (ODIN Consortium)

DataCite, CrossRef (DOI)

DOI, ePIC (Handle)
LSID, NBN (URN)

Handle, URN, ARK, XRI

Сущности в процессе научного исследования

- 1. Агенты: личности (такие как исследователи и кураторы); организации(такие как фонды, научные институты, датацентры, издательства и архивы), другие юрлица (такие как консорциумы, организованные для проведения исследования)
- **2. Ресурсы**: публикации, исследовательские данные, артефакты и средства исследования, такие как компьютеры, ПО, лабораторное оборудование, физические образцы
- 3. Права (правоообладание): условия грантов, лицензии, патенты
- 4. События: процессы/явления, релевантные природе ресурсов (объектов исследования), такие как создание, курирование, доступ, предъявление прав, обновление, цитирование, рецензирование
- 5. Производные сущности: такие как проекты, объединяющие вовлеченные организации-исполнители, финансирующие организации, полученные результаты и условия доступа к РИД

ПИД для исследователей и всего происходящего в научном сообществе

PIDs for people (researchers) include ISNIs and ORCID iDs



PIDs for places (research organizations) include GRID and ROR



PIDs for things (research outputs/inputs like grants, reviews, preprints, projects, etc.) include Crossref and DataCite DOIs, IGSNs, RAiDs, and more



ПИД для сущностей в научном процессе

Постоянный идентификатор	Чему/кому присваивается	Где зарегистрировать
DOI	Pecypc	DataCite, CrossRef и другие
ORCID iD	Исследователь (физлицо)	ORCID https://orcid.org/
ROR	Организация	Research Organization Registry https://ror.org/
Funder ID	Финансирующая организация, фонд	Funder Registry https://www.crossref.org/services/funder-registry/













Provenance

Metadata

What's what?

Who's who?

Can you tell me more about it?

What?

Who?

A persistent identifier is a long-lasting reference to a digital resource

How long is long?

How do I know?

Policies and Guarantees

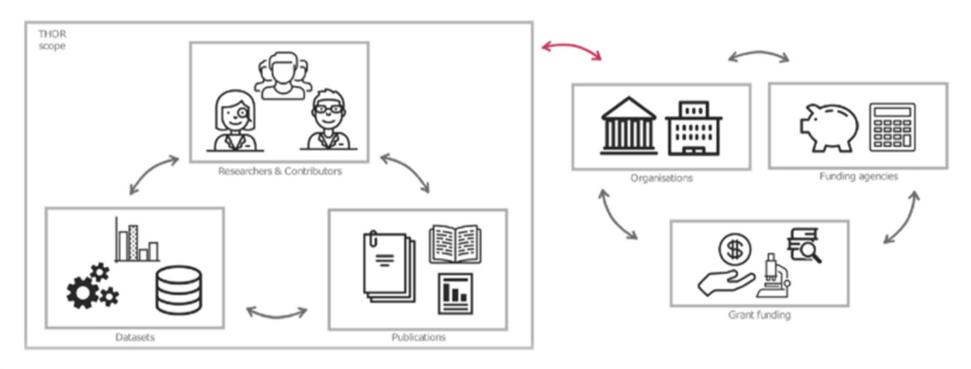
Where can my machine find it?

Where can I find it?

Machine-readability

Инфраструктура для любых (не только цифровых) ресурсов

Пространство, основанное на ПИД



Полнота графа знаний зависит от **качества и полноты метаданных,** указанных при регистрации ПИД

Пример https://commons.datacite.org/orcid.org/0000-0003-3484-6875

Что есть уже сейчас В DataCite Commons The following main data sources are used in DataCite Commons for a total of currently 38,472,046 records:

DataCite

19,897,461 Works 100% of identifiers and metadata.

Crossref

8,771,341 Works
7.48% of identifiers and
metadata. Import is ongoing.

ORCID

9,704,912 People 100% of identifiers. Personal and employment metadata.

ROR

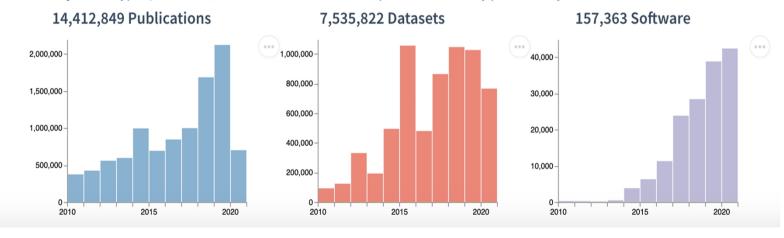
98,332 Organizations 100% of identifiers and metadata.

Additional information comes from these data sources:

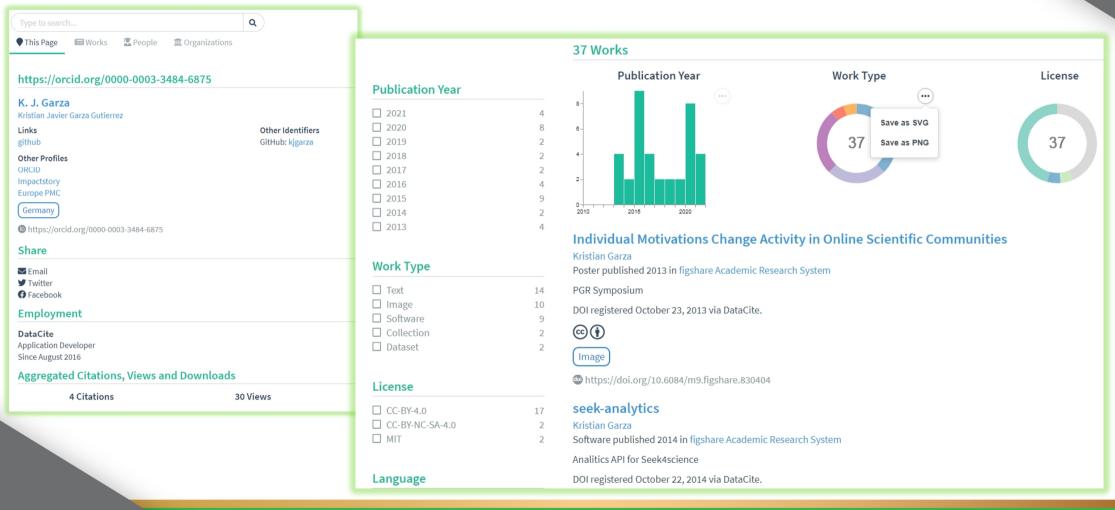
- Wikidata: inception year, geolocation and Twitter account for organizations
- Unpaywall: download link for Open Access content via Crossref

Works

DataCite Commons currently includes 28,668,802 works, with identifiers and metadata provided by DataCite and Crossref. For the three major work types publication, dataset and software, the respective numbers by publication year are shown below.



DataCite Commons. Проект FREYA (завершен в 2020)



Views, Downloads and Citations

Число показателей будет расширено - Make Data Count grant (Alfred P. Sloan Foundation), начало проекта в августе 2021

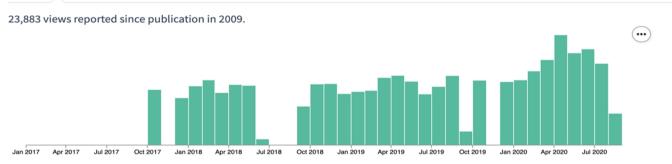


2020	2
2018	12
2017	12
2016	13
2015	1
2013	1
2009	1

Work Type



23,883 Views 14,652 Views



1 Reference 42

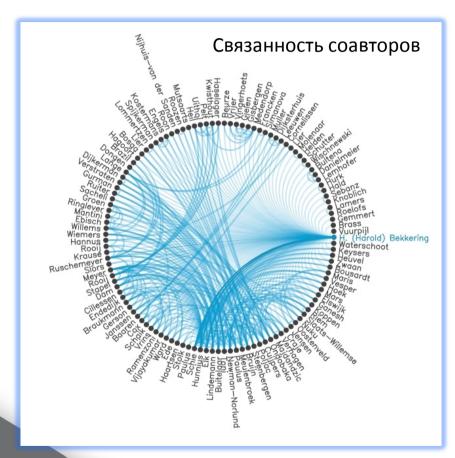
42 Citations



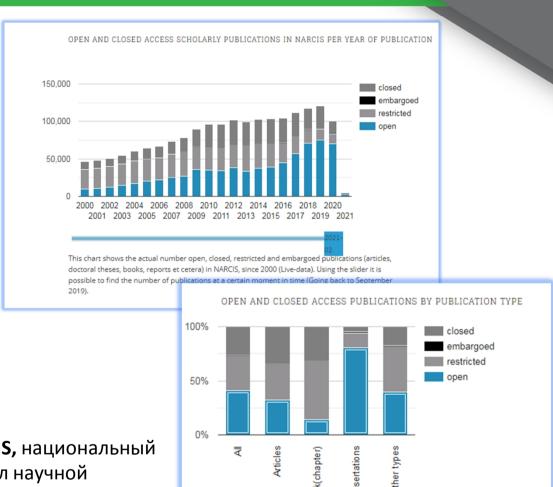
Long-Term Hydraulic Adjustment of Three Tropical Moist Forest Tree Species to Changing Climate

Mahmuda Islam, Mahmuda Islam & Mahmuda Islam Journal Article published 2018 in Frontiers in Plant Science

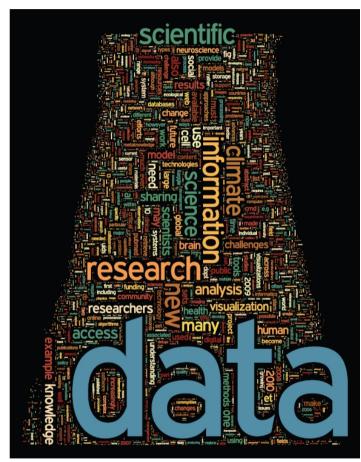
Что могут ПИД?



NARCIS, национальный портал научной информации Нидерландов https://www.narcis.nl/



Создание надежной мировой инфраструктуры, в которой каждая сущность, вовлеченная в научное исследование, имеет постоянное уникальное имя, по которому можно не только точно и однозначно ее найти, но также точно и однозначно процитировать, учесть все ссылки на ресурс при расчете различных показателей



Цифровой идентификатор объекта (DOI)

- Постоянное имя и гиперссылка
- Обычно выражается через URL
- Определяет ресурс не место его размещения
- Имеет связанные с ним метаданные
- Стандарт ISO 26324 (2012)
- Уникальная алфавитно-цифровая строка, назначенная регистрирующим агентством



Назначение DOI

DOI обеспечивает разрешимую, постоянную и интероперабельную ссылку:

- разрешимая стандартный синтаксис идентификатора и механизм отображения ссылок в сети (Handle System)
- постоянная на основе:
 - инфокоммуникационной инфраструктуры (реестр БД, поддержка proxy, пр.)
 - социальной инфраструктуры (обязательства регистрирующих агентств)
- интероперабельная благодаря модели данных (семантическая интероперабельность)



Модель управления

Международный фонд DOI (International DOI Foundation, IDF)

- Обеспечивает техническую и социальную инфраструктуру
- Решение общих задач (стандарты, лицензии, пр.)

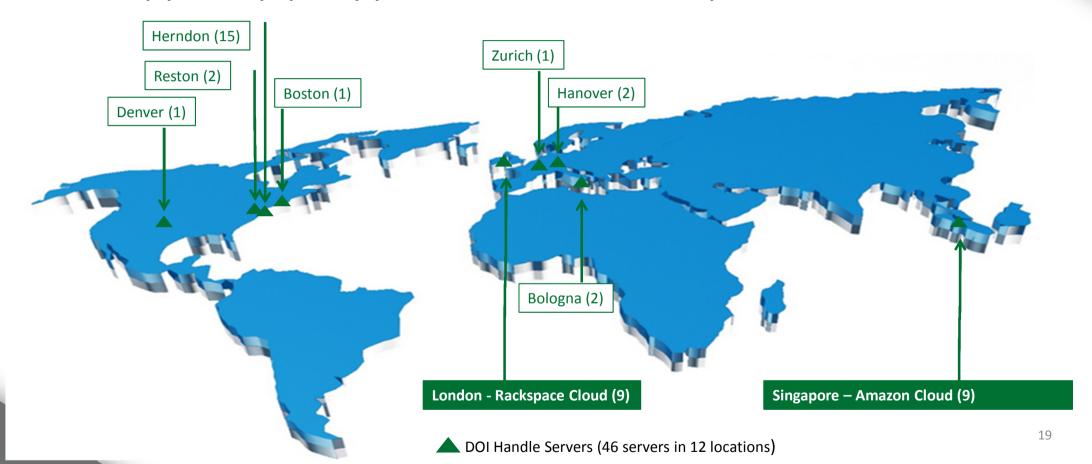
Регистрирующие агентства (РА)

- Автономны и независимы друг от друга
- Самостоятельно определяют бизнес модель взаимодействия с пользователями
- Все РА разделяют затраты на поддержание инфраструктуры (IDF)
- Все РА поддерживают стандарты системы DOI
- РА могут иметь собственные доп. сервисы

Пользователи сервисов регистрирующих агентств



Инфраструктура DOI Handle System





Регистрирующие агентства







Фокус:

Цифровые библиотеки

Миссия:

Содействие совместному использованию данных

Фокус:

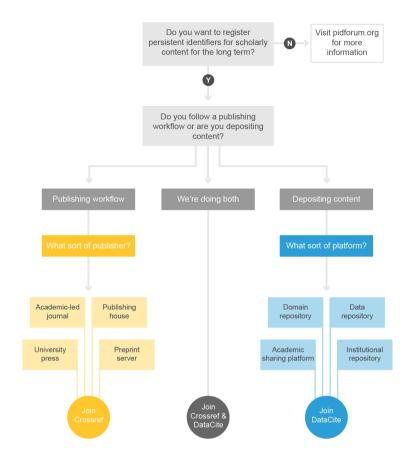
Эл. архивы журналов Миссия: устойчивая инфраструктура для научного обмена

CrossRef и DataCite

Crossref makes research outputs easy to **find**, **cite**, **link**, **assess**, and reuse.

DataCite's mission is to be the world's leading provider of persistent identifiers for research. Through our portfolio of services, we provide the means to create, find, cite, connect, and use research.

https://www.crossref.org/community/datacite/ https://support.datacite.org/docs/datacite-or-crossref





Crossref is a membership organisation for scholarly publishers providing services to help the scholarly community find, cite, link, assess and re-use published content.



Datacite is a membership organisation working with the respository community to register DOIs for a range of content and provides services to create, find, cite, connect, and use research.

DOI в библиотеках, музеях, архивах



Научные данные

Научные данные - информационные ресурсы любого вида и формата, собранные и/или сгенерированные в ходе научных исследований, представляющие значимость для научнообразовательной деятельности, например, научные статьи, материалы конференций, данные результатов эксперимента/опроса/анализа, тексты программ для ЭВМ, рисунки, схемы, фотографии, отчетная научно-техническая документация, авторефераты, диссертации, файлы, Web-страницы, учебная и учебно-методическая литература и т.д.

Research data, взгляд издателя



ELSEVIER

What is research data?

https://www.elsevier.com/authors/t ools-and-resources/research-data While the definition often differs per field, generally, research data refers to the results of observations or experiments that validate your research findings. These span a range of useful materials associated with your research project, including:

- •Raw or processed data files
- Software
- •Code
- Models
- Algorithms
- Protocols
- Methods

Research data *does not* include text in manuscript or final published article form, or data or other materials submitted and published as part of a journal article.

Взгляд NATURE

CAREER FEATURE · 13 MAY 2019

Data sharing and how it can benefit your scientific career

Open science can lead to greater collaboration, increased confidence in findings and goodwill between researchers.

Gabriel Popkin







Find a new job



Data sharing can be complex for scientists to navigate, but the rewards are often careel enhancing. Credit: Hero Images/Getty

> PDF version

RELATED ARTICLES

"A love letter to your future self": What scientists need to know about FAIR data



Why Jupyter is data scientists' computational notebook of choice



NatureTech



How the way scientists share data is changing

OPENING UP OPEN DATA

A growing number of websites provide tools, information and training regarding requirements and practices for open data.

Online tools

The journal Scientific Data recommends a number of general and discipline-specific data repositories.

Altmetric: An online platform that tracks data on the impact of research.

figshare: An open-access repository for preserving and sharing research outputs, including figures and images. (figshare is owned by Digital Science, a company operated by Holtzbrinck, which has a share in *Nature's* publisher.)

Open Science Badges: The Centre for Open Science's web badges show readers that content is openly available.

 $\label{lem:Jupyter Notebook: An open-source web app for creating and sharing scientific data and text.$

Information and training

 $\label{thm:continuous} European Union projects FOSTER Plus and OpenAIRE provide training on open science and open data.$

The New England Journal of Medicine has published advice from researchers on sharing clinical data responsibly (M. M. Mello et al. N. Engl. J. Med. 369, 1651–1658; 2013).

A summary of the US National Academies of Sciences, Engineering, and Medicine 2018 report *Open Science by Design*.

https://www.nature.com/articles/d41586-019-01506-x

FAIR научные данные

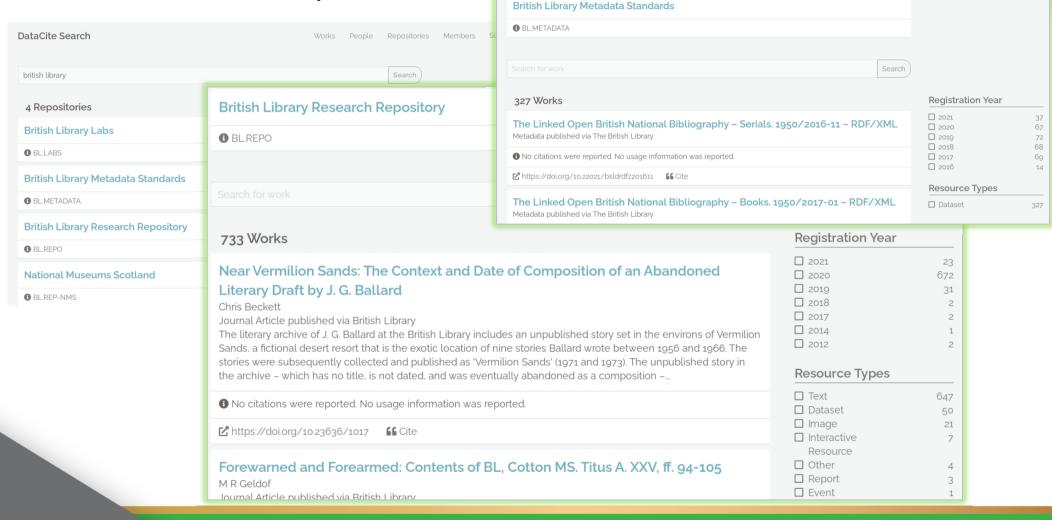


standards

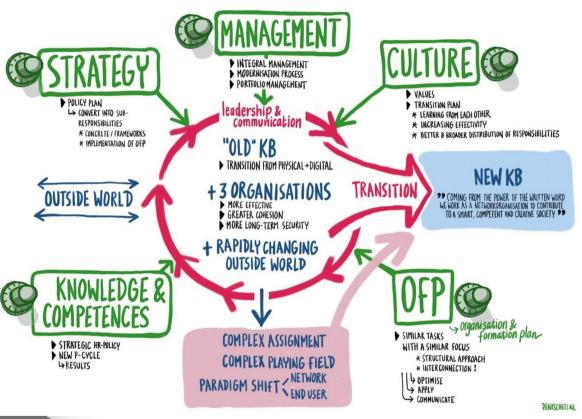


Finda	able	Acce	ssible	Inte	roperable	Reus	able
1	(мета)данные имеют уникальный постоянный идентификатор	1	(мета)данные могут быть найдены по идентификатору с использованием стандартного коммуникационного протокола	1	(мета)данные используют формальный, доступный и широко распространенный язык для представления знаний	1	(мета)данные являются «богатыми», с множеством точных и релевантных атрибутов
2	данные описаны «богатыми» метаданными	1.1	протокол доступа является открытым и широко распространенным	2	(мета)данные используют словари, следующие FAIR принципам	1.1	(мета)данные опубликованы с ясной лицензией, дающей право на использование данных
3	метаданные включают ясный и однозначный и однозначный идентификатор объекта, который они описывают	1.2	протокол обеспечивает аутентификацию и авторизацию, когда это требуется	3	(мета)данные включают применимые ссылки на другие (мета)данные	1.2	(мета)данные однозначно ассоциированы с источником их происхождения
4	(мета)данные зарегистрированы/ проиндексированы в поисковых системах	2	метаданные доступны даже тогда, когда сами данные становятся недоступны	4	(мета)данные должны быть в стандартном формате обмена	1.3	(мета)данные соответствуют стандартам предметной области

British Library



Национальная библиотека Нидерландов



Dutch Digital Heritage Network

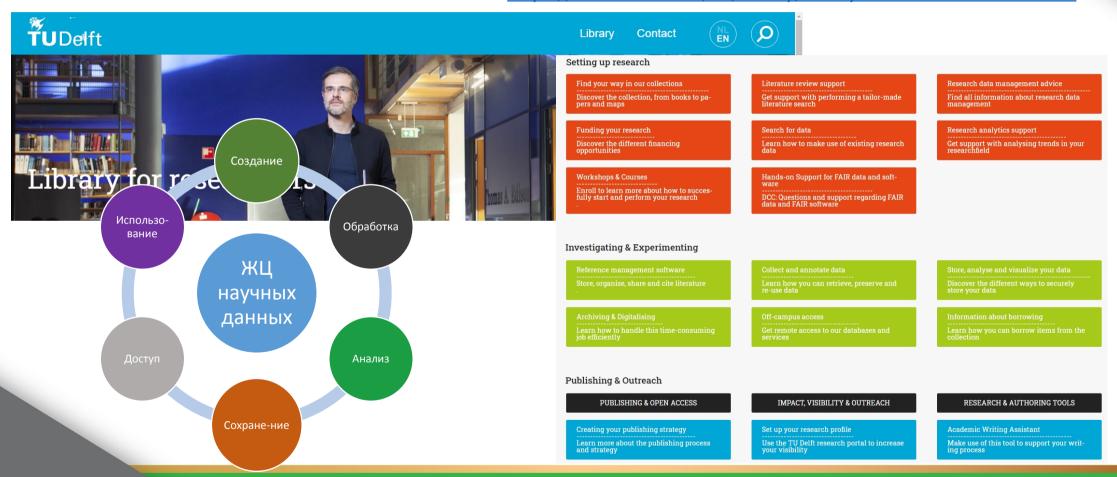
The Dutch Digital Heritage Network (Netwerk Digitaal Erfgoed) is a partnership in the Netherlands that focuses on developing a system of national facilities and services for improving the visibility, usability, and sustainability of digital heritage. The network is open to all institutions and organisations in the digital heritage field. Together we can make the most of our digital heritage and preserve it for future generations.





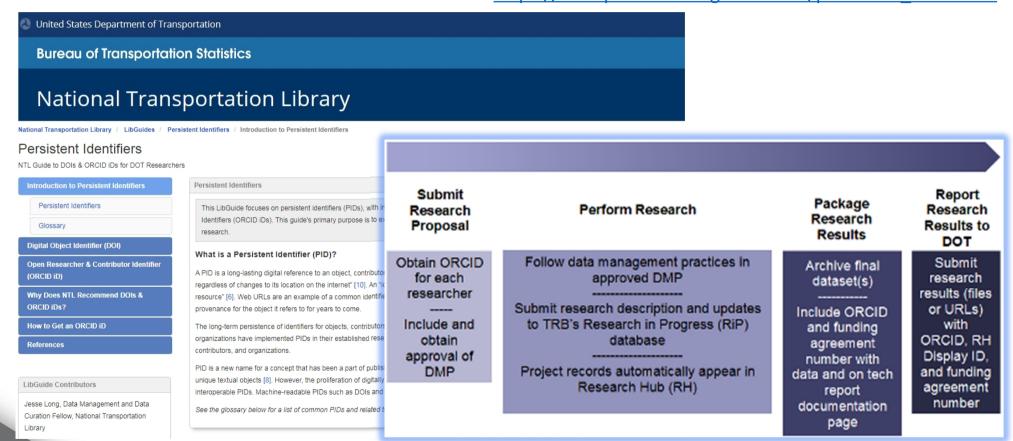
Библиотека TU Delft

https://www.tudelft.nl/en/library/library-for-researchers#c225799



Библиотека Министерства транспорта США

https://transportation.libguides.com/persistent_identifiers

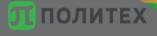




Что это дает регистрация DOI?

- Повышение статуса ресурса
- Подтверждение достоверности исследования
- Сохранение и распространение неопубликованных материалов, артефактов и пр.
- Облегчение процесса цитирования
- DOI основа современных научных коммуникаций
- Содействие поиску и цитированию ресурса, в том числе в зарубежных работах

Новая роль и новые услуги библиотеки в цифровой среде



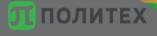
Повышает ли DOI цитируемость публикации?

Напрямую – НЕТ

• DOI не связан с цитируемостью. Но DOI связан с цитированием.

Косвенно – ОБЯЗАТЕЛЬНО!

- DOI повышает авторитет издания
- DOI значительно облегчает процедуры цитирования, поиска и локализации научной публикации
- DOI является сейчас неотъемлемым атрибутом современной системы научной коммуникации за счет эффективного обеспечения процессов обмена научной информацией



Как работают DOI при учете цитирований

Crossref as a new source of citation data: A comparison with Web of Science and Scopus

Nees Jan van Eck, Ludo Waltman, Vincent Larivière, Cassidy Sugimoto

January 17th, 2018

No comments

WoS и Scopus создают собственные базы метаданных.

DOI является желательным атрибутом — дает более точный учет использований. Не все DOI для ресурсов WoS и Scopus назначены CrossRef.

Table 1. Number of publications in WoS and Scopus, with a breakdown based on whether a publication has a DOI and whether it has a Crossref match (in millions; period 2012–2016).

	All docum	nent types	Research and review articles		
	WoS	Scopus	WoS	Scopus	
All pub.	11.9 (100.0%)	13.9 (100.0%)	7.6 (100.0%)	9.9 (100.0%)	
Pub. with DOI	8.3 (69.6%)	11.3 (80.9%)	6.8 (90.2%)	8.3 (83.8%)	
Pub. with Crossref match	8.2 (68.3%)	10.7 (76.9%)	6.7 (88.9%)	7.9 (79.7%)	



Missing references («потерянные» ссылки)

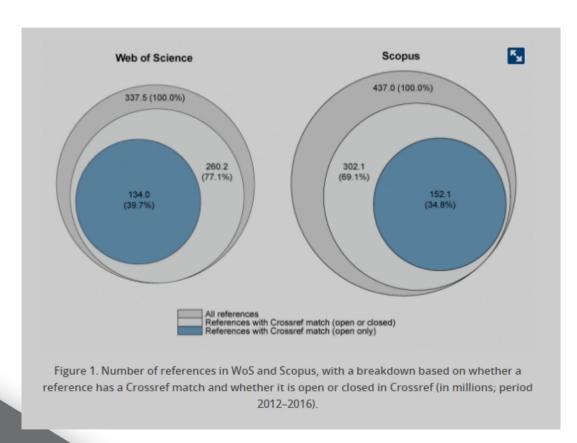


Table 2. Number of references in Crossref, and number of missing references, based on comparisons with WoS and Scopus (in millions; period 2012–2016).

References in Crossref, both open and closed	339.2
Missing references in Crossref, based on a comparison with WoS	34.2
Missing references in Crossref, based on a comparison with Scopus	64.5

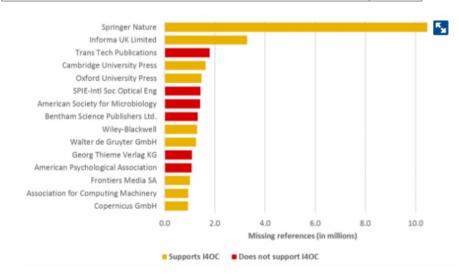
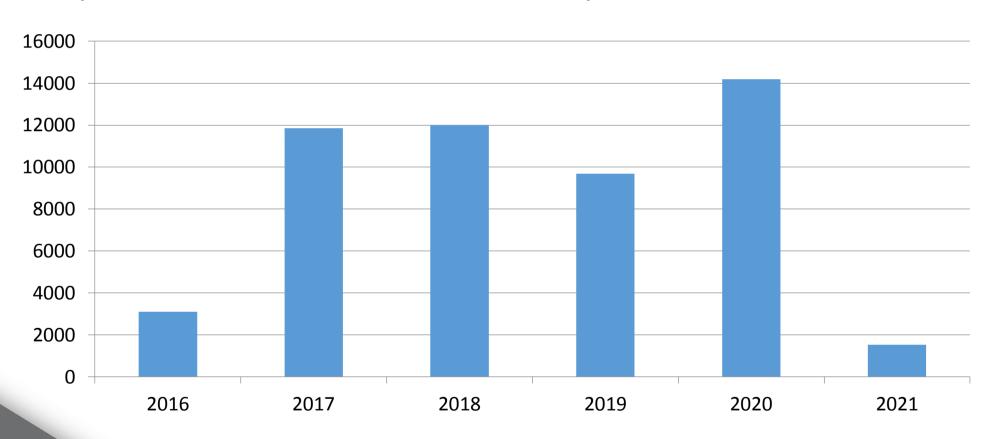
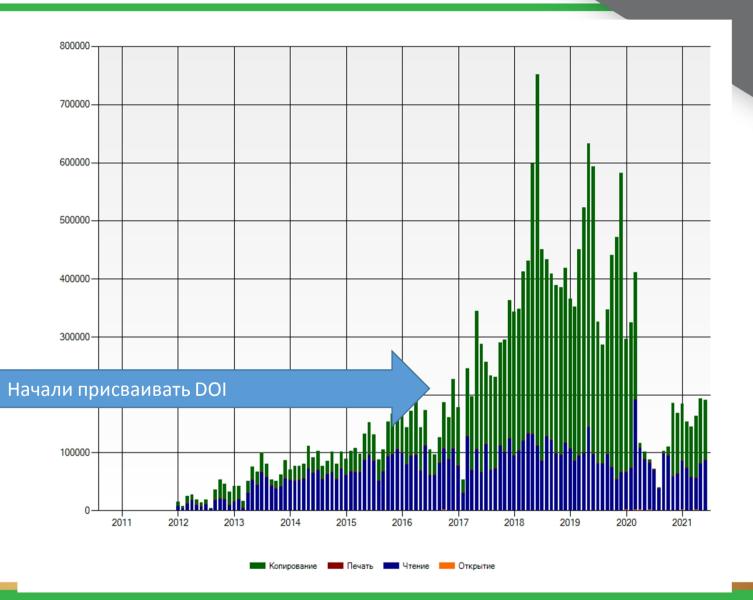


Figure 3. Top 15 publishers with the largest number of missing references in Crossref, based on a comparison with Scopus (in millions; period 2012–2016).

Национальный центр идентификации научных данных СПбПУ, присвоено DOI



Электронная библиотека СПбПУ. Статистика использования



Приглашение к сотрудничеству



Спасибо за внимание!

Соколова Наталия Викторовна natalia@unilib.spbstu.ru

- https://doi.spbstu.ru Национальный центр идентификации научных данных
- https://library.spbstu.ru Информационно-библиотечный комплекс СПбПУ
- https://elib.spbstu.ru Электронная библиотека СПбПУ