

Открытая наука как глобальное пространство связанных результатов исследований

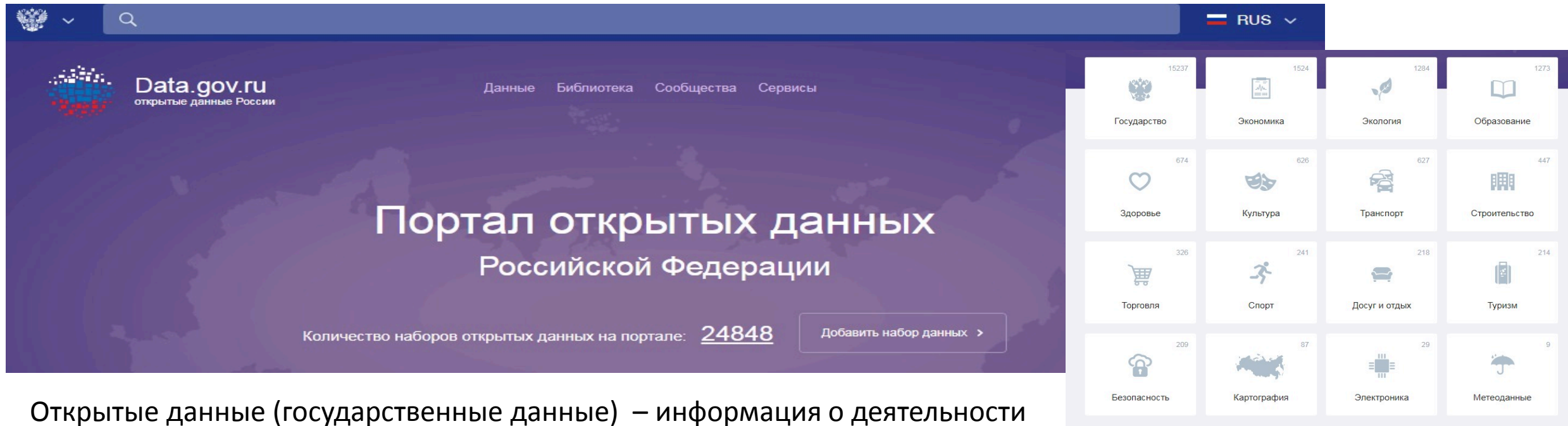
Соколова Наталия Викторовна, директор центра информационных систем
Санкт-Петербургский политехнический университет Петра Великого



Глобальное облако
открытой науки

<https://forumspb.com/programme/business-programme/97149/#broadcast>

Открытые данные России



The screenshot shows the homepage of the Data.gov.ru portal. The header includes the Russian coat of arms, a search bar, and a language selector set to 'RUS'. The main banner features the 'Data.gov.ru' logo and the text 'Портал открытых данных Российской Федерации'. Below this, it states 'Количество наборов открытых данных на портале: 24848' and provides a button to 'Добавить набор данных'. On the right side, there is a grid of 16 categories, each with an icon, a name, and a count of datasets.

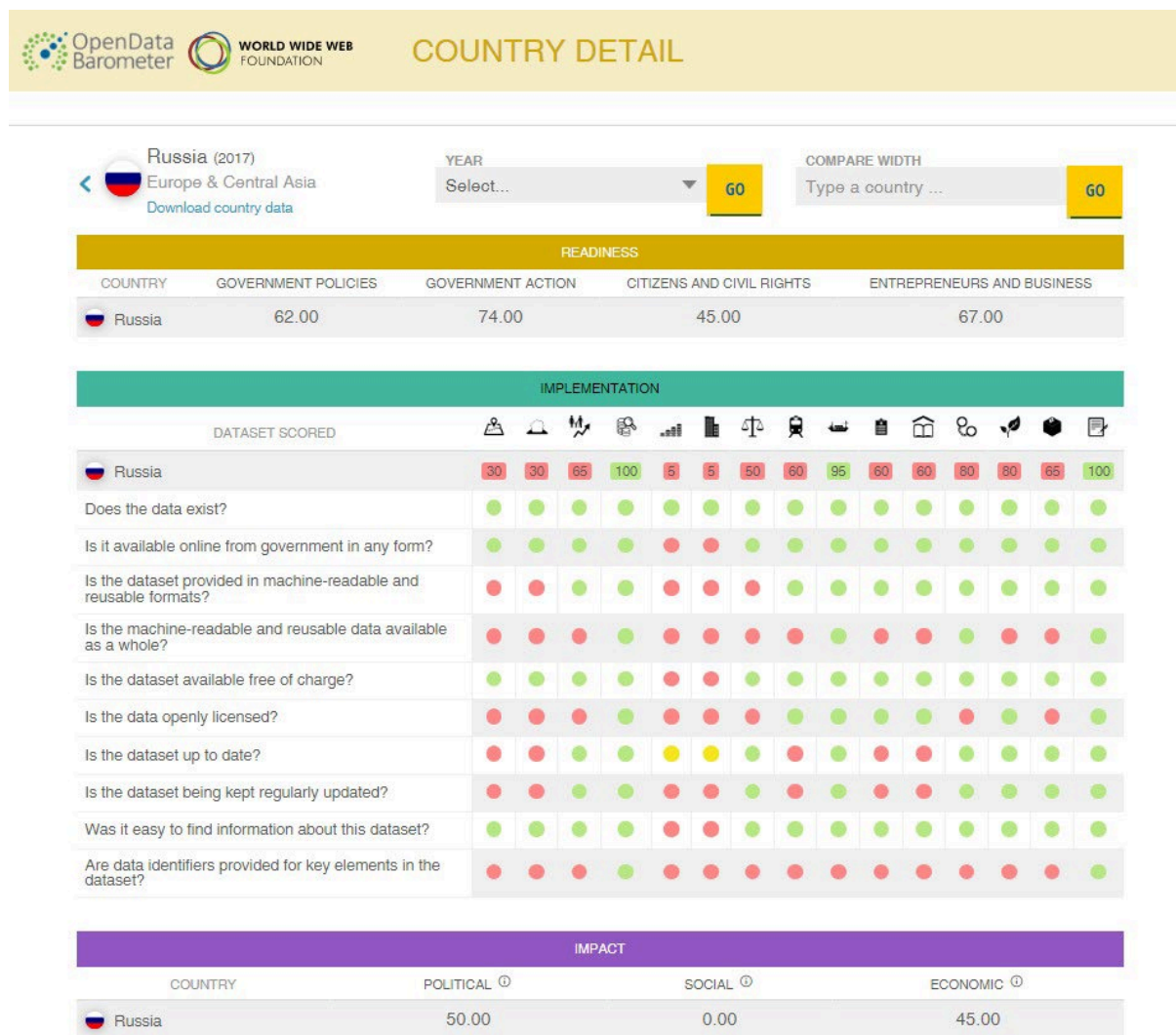
Category	Count
Государство	15237
Экономика	1524
Экология	1284
Образование	1273
Здоровье	674
Культура	626
Транспорт	627
Строительство	447
Торговля	326
Спорт	241
Досуг и отдых	218
Туризм	214
Безопасность	209
Картография	87
Электроника	29
Метеоданные	9

Открытые данные (государственные данные) – информация о деятельности государственных органов и органов местного самоуправления, размещенная в сети Интернет в виде массивов данных в формате, обеспечивающем их автоматизированную обработку в целях повторного использования без предварительного изменения человеком (машиночитаемый формат), и на условиях ее свободного (бесплатного) использования.

Данные сгруппированы по 16 областям деятельности.

Науки среди них нет

Open data barometer



Глобальная система, показывающая уровень поддержки государством использования открытых данных в области статистики/учета, внедрения инноваций и социальной политики

<https://opendatabarometer.org/>

Открытая наука



Источник: UNESCO Recommendation on Open Science, 2021

Основные цели:

- Публичная доступность и полная прозрачность научных коммуникаций.
- Публичная доступность и повторное использование научных данных.
- Прозрачность методов проведения экспериментов, наблюдений, сбора данных.
- Полная научная коллаборация.

Основные потребности:

- Укрепление диалога между наукой и обществом.
- Связанность исследователей с формированием политики развития науки.
- Развитие надлежащих е-инфраструктур, цифровых средств и сервисов для открытой науки.
- Изменение законодательства и политики лицензирования для открытой науки.

Создание новой устойчивой инфраструктуры:

- Подготовки квалифицированных специалистов в области открытой науки.
- Ответственное отношение исследователей, соответствующее принятым в науке ценностям и вызывающего доверие – научная частота и честность

Рекомендации группы Research, Innovation and Science Policy Experts (RISE)

Эволюция открытости

Хотя открытость данных полезна с позиции ускорения науки и инноваций, требуется **достижение баланса** и учетом стоимости, приватности, безопасности, защиты интеллектуальной собственности, предотвращения недобросовестного использования данных

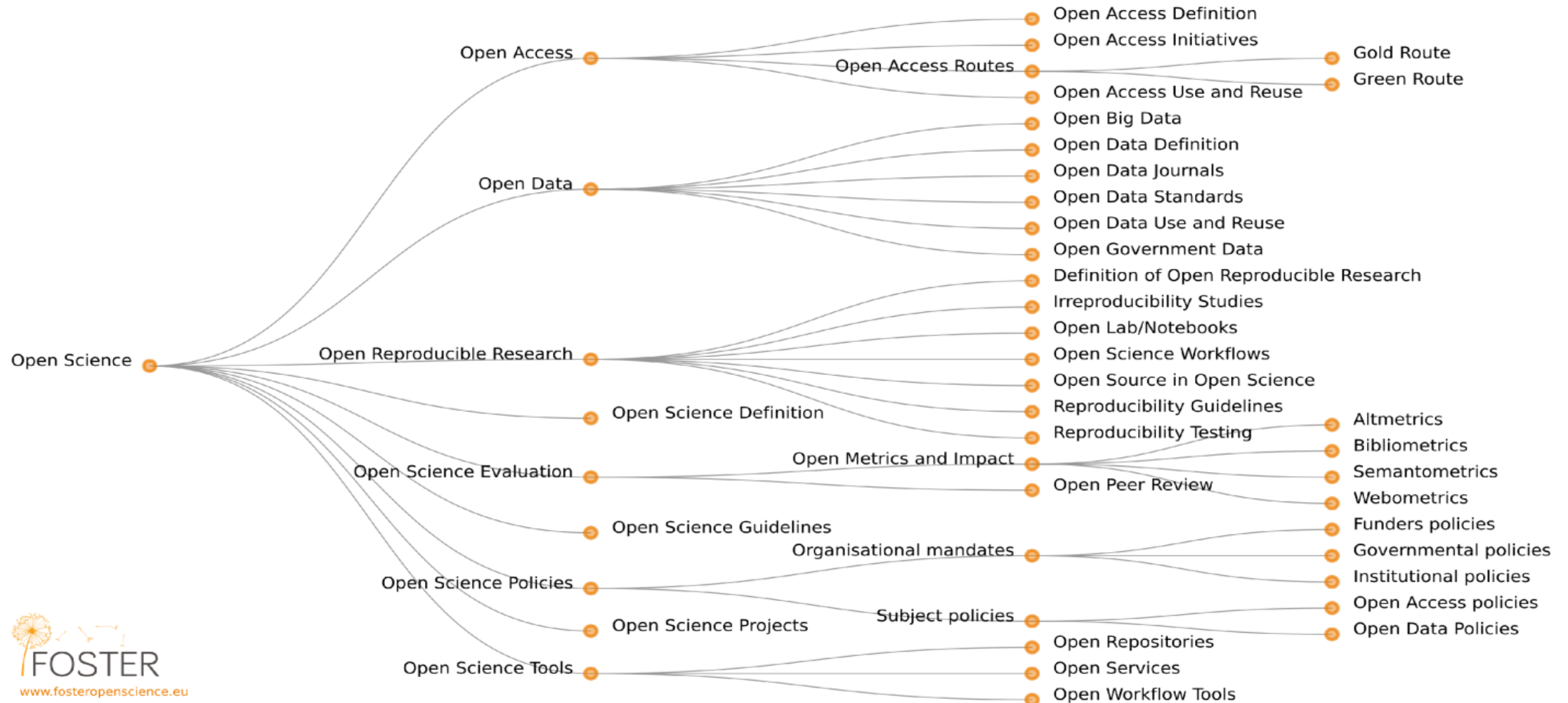
Термин **«улучшенный (enhanced)»** доступ к данным используется все чаще в отношении научных данных, полученных на средства госбюджета. Данный подход фиксирует важные предупреждения в части возможных негативных последствий от полной открытости всех данных.

Различные уровни открытости могут включать:

- открытый доступ с открытой лицензией;
- публичный доступ со специфической лицензией, ограничивающей использование данных;
- ограниченный доступ только группы аутентифицированных пользователей;
- ограниченный доступ только лицам, указанным в контракте

Простое определение открытых данных говорит, что право доступа и повторного использования имеют **все без каких-либо технических или юридических ограничений. Это совсем не обязательно означает, что доступ бесплатен**, хотя в контексте открытой науки считается нормой, что пользователь данных не несет расходов

Таксономия открытой науки



Цифровая трансформация науки

Сейчас конфликт: новые методы/объекты/инструменты исследований, но традиционные методы научных коммуникаций.

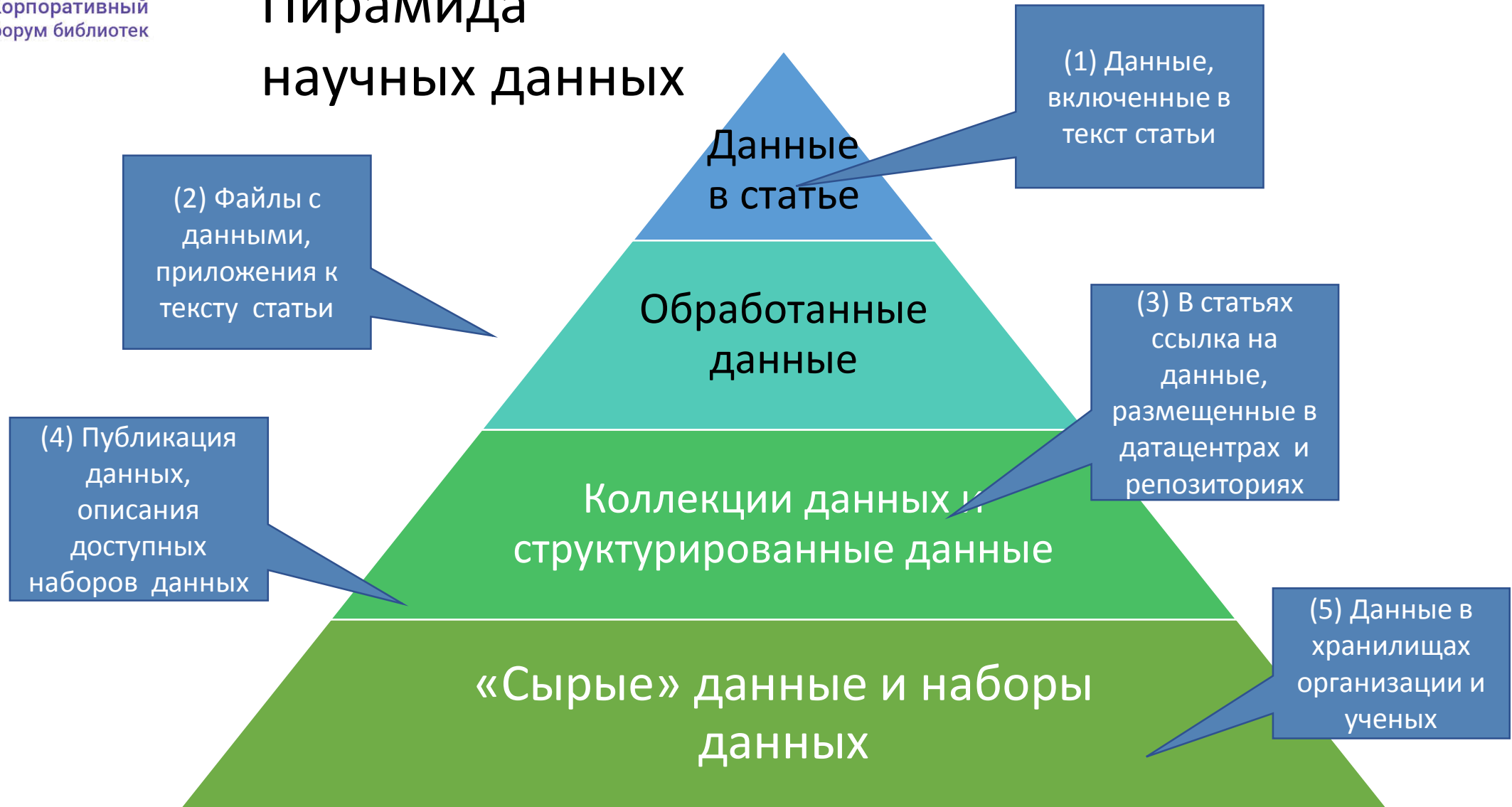
Цель: переосмысление/реорганизация форм и методов представления знаний, результатов исследований для обеспечения быстрого распространения, повторного использования и учета РИД

Методы:

- Использование исследовательских данных наряду с научными статьями
- Повторное использование и цитирование результатов
- Верификация результатов исследований (финансирующей организацией, другими исследователями)
- Создание национальной облачной инфраструктуры науки
- Вхождение в глобальное научно-образовательное пространство

В России: нет практики сохранения, повторного использования и учета исследовательских данных

Пирамида научных данных



Источник: Report on Integration of Data and Publications, 2011

https://www.stm-assoc.org/2011_12_5_ODE_Report_On_Integration_of_Data_and_Publications.pdf

Что сохраняем? Как находим?

**ЧЕМ БОЛЬШЕ ДАННЫХ,
ТЕМ ЛУЧШЕ
????**



ПРОБЛЕМЫ:

- Дублетность
- Разные схемы метаданных
- Разная лингвистика
- Разное качество описаний данных
- Сложность регулирования доступа в гетерогенной среде
- ...

Основные государственные и ведомственные системы:

- Ведомственная система Министерства науки и высшего образования (Система управления НИР и ГЗ wnir.fano.gov.ru)
- Информационные системы РФФИ и РНФ
- Информационная система ФЦП «Исследования и разработки»
- ЕГИСУ НИОКТР.

Системы **не включены в состав открытых государственных данных** и не предусматривает возможность использования данных третьими сторонами.

Серьезным недостатком является **использование проприетарных систем идентификации** организаций и ученых.

Проекты интеграции РФ

- **Единая точка доступа к электронным каталогам библиотек сферы образования и науки (ЭКБСОН), 2011** <http://www.vlibrary.ru/>

*Сейчас в системе собрано **22 млн. записей** из библиотечных каталогов из 600 организаций. Однако использование традиционных подходов для информационного поиска не дает нового качества при анализе собранных данных – можно только получить информацию, в какой библиотеке есть искомый документ.*

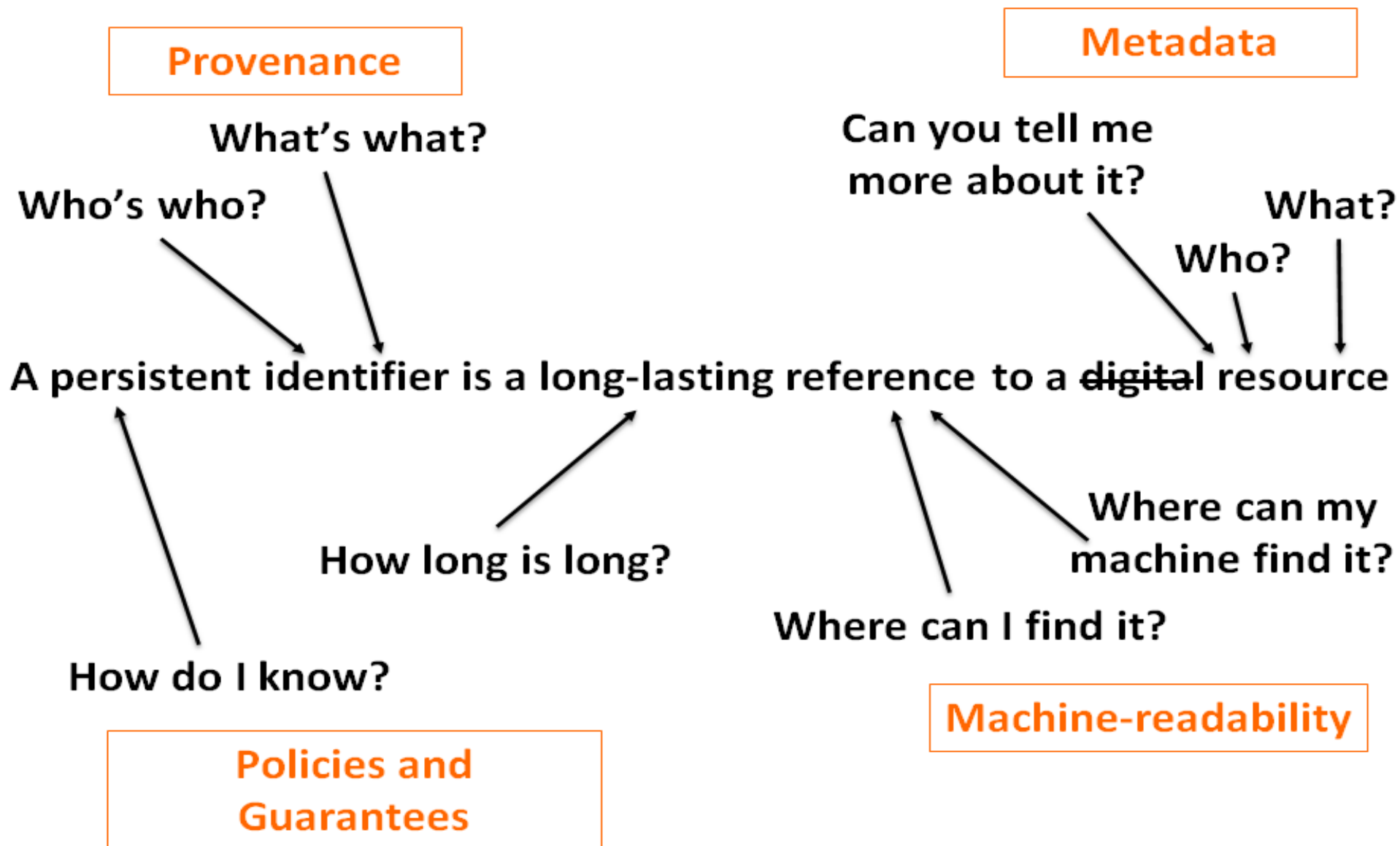
- **Национальный агрегатор открытых репозиторий (НОРА), 2018**
<https://www.openrepository.ru/>

*Собрано около **650 тыс. ресурсов** (некоторые – дублетные) из 20 институциональных репозиторий России и 4 из Белоруссии. Поступления за 2021 год – менее 22 тыс.*

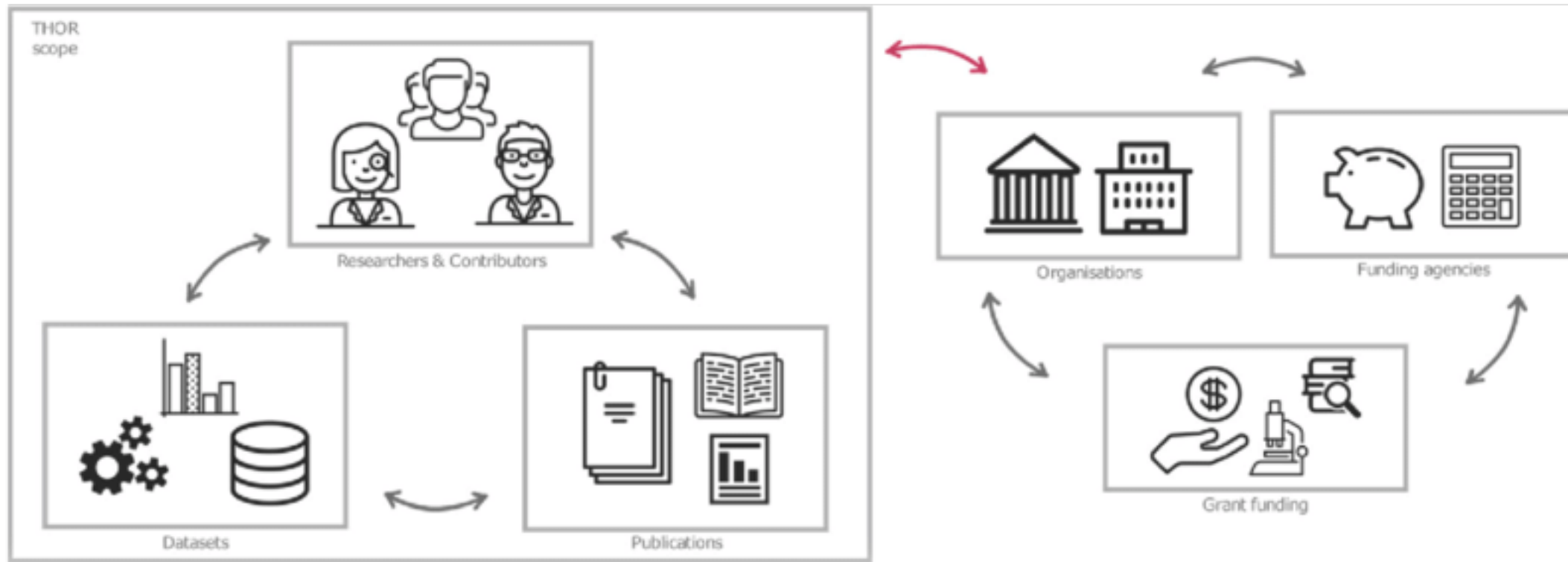
- **Программа «Научно-технологическое развитие РФ», 2020**

Мероприятие 5.2 «Интеграция ресурсов библиотек, архивов, иных организаций, в том числе осуществляющих функции государственных депозитариев, в Национальную электронную библиотеку»

Структурируем, выделяем концепты



Пространство, основанное на постоянных идентификаторах



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

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

Type to search...

This Page
 Works
 People
 Organizations

<https://orcid.org/0000-0003-3484-6875>

K. J. Garza
 Kristian Javier Garza Gutierrez

Links
[github](#)

Other Profiles
[ORCID](#)
[Impactstory](#)
[Europe PMC](#)
[Germany](#)

Other Identifiers
 GitHub: [kjgarza](#)

 <https://orcid.org/0000-0003-3484-6875>

Share

 Email
 Twitter
 Facebook

Employment

DataCite
 Application Developer
 Since August 2016

Aggregated Citations, Views and Downloads

4 Citations
 30 Views

37 Works

Publication Year

☐ 2021	4
☐ 2020	8
☐ 2019	2
☐ 2018	2
☐ 2017	2
☐ 2016	4
☐ 2015	9
☐ 2014	2
☐ 2013	4

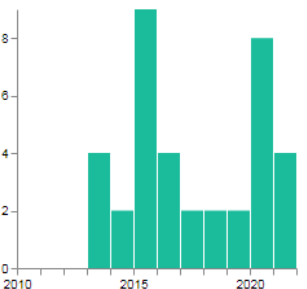
Work Type

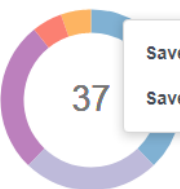
☐ Text	14
☐ Image	10
☐ Software	9
☐ Collection	2
☐ Dataset	2

License

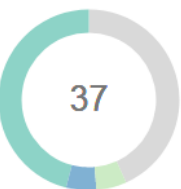
☐ CC-BY-4.0	17
☐ CC-BY-NC-SA-4.0	2
☐ MIT	2

Language

Publication Year


Work Type


Save as SVG
 Save as PNG

License


Individual Motivations Change Activity in Online Scientific Communities
 Kristian Garza
 Poster published 2013 in [figshare Academic Research System](#)
 PGR Symposium
 DOI registered October 23, 2013 via DataCite.

[Image](#)
 <https://doi.org/10.6084/m9.figshare.830404>

seek-analytics
 Kristian Garza
 Software published 2014 in [figshare Academic Research System](#)
 Analytics API for Seek4science
 DOI registered October 22, 2014 via DataCite.

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.

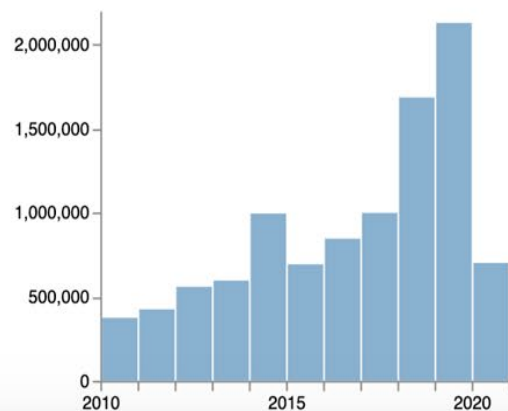
Additional information comes from these data sources:

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

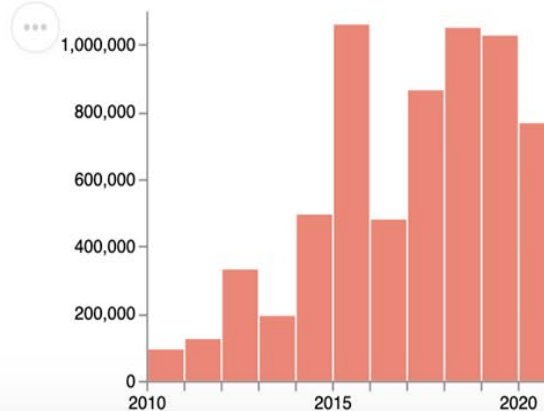
Works

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.

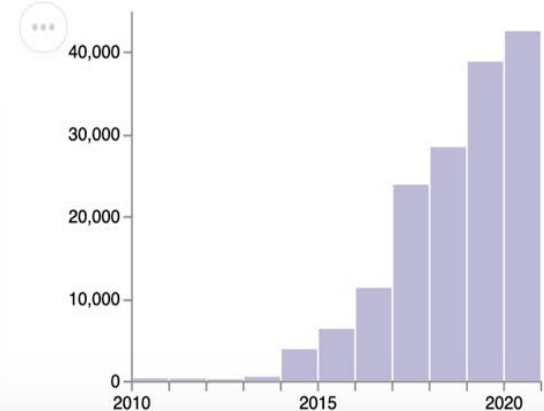
14,412,849 Publications



7,535,822 Datasets



157,363 Software



Views, Downloads and Citations

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



Publication Year

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

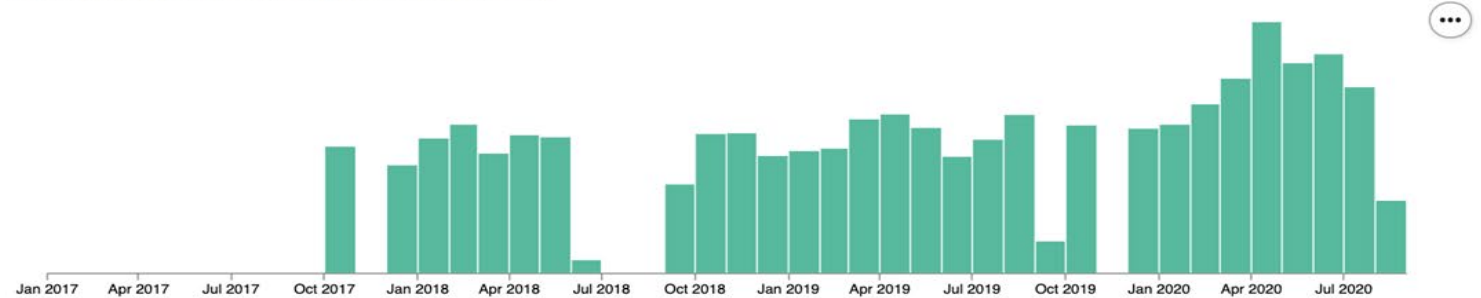
Work Type

☐ Text	41
☐ Dataset	1

23,883 Views

14,652 Views

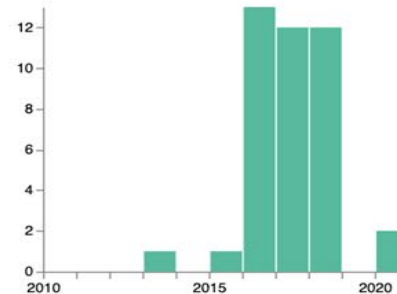
23,883 views reported since publication in 2009.



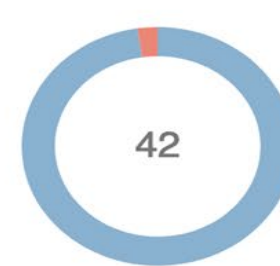
1 Reference

42 Citations

Publication Year



Work Type



License



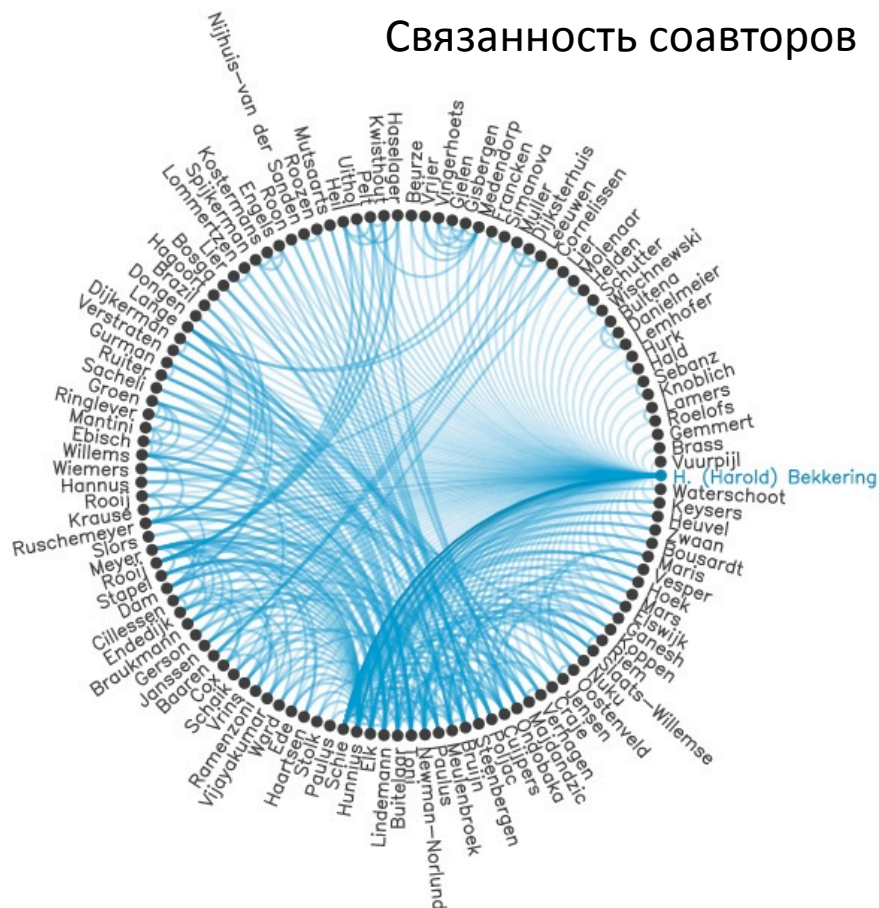
Long-Term Hydraulic Adjustment of Three Tropical Moist Forest Tree Species to Changing Climate

Mahmuda Islam, Mahmuda Islam & Mahmuda Islam

Journal Article published 2018 in [Frontiers in Plant Science](#)

Генерируем карту науки?

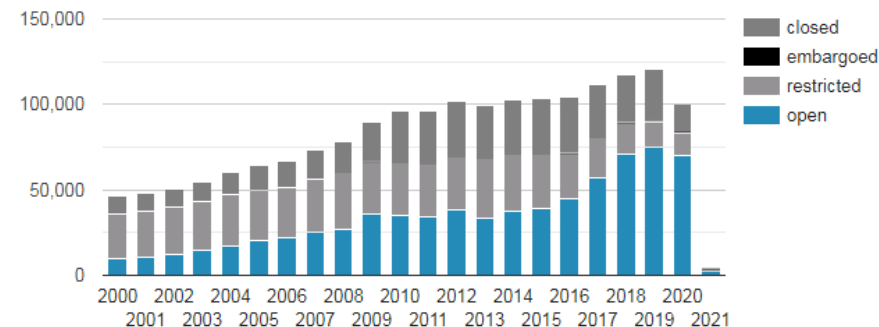
Связанность соавторов



Пример:
NARCIS, национальный
портал научной
информации
Нидерландов

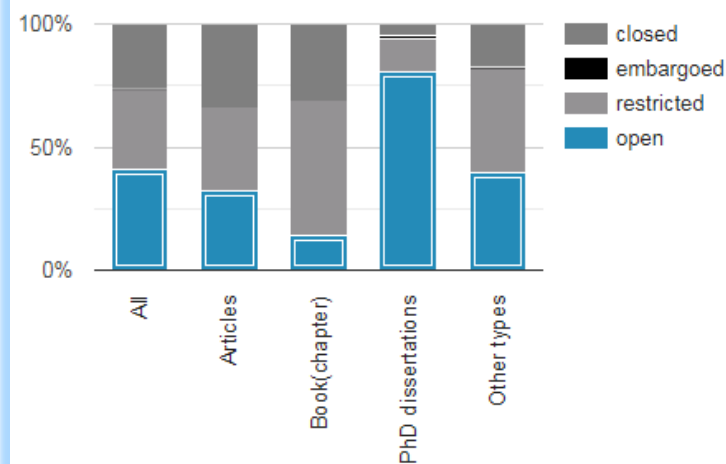
<https://www.narcis.nl/>

OPEN AND CLOSED ACCESS SCHOLARLY PUBLICATIONS IN NARCIS PER YEAR OF PUBLICATION



This chart shows the actual number open, closed, restricted and embargoed publications (articles, doctoral theses, books, reports et cetera) in NARCIS, since 2000 (Live-data). Using the slider it is possible to find the number of publications at a certain moment in time (Going back to September 2019).

OPEN AND CLOSED ACCESS PUBLICATIONS BY PUBLICATION TYPE

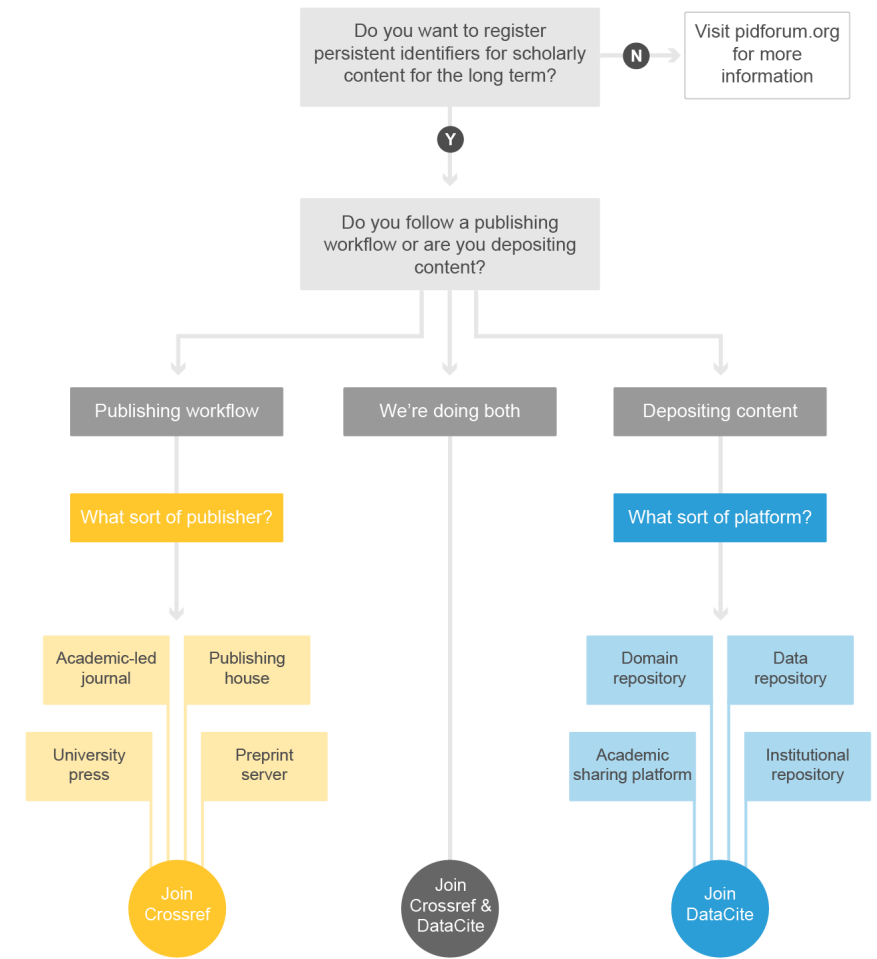


CrossRef и DataCite

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>



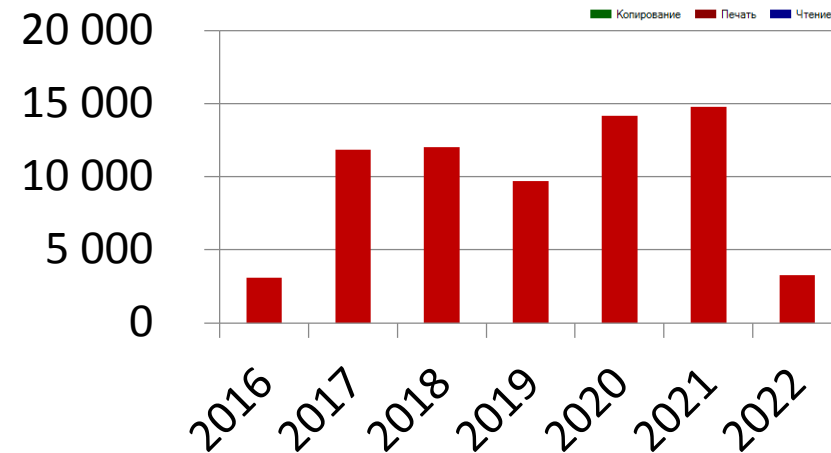
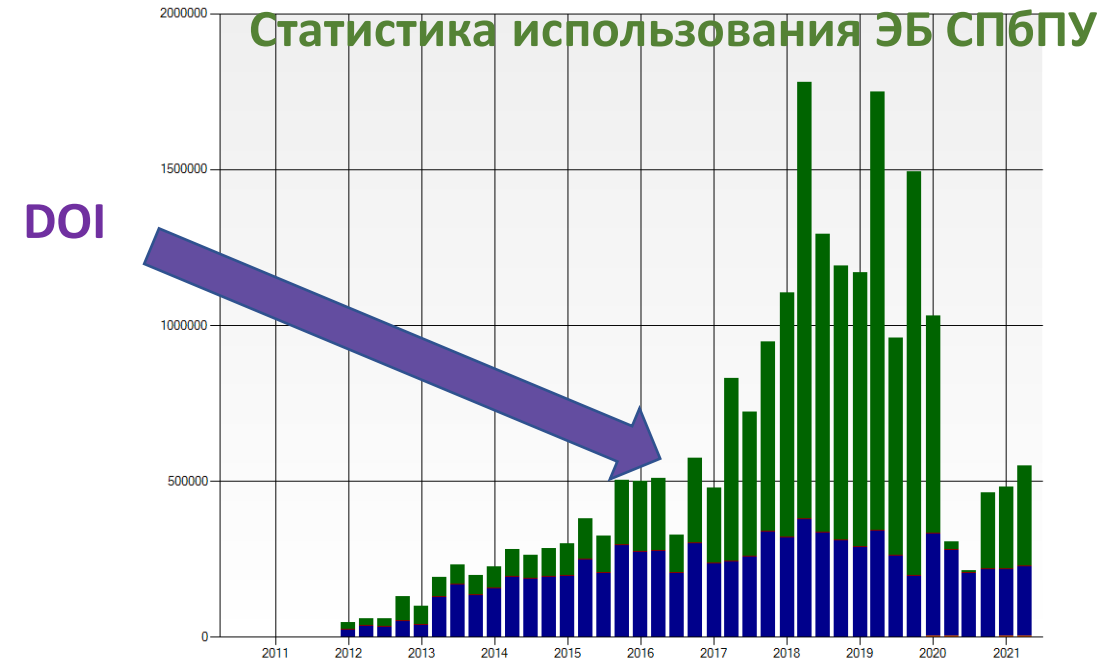
Crossref is a membership organisation for scholarly publishers providing services to help the scholarly community find, cite, link, assess and re-use published content.



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

Национальный центр идентификации научных данных (СПбПУ)

- 2017 - создан Национальный центр идентификации научных данных
- 2020 - в СПбПУ регистрируется около 14 тыс. объектов, всем ресурсам ЭБ СПбПУ назначаются DOI
- 2021 - создан национальный консорциум для интеграции научных инженерных знаний
- 2021 – создан национальный репозиторий научных данных



**Кол-во DOI,
зарегистрированных
Национальным
центром
идентификации
Научных
данных**

Национальный репозиторий научных данных

repository.spbstu.ru

[Начало](#) / [DOI](#) / Создание

DOI:

URL:

Заглавие:

Язык: Тип заглавия:

[Добавить ещё заглавие](#)

Автор:

Тип: ☒ Персона ☐ Организация ☐ Другое

Фамилия: Имя Отчество: Полное имя:

Принадлежность:

[Добавить ещё автора](#)

Издатель:

Год публикации:

Тип ресурса:

[Справочник видов данных DataCite Commons](#)

Рекомендуемые поля

Тема/Предмет:

[Добавить тему](#)

Аннотация / дополнительная:

[Добавить описание](#)

[Рубрикатор OECD](#)

[Начало](#) / [DOI](#) / Создание

DOI:

URL:

Заглавие:

Язык:

[Добавить ещё заглавие](#)

Автор:

Тип: ☒ Персона ☐ Организация ☐ Другое

Фамилия: Имя Отчество: Полное имя:

Принадлежность:

[Добавить ещё автора](#)

Издатель:

Год публикации:

Тип ресурса:

[Справочник видов данных DataCite Commons](#)

Рекомендуемые поля

Тема/Предмет:

[Добавить тему](#)

Аннотация / дополнительная:

[Добавить описание](#)

Тип ресурса

[Найти](#)

- Audiovisual
- Book
- BookChapter
- Collection
- ComputationalNotebook
- ConferencePaper
- ConferenceProceeding
- DataPaper
- Dataset
- Dissertation
- Event
- Image
- InteractiveResource
- Journal
- JournalArticle
- Model
- OutputManagementPlan
- PeerReview
- PhysicalObject
- Preprint
- Report
- Service
- Software
- Sound
- Standard

Предложение – новый подход: инфраструктура **связанных** научных данных

- Размещение в **отечественных репозиториях** научных данных результатов исследований, выполненных за средства госбюджета и научных фондов
- Создание **национальной спецификации** описания научных данных на основе используемого в мире формата, совместимой с глобальными системами, регистрирующими постоянные идентификаторы
- Создание **сети FAIR репозиториев научных данных**, как институциональных, так и тематических
- Использование глобальных **реестров и схем метаданных** для создания национальных научных данных и их описаний высокого качества
- Система рекомендаций/требований **по разграничению доступа** к научным данным, созданных на средства госбюджета и при финансировании российских научных фондов