

# **Що треба знати, перш ніж поширювати дані?**

Тетяна Борисова  
9 жовтня 2017 р., м. Київ, НаУКМА

# Які дані поширюють

Не всі дані, що створюються під час дослідження, потребують поширення.

Можливе поширення як неструктурованих (первинних даних), так і структурованих.

Дані можуть бути як відкритого, так і закритого чи обмеженого доступу.

# RESEARCHER DATA SHARING INSIGHTS

WILEY

- Wiley's Researcher Data Insights Survey was launched earlier this year to understand how and why researchers make their research data publicly available. The study's results, highlighted below, are intended to advance the global conversation about data sharing and help Wiley better meet the needs of our researchers, authors, and partners in the rapidly evolving landscape of scientific research and communications.
- The survey was deployed in March 2014 and received more than 2,250 responses from researchers around the world.

## GLOBAL DATA SHARING TRENDS

Data sharing practices vary widely across research fields and geographic areas. Just over half of researchers report making their data publicly available, though archiving results in repositories is not yet the norm.

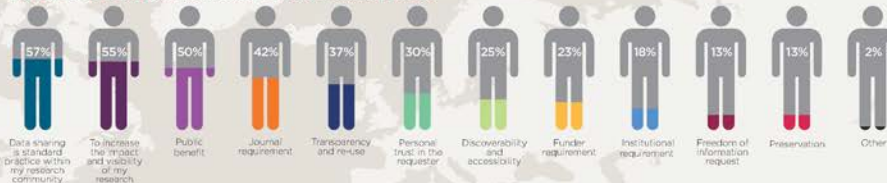


### WAYS DATA IS SHARED

- 67% As supplementary material in a journal
- 37% Personal, institutional or project webpage
- 26% Institutional data repository (i.e. university or institute-sponsored)
- 19% Discipline-specific data repository
- 6% General-purpose data repository (e.g. Dryad, Figshare)
- 5% Other

Globally, researchers also report sharing their data in limited and non-permanent ways: 57% are sharing data at a conference while 42% of researchers share their data upon informal request (e.g. email, direct contact, etc.).

## RESEARCHER MOTIVATIONS FOR SHARING DATA



## DATA SHARING TRENDS BY COUNTRY



## REASONS WHY RESEARCHERS ARE HESITANT TO SHARE THEIR DATA

- 42% Intellectual property or confidentiality issues
- 36% My funder/institution does not require data sharing
- 26% I am concerned that my research will be scooped
- 26% I am concerned about misinterpretation or misuse
- 23% Ethical concerns
- 22% I am concerned about being given proper citation credit or attribution
- 21% I did not know where to share my data
- 20% Insufficient time and/or resources
- 16% I did not know how to share my data
- 12% I don't think it is my responsibility
- 12% I did not consider the data to be relevant
- 11% Lack of funding
- 7% Other

## DATA SHARING BY DISCIPLINE

Data sharing, specifically by way of data repositories, is most prevalent amongst life scientists, particularly those in the earth and environmental and agriculture and food sciences.



A typical *Health Science* researcher says she would be motivated to share her data in the future in order to benefit the public, so long as privacy and ethical concerns are assured.



A typical *Life Science* researcher says she would be motivated to share more of her data in the future if she was guaranteed proper credit.



A typical *Physical Science* researcher says she would be motivated to share her data in the future because it is standard practice within her research community and because it increases the impact and visibility of her work.



A typical *Social Science and Humanities* researcher says she would be motivated to share her data in the future if it increased the impact and visibility of her work or if she was required to by her funder.

# Мотивація поширювати дані

57% - поширення даних стандартна практика в дослідницькому середовищі.

55% - збільшити видимість та внесок від дослідження

50% - публічні переваги

42% - вимоги журналів

37% - прозорість та повторне використання

30% - персональна довіра до того, хто запитує дані



# Мотивація поширювати дані

25% - доступність та пошук

23% - вимоги грантодавців

18% - вимоги інституції

13% - зберігання

2% - інше



# Де поширюють дані (за дисциплінами)

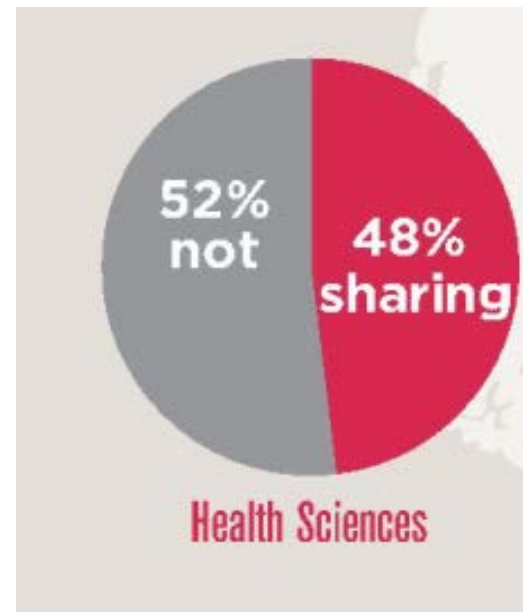
68% - як додаткові матеріали у журналі

29% - персональний чи інституційний сайт, сайт лабораторії

29% - інституційний репозитарій даних

21% - тематичний репозитарій даних

5% - загальний репозитарій даних (Dryad, figshare)



# Де поширюють дані (за дисциплінами)

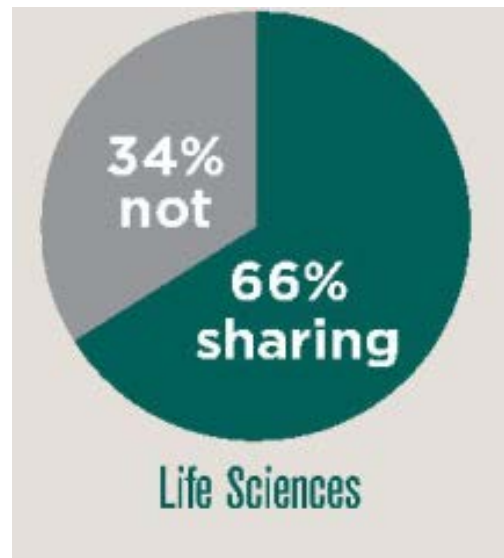
76% - як додаткові матеріали у журналі

42% - тематичний репозитарій даних

29% - персональний чи інституційний сайт, сайт лабораторії

23% - інституційний репозитарій даних

13% - загальний репозитарій даних (Dryad, figshare)



# Де поширюють дані (за дисциплінами)

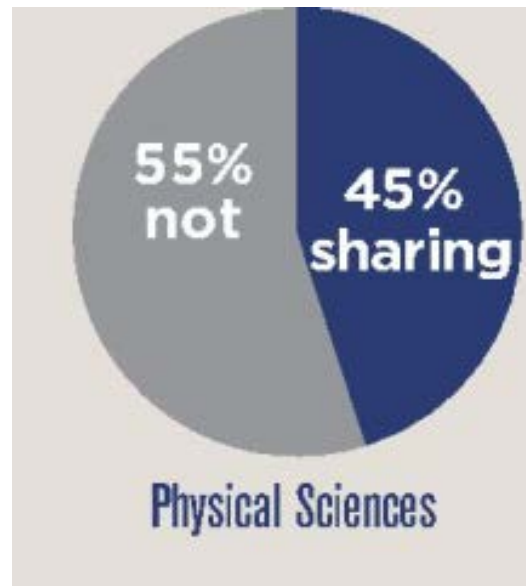
69% - як додаткові матеріали у журналі

41% - персональний чи інституційний сайт, сайт лабораторії

28% - інституційний репозитарій даних

10% - тематичний репозитарій даних

3% - загальний репозитарій даних (Dryad, figshare)





# Де поширюють дані (за дисциплінами)

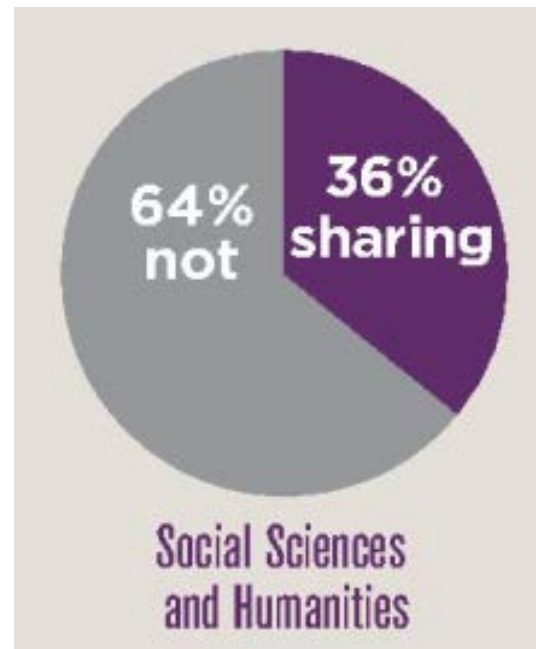
52% - як додаткові матеріали у журналі

51% - персональний чи інституційний сайт, сайт лабораторії

25% - інституційний репозитарій даних

3% - загальний репозитарій даних (Dryad, figshare)

2% - тематичний репозитарій даних



# Data паразити



The NEW ENGLAND  
JOURNAL of MEDICINE

HOME ARTICLES & MULTIMEDIA ▾ ISSUES ▾ SPECIALTIES & TOPICS ▾ FOR AUTHORS ▾ CME ▾

## EDITORIAL

### Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.

N Engl J Med 2016; 374:276-277 | January 21, 2016 | DOI: 10.1056/NEJMe1516564

Share:     

Article References Citing Articles (79) Letters Metrics

The aerial view of the concept of data sharing is beautiful. What could be better than having high-quality information carefully reexamined for the possibility that new nuggets of useful data are lying there, previously unseen? The potential for leveraging existing results for even more benefit pays appropriate increased tribute to the patients who put themselves at risk to generate the data. The moral imperative to honor their collective sacrifice is the trump card that takes this trick.

However, many of us who have actually conducted clinical research, managed clinical studies and data collection and analysis, and curated data sets have concerns about the details. The first concern is that someone not involved in the generation and collection of the data may not understand the choices made in defining the parameters. Special problems arise if data are to be combined from independent studies and considered comparable. How heterogeneous were the study populations? Were the eligibility criteria the same? Can it be assumed that the differences in study populations, data collection and analysis, and treatments, both protocol-specified and unspecified, can be ignored?

Перестороги, під час поширення даних:  
- хтось, хто не займається створенням та збором даних, може не зрозуміти вибору, зробленого при визначенні параметрів;  
- з'явиться новий клас науковців - люди, які не мали нічого спільного з виконанням дослідження, але використовують дані інших груп для своїх цілей, можливо, крадуть продукти досліджень або навіть використовують дані, щоб спробувати спростувати те, що встановило оригінальне дослідження («research parasites»).

<http://www.nejm.org/doi/full/10.1056/NEJMe1516564#t=article>



Твіти **545** Читас **1** Читачі **1 259** Вподобання **4 460**

## Research Parasite

@dataparasite

Reanalyzing your data. Disproving what you posited. Stealing ideas you haven't yet had.

📍 In your data

🌐 [for.researchparasites.com](http://for.researchparasites.com)

📅 Дата приєднання: січень 2016

Твітнути

Повідомле...

🖼️ 31 фото чи відео



Твіти **Твіти й відповіді** Медіафайли

📌 Закріплений твіт



**Research Parasite** @dataparasite · 4 лют. 2016



🗨️ 74 ❤️ 88 ✉️

<https://twitter.com/dataparasite>



[MISSION](#) [APPLY](#) [PRIZE & SUPPORTERS](#) [AWARD RECIPIENTS](#) [COI RULES](#)

# THE PARASITE AWARDS

*Celebrating rigorous secondary data analysis*

 [Tweet](#)

<http://researchparasite.com/#prize>

# AWARD RECIPIENTS

*Exemplars of research parasitism.*



**Kun-Hsing Yu**  
2017 Junior Parasite



**Erick Turner**  
2017 Sustained Parasitism

<http://researchparasite.com/#prize>

# Способи поширення даних

- Завантажувати як додаткові матеріали до публікації в журналі (якщо журнал має політики щодо поширення даних);
- Депонувати в репозитарії даних (тематичні, інституційні чи мультидисциплінарні);
- Розміщувати на сайтах (персональних, проекту, лабораторії);
- Публікувати у журналах для даних.
- Неофіційне поширення (персонально, на запит інших дослідників)

# Data journals

[SCIENTIFIC DATA JOURNAL](#)

[DATA SCIENCE JOURNAL](#)

[DATASET PAPERS IN SCIENCE](#)

[JOURNAL OF OPEN HUMANITIES DATA](#)

[JOURNAL OF OPEN PSYCHOLOGY DATA](#)

[JOURNAL OF OPEN ARCHAEOLOGY DATA](#)

[GEOSCIENCE DATA JOURNAL](#)

[EARTH SYSTEM SCIENCE DATA \(ESSD\)](#)

[JOURNAL OF CHEMICAL AND ENGINEERING DATA](#)

[THE INTERNATIONAL JOURNAL OF ROBOTICS RESEARCH - DATA PAPERS](#)

[JOURNAL OF PHYSICAL AND CHEMICAL RESEARCH DATA](#)

[ECOLOGICAL ARCHIVES](#)

[IUCRDATA](#)

[BMC RESEARCH NOTES](#)

[GIGASCIENCE](#)

[JOURNAL OF OPEN PUBLIC HEALTH DATA](#)



# Journal of Open Humanities

**Journal of Open Humanities Data**

Start Submission Become a Reviewer

Reading: Writing About Shakespeare: 1960–2010 Share: f t g+ in Download A

## Data Papers

### Writing About Shakespeare: 1960–2010

**Authors:** [Laura Estill](#) ✉, [Dominic Klyve](#)

#### Abstract

This dataset quantifies writing (in number of articles, books, dissertations, and monographs) about each of Shakespeare's plays during each year in the period 1960–2010. The information was extracted from the *World Shakespeare Bibliography* ([www.worldshakesbib.org](http://www.worldshakesbib.org)).

**Keywords:** literature, statistical data, scholarly publishing, academic publishing, digital humanities, Shakespeare

**How to Cite:** Estill, L. & Klyve, D., (2016). Writing About Shakespeare: 1960–2010. *Journal of Open Humanities Data*. 2, p.e3. DOI: <http://doi.org/10.5334/johd.4>

197 Views	21 Downloads	7 Twitter
--------------	-----------------	--------------

Published on 25 Nov 2016 Peer Reviewed CC BY 4.0

JUMP TO

DISCUSSIONS

Abstract

(1) Overview

(2) Methods

(3) Dataset description

(4) Reuse potential

Notes

Additional Files

Acknowledgements

Competing Interests


References

<https://openhumanitiesdata.metajnl.com/articles/10.5334/johd.4/>



# Scientific Data

A global reference database of crowdsourced cropland data collected using the Geo-Wiki platform

Juan Carlos Laso Bayas , Myroslava Lesiv [...] John W Wilson

*Scientific Data* **4**, Article number: 170136

(2017)

doi:10.1038/sdata.2017.136

Download Citation

Agriculture Environmental impact

Geography Sustainability

Received: 18 April 2017

Accepted: 02 August 2017

Published online: 26 September 2017

## Abstract

A global reference data set on cropland was collected through a crowdsourcing campaign using the Geo-Wiki crowdsourcing tool. The campaign lasted three weeks, with over 80 participants from around the world reviewing almost 36,000 sample units, focussing on cropland identification. For quality assessment purposes, two additional data sets are provided. The first is a control set of 1,793 sample locations validated by students trained in satellite image interpretation. This data set was used to assess the quality of the crowd as the campaign progressed. The second data set contains 60 expert validations for additional evaluation of the quality of the contributions. All data sets are

## Associated Content

*Remote Sensing* | Article

[Integrating Crowdsourced Data with a Land Cover Product: A Bayesian Data Fusion Approach](#)

Patrick Bogaert & Sarah Gengler

*Global Change Biology* | Article

[Mapping global cropland and field size](#)

Adeaga Olusegun, Adriana Gomez [...] Aline Mosnier

Sections Figures References

Abstract

Background & Summary

Methods

Data Records

Technical Validation

Usage Notes

Additional Information

References

Data Citations

Acknowledgements

Author information

Supplementary information

<https://www.nature.com/articles/sdata2017136>

# Re3data

Глобальний реєстр репозитаріїв даних досліджень з усіх дисциплін.

Запущений у травні 2013 року. На сьогодні надає інформацію понад 1200 репозитаріїв.





The research data repository provides additional information on its service.



The research data repository was reviewed by the re3data.org Team.



The research data repository is either certified or supports a repository standard.



The research data repository provides a policy.



The research data repository provides open/restricted/closed access to its data.



The terms of use and licenses of the data are provided by the research data repository.



The research data repository uses a persistent identifier system to make its provided data persistent, unique and citable.

# На що звернути увагу, обираючи сховище для розміщення даних

Які формати підтримує, чи змінює якість даних під час розміщення?

Які метадані присвоює?

Чи є політики авторства, чи підтримує ліцензії?

Чи робить бекапи і як регулярно?

Чи присвоює ідентифікатор?

Чи рецензується?

Які можливості доступу передбачає?



# Склад Re3data (за темою)

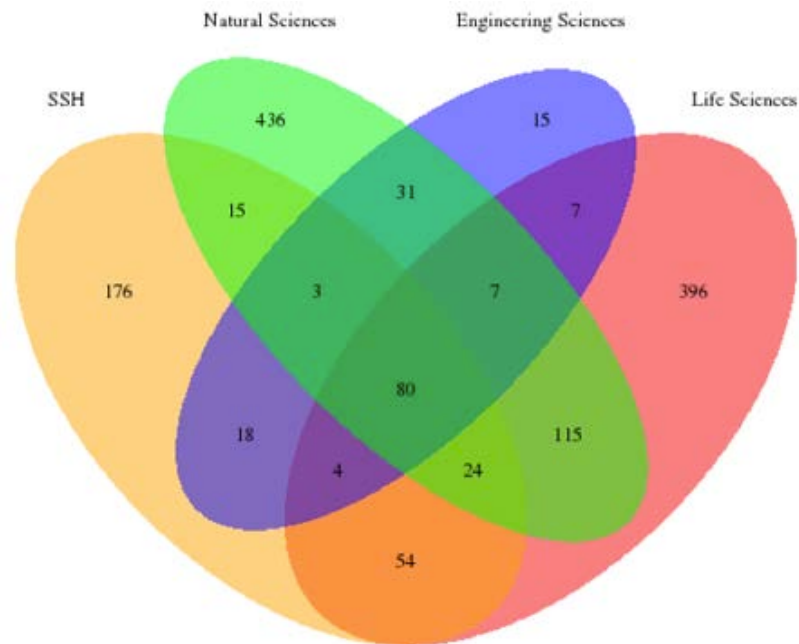
4 основні предметні категорії, на які розподіляються репозитарії в реєстрі:

Natural Sciences (51.5 %)

Life Sciences (49.8 %)

Humanities and Social Sciences (27.1 %)

Engineering Sciences (12.0 %)



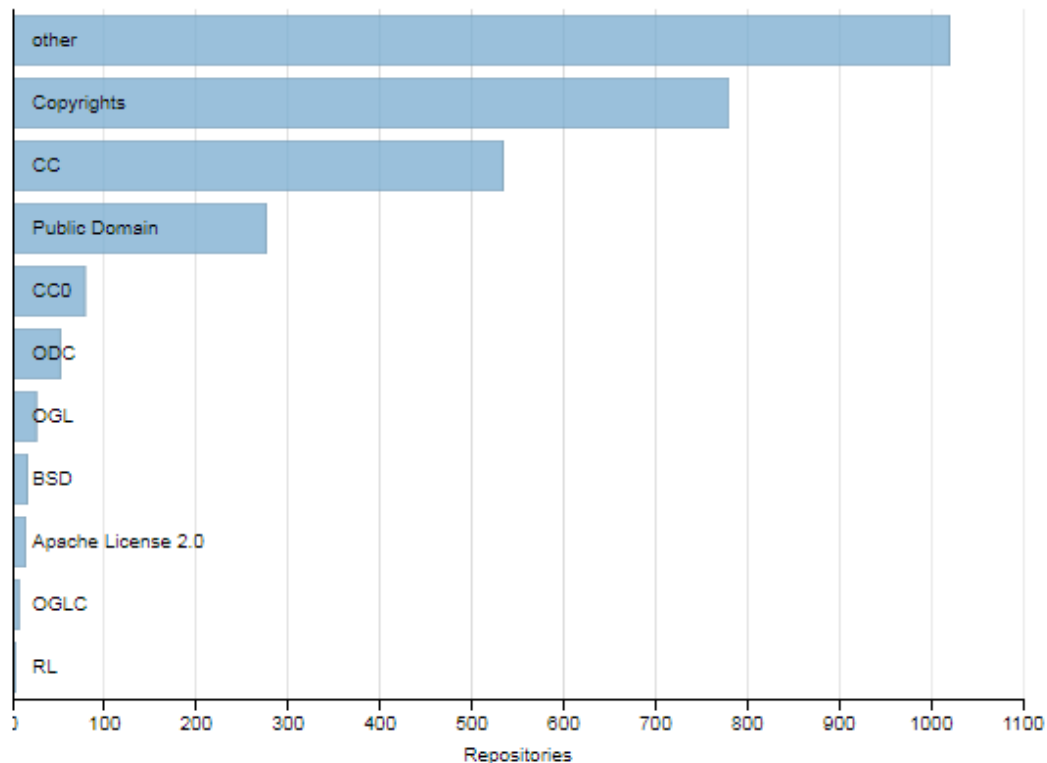
The Landscape of Research Data Repositories in 2015: A re3data Analysis -

<http://www.dlib.org/dlib/march17/kindling/03kindling.html>

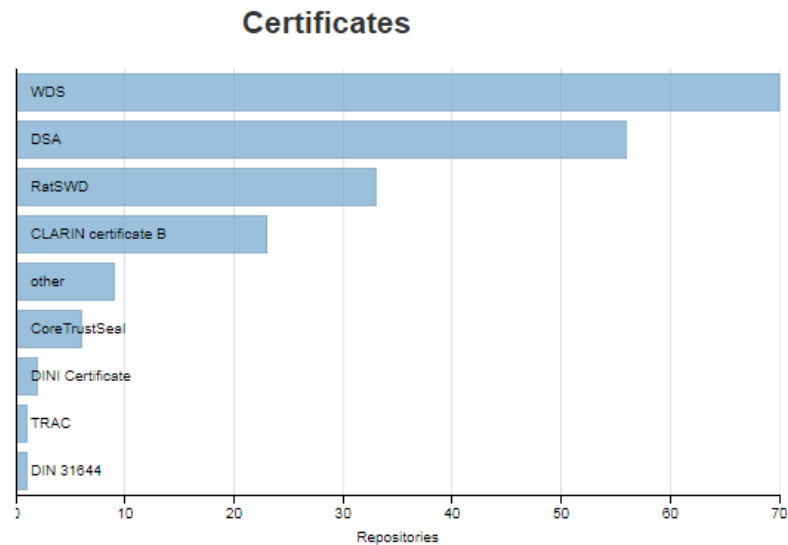
# Ліцензії

Авторське право  
Creative Commons  
Public Domain  
CC0  
Open Data Commons  
Open Game License  
BSD  
Apache License 2.0  
Open General License

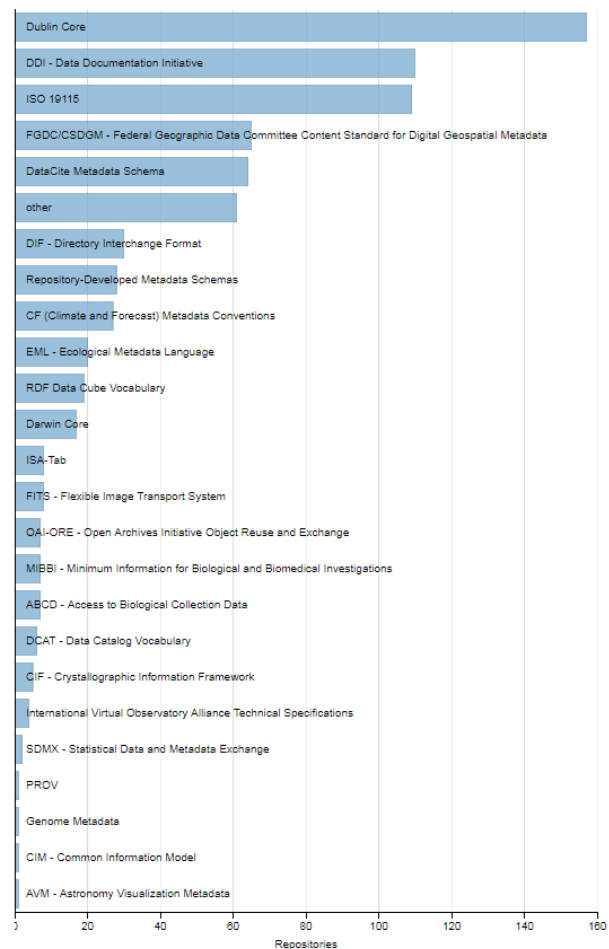
## Data licenses



# Сертифікати



# Стандарти метаданих





# Мультидисциплінарні репозитарії



# Дякую за увагу!

**Тетяна Борисова,**

Завідувачка відділу електронних ресурсів

Наукова бібліотека Національного університету

«Києво-Могилянська академія»

[borysovato@ukma.edu.ua](mailto:borysovato@ukma.edu.ua)

Семінар «Цифрова підтримка наукових досліджень»

9 жовтня 2017 р., м. Київ

Національний університет «Києво-Могилянська академія»