

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

Research Collection Library

SMU Libraries

---

10-2019

### The rise of open scholarly data & possible implications for libraries

Aaron TAY

Singapore Management University, aarontay@smu.edu.sg

Follow this and additional works at: [https://ink.library.smu.edu.sg/library\\_research](https://ink.library.smu.edu.sg/library_research)



Part of the [Scholarly Communication Commons](#)

---

#### Citation

TAY, Aaron. The rise of open scholarly data & possible implications for libraries. (2019). *9th iGroup China Librarian Conference 2019, October 28-31*. 1-23.

Available at: [https://ink.library.smu.edu.sg/library\\_research/150](https://ink.library.smu.edu.sg/library_research/150)

This Presentation is brought to you for free and open access by the SMU Libraries at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Library by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email [cherylds@smu.edu.sg](mailto:cherylds@smu.edu.sg).

# THE RISE OF OPEN SCHOLARLY DATA & POSSIBLE IMPLICATIONS

9<sup>TH</sup> IGCLC 2019, October 28-31, Wuhan

开放学术数据的兴起及其潜在影响

AARON TAY

LIBRARY ANALYTICS MANAGER

图书馆分析部经理

SINGAPORE MANAGEMENT UNIVERSITY

新加坡管理大学

AARONTAY@GMAIL.COM

# THE FOCUS ON OPEN ACCESS 开放访问的焦点

## **The Future of OA: A large-scale analysis projecting Open Access publication and readership**

 Heather Piwowar,  Jason Priem,  Richard Orr

doi: <https://doi.org/10.1101/795310>

**Given existing trends, we estimate that by 2025:**

- **44% of all journal articles will be available as OA**
- **70% of all article views will be to OA articles**

就现在的趋势看，我们预测在2025年来临前，44%的期刊文章将被开放发表、70%的访问量将来自开放访问文章

[//www.biorxiv.org/content/10.1101/795310v1](https://www.biorxiv.org/content/10.1101/795310v1)

# QUIET REVOLUTION IN SCHOLARLY METADATA

## 学术文章的元数据正在悄悄改变

Scopus Search Sources Alerts Lists Help SciVal Create account Sign in

### Document details

< Back to results | < Previous 2 of 2

Export Download Print Find it @ SMU Libraries View at Publisher

unpaywall Crossref

BMC Women's Health Open Access Volume 17, Issue 1, 10 October 2017, Article number 98

#### Stress, health and quality of life of female migrant domestic workers in Singapore: A cross-sectional study (Article) (Open Access)

Anjara, S.G.<sup>1</sup> Nellums, L.B.<sup>2</sup> Bonetto, C.<sup>3</sup> Van Bortel, T.<sup>3,4</sup>

<sup>1</sup>University of Cambridge School of Clinical Medicine, Cambridge Institute of Public Health, Box 113 Cambridge Biomedical, Forvie Site, Robinson Way, Cambridge, CB2 0SR, United Kingdom  
<sup>2</sup>Imperial College London, Section of Infectious Diseases and Immunity, Department of Medicine, Commonwealth Building, Hammersmith Hospital, Du Cane Road, London, W12 0NN, United Kingdom  
<sup>3</sup>University of Verona, Department of Neurosciences, Biomedicine and Movement Sciences, Section of Psychiatry, P.le L.A. Scuro 10, Verona, 37134, Italy

View additional affiliations

#### Abstract

Background: There is a global increase in migrant workers. In Singapore, there are over 230,000 migrant domestic workers (MDWs). Female MDWs may experience high levels of stress and social isolation, which may negatively impact on their health and quality of life. There have also been documented cases of abuse and exploitation. However, there is a lack of empirical research with this population. This study aimed to investigate factors impacting on the health and quality of life of female MDWs in Singapore, including socio-demographic and job related characteristics, stress, social isolation, and working management style. Methods: A cross-sectional survey was carried out with 182 female MDWs in Singapore. The survey examined health and quality of life (WHOQoL-Bref), social connectedness (the Friendship Scale), and preferred and experienced working management style (the Theory X and Theory Y Questionnaire). Descriptive analyses were carried out in addition to ANOVA, t-tests, and chi-square tests, followed by a multivariate analysis using linear regression. Results: Participants were found to have good overall quality of life and satisfaction with health. Age and working experience were found to be significantly ( $p < 0.05$ ) associated with overall quality of life and three domains (psychological, social, and environmental health). Agreement between experienced and preferred working management style was also found to be associated with higher quality of life scores (with the exception of the social relationships domain). Though women reported relatively good overall quality of life, more than half of participants reported feeling stressed. In addition, nearly 20% of participants reported being isolated or very isolated. Stress was identified to be associated with isolation. In the multivariate analysis, stress was found to contribute to worse quality of life in all domains except social relationships, after adjusting for confounders. Social connectedness was positively associated with all domains of quality of life, and agreement of working management style was positively associated with physical health, psychological health and environmental quality of life. Conclusions: The findings serve as an evidence-base pointing to the need for policies aimed at decreasing stress and social isolation among female MDWs in order to improve their health and quality of life. © 2017 The Author(s).

View references (81)

#### SciVal Topic Prominence

Topic: Loneliness | Social Isolation | Feeling lonely

Prominence percentile: 98.137

#### Indexed keywords

EMTREE medical terms:

adult analysis of variance chi square test cross-sectional study empirical research environmental health female friendship human human experiment linear regression analysis management style migrant multivariate analysis psychological well-being quality of life questionnaire satisfaction Singapore social interaction social isolation Student t test worker cross-sectional study housekeeping male mental health middle aged migration psychology quality of life Singapore standards statistics and numerical data work schedule workload

MeSH:

Adult Cross-Sectional Studies Female Housekeeping Humans Male Mental Health Middle Aged Quality of Life Singapore Transients and Migrants Work Schedule Tolerance Workload

#### Metrics

8 Citations in Scopus

2.12 Field-Weighted Citation Index

PlumX Metrics Usage, Captures, Mentions, Social Media and Citations beyond Scopus

paperbuzz

#### Cited by 8 documents

Occupational health outcomes among international migrant workers: a systematic review and meta-analysis  
Hargreaves, S., Rustage, K., Nellums, L.B. (2019) *The Lancet Global Health*

Family networks during migration and risk of non-affective psychosis: A population-based cohort study  
Dykoohorn, J., Hollander, A.-C., Lewis, G. (2019) *Schizophrenia Research*

Perceived stressors and coping mechanisms of female migrant domestic workers in Singapore  
Van Bortel, T., Martin, S., Anjara, S. (2019) *PLoS ONE*

View all 8 citing documents

Inform me when this document is cited in Scopus:

Set citation alert Set citation feed

#### Related research data

Stress, health and quality of life of female migrant domestic workers in Singapore: a cross-sectional study  
Springer Nature

Stress, health and quality of life of female migrant domestic workers in Singapore: a cross-sectional study  
Springer Nature

Stress, health and quality of life of female migrant domestic workers in Singapore: a cross-sectional study  
Springer Nature

DataCite FIND, ACCESS, AND REUSE DATA

# QUIET REVOLUTION IN SCHOLARLY METADATA

## 学术文章的元数据正在悄悄改变

1. **Item basic metadata - title/abstract/author/year of pub etc e.g. Crossref metadata search, Datacite metadata search**

基本元数据——文章标题、摘要、作者、发表年份，可通过Crossref或Datacite进行元数据检索。

2. **References of article e.g. COCI (<http://opencitations.net/index/coci>), OCC - (<http://opencitations.net/corpus>), Microsoft Academic Graph (<https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/>)**

参考书目可通过COCI、OCC或微软学术Graph（引用关系图）检索。

3. **Metrics & Altmetrics of article - number of times tweeted, cited in Wikipedia, patents etc - e.g. Crossref Events Data API (<https://www.crossref.org/services/event-data/>)**

文章的主要指数和可替代指数——被推特转发的次数、被维基百科或专利引用的次数，可通过Crossref的Events Data API（事件数据插件）计算。

4. **OA availability status - e.g. Unpaywall (<http://unpaywall.org>)**

开放获取的状态，可通过Unpaywall网站查询。

5. **Full text for Text data mining and research dataset e.g. CORE (<https://www.jisc.ac.uk/core>)**

用于数据挖掘和科研数据集的全文，可通过CORE查询。

# PLAYERS IN THE "OPEN" SCHOLARLY DATA ECOSYSTEM

## “开放”学术数据生态圈的参与者



<https://www.crossref.org/>



<https://datacite.org/>



<https://www.ncbi.nlm.nih.gov/pubmed/>



<https://orcid.org/>



<https://ror.org/>

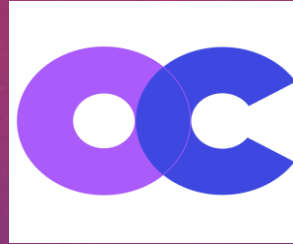
# PLAYERS IN THE "OPEN" SCHOLARLY DATA ECOSYSTEM

## “开放”学术数据生态圈的参与者



I4OC

<https://i4oc.org/>



<https://opencitations.net/index>



<https://www.wikidata.org>



<https://core.ac.uk/>



<https://www.base-search.net>



<https://www.microsoft.com/en-us/research/project/academic/>



iCite

<http://unpaywall.org/>



unpaywall

<http://unpaywall.org/>



paperbuzz

<https://paperbuzz.org/>

# MICROSOFT (微软) ACADEMIC

Microsoft Academic ” [Sign up / Sign in](#)

## Research more, search less

Search any topic, author, journal, etc. or any combination of these

- 229,848,459 Papers
- 232,731,853 Authors
- 664,910 Topics
- 4,418 Conferences
- 48,841 Journals
- 25,579 Institutions

**Top Authors in Mathematics**  
Items are sorted by saliency. [?](#)  
1. Emmanuel J. Candès

**Top Institutions in Mathematics**  
Items are sorted by saliency. [?](#)  
1. Stanford University

Unleash the Power of Semantic Search

<https://academic.microsoft.com/home>

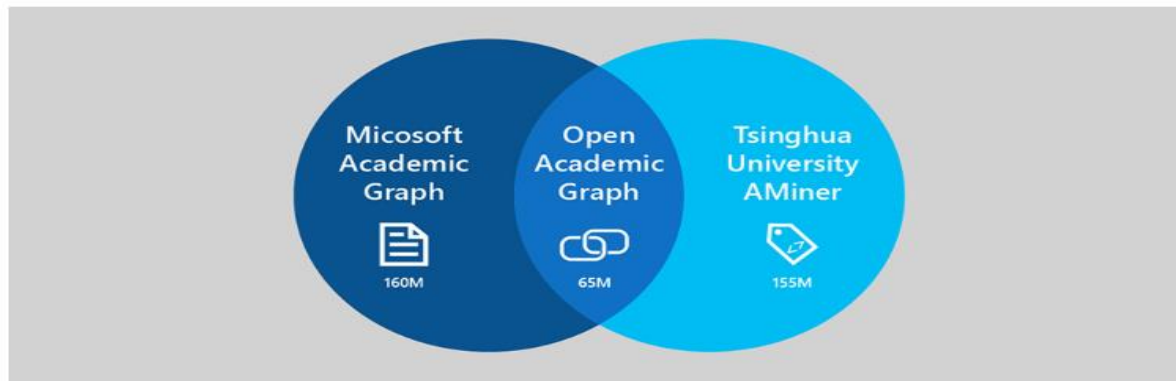


# MICROSOFT ACADEMIC PARTNERSHIP WITH TSINGHUA UNIVERSITY

## 微软学术与清华大学的合作

### Microsoft and Tsinghua University Work Together on Open Academic Data Research

March 22, 2018 | By Microsoft blog editor



In a recent collaboration, Microsoft and China's Tsinghua University released an academic graph, named **Open Academic Graph (OAG)**. This billion-scale academic graph integrates the current **Microsoft Academic Graph (MAG)** and Tsinghua's **AMiner** academic graph. Specifically, it contains the metadata information of 155 million academic paper metadata from AMiner and 166 million papers from MAG. By consolidating metadata information of each, it generates nearly 65 million matching relationships between the two academic graphs [1].

微软和清华大学  
共同建立一个开放学术数  
据研究的工具  
Open Academic  
Graph

<https://www.microsoft.com/en-us/research/blog/microsoft-tsinghua-university-work-together-open-academic-data-research/>

# WHAT CAN YOU DO IN A WORLD WHERE ALL THIS SCHOLARLY INFO IS OPEN AND AVAILABLE?

Title/author/abstract  
标题/作者/摘要

Affiliations  
作者归属机构

References  
(relationships!)  
参考目录（引用关系）

Altmetrics  
各种评价指数

OA Status  
开放获取状态

Full text?  
全文？

# PID GRAPHS – CREATING RELATIONSHIPS

## 永久识别号关系图——建立关系

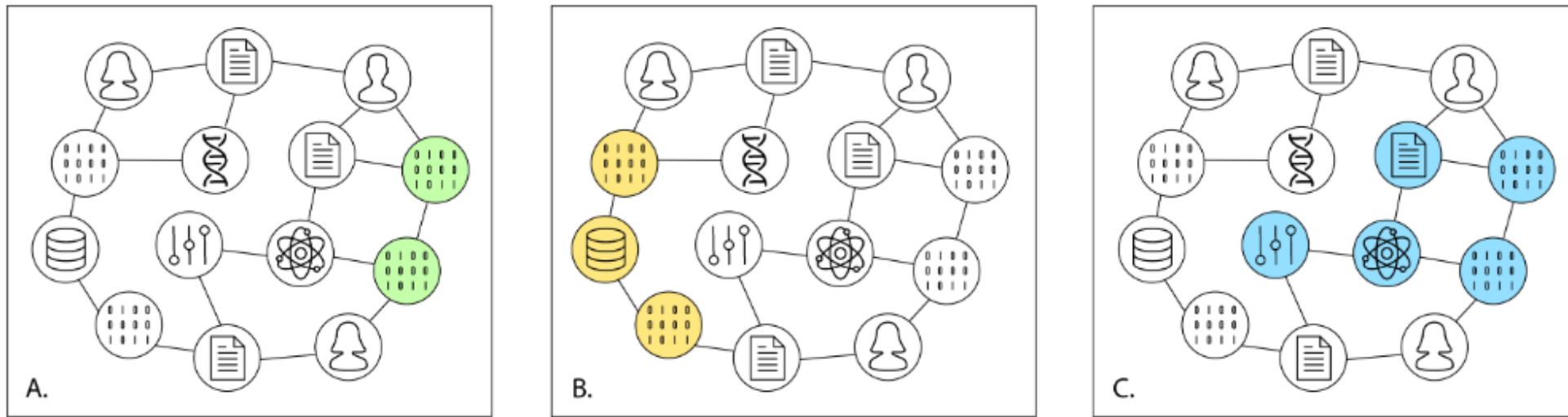


Fig 1. A schematic representation of the PID graph with digital objects connected by PIDs, showing three use cases: A: Different versions of software code, B: Datasets hosted by a particular repository, C: All digital objects connected to a research object.

<https://blog.datacite.org/introducing-the-pid-graph/>

# USE OPEN DATA TO IMPROVE PROCESSES & DECISIONS MAKING

## 利用开放数据改进流程和决策

Monitor open access  
监测开放访问

Negotiation with  
publishers  
出版社谈判

Power search and  
recommendations  
助力资源检索和推  
荐

Evaluate grant  
performance  
评估研究补助金

Other AI applications  
其他人工智能应用

<https://force2019.sched.com/event/U2my/stronger-together-a-single-open-dataset-to-map-every-facet-of-global-research>

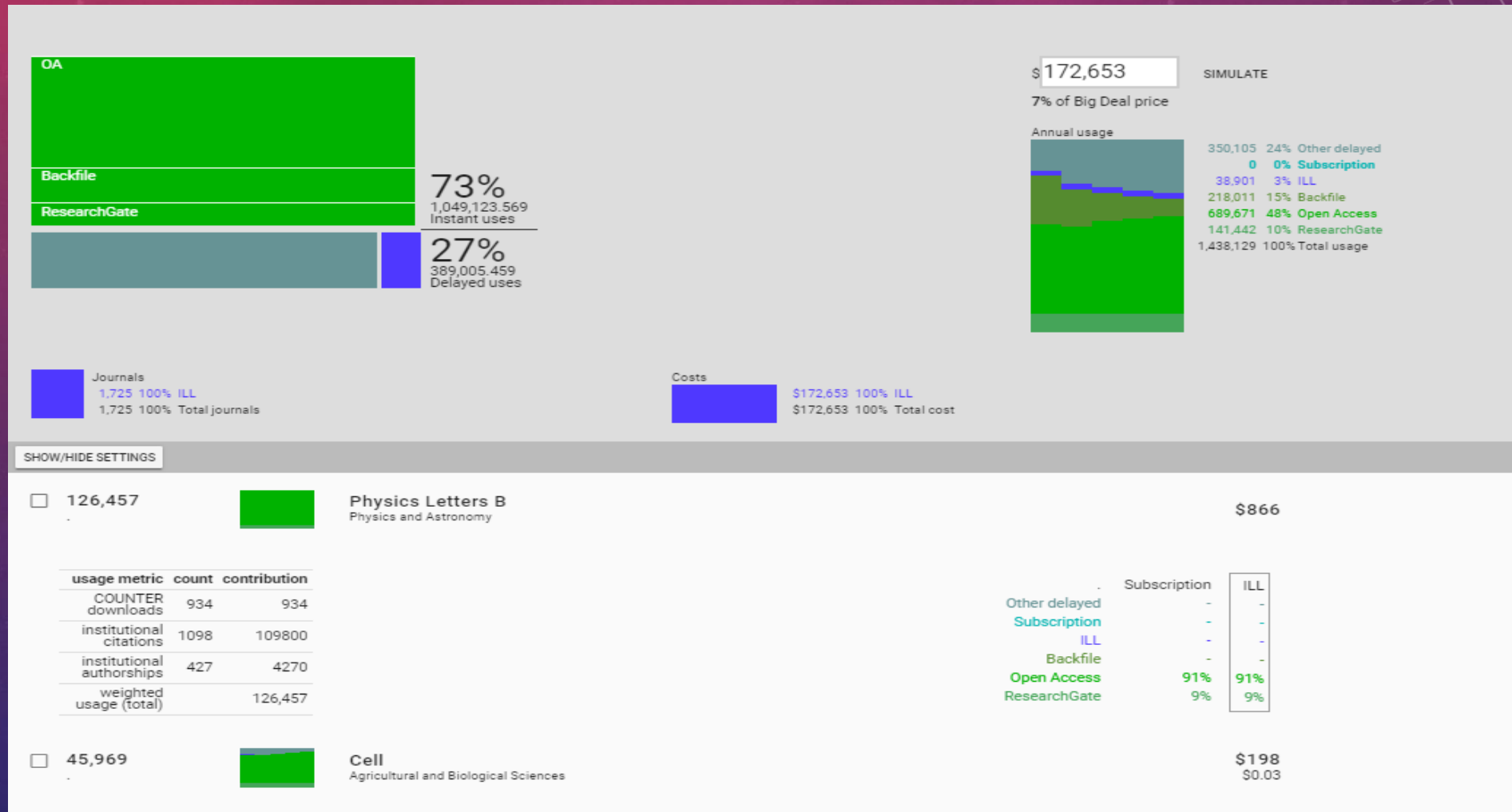
# COMBINING OPEN DATA + INSTITUTIONAL DATA FOR DECISION MAKING – UNPAYWALL JOURNALS (为决策而合并的开放数据+ 机构数据)

We're building a web-based tool called Unpaywall Journals. It creates a five-year projection of how cancellations will impact access at your institution. We do this by mashing up lots of historical data, all at the level of each individual journal title:

- your COUNTER usage numbers
- access logs from the widely-used Unpaywall extension (unlike COUNTER, this lets us bring *article-level* stats to the party)
- citations and authorship from your faculty
- hybrid, bronze, and green OA current status and trends
- perpetual access you already have to historical content
- costs of document delivery and ILL
- per-title costs (if purchased a la carte, outside of a Big Deal bundle)

Unpaywall Journals – New Product by Our Research (formerly Impactstory)

# COMBINING OPEN DATA + INSTITUTIONAL DATA FOR DECISION MAKING – UNPAYWALL JOURNALS (为决策而合并的开放数据+ 机构数据)



# NEW ACADEMIC SEARCH ENGINES BENEFITING FROM "OPEN" SCHOLARLY DATA 得益于“开放”学术数据的检索引擎



# LENS – BLENDING OF MULTIPLE OPEN SOURCES

## 融合多种开放资源

Scholarly records from PubMed (28.6M)

Scholarly records from Crossref (98.3M)

Scholarly records from Microsoft Academic(164.5M)

PubMed、Crossref和微软学术分别能检索到多少资源

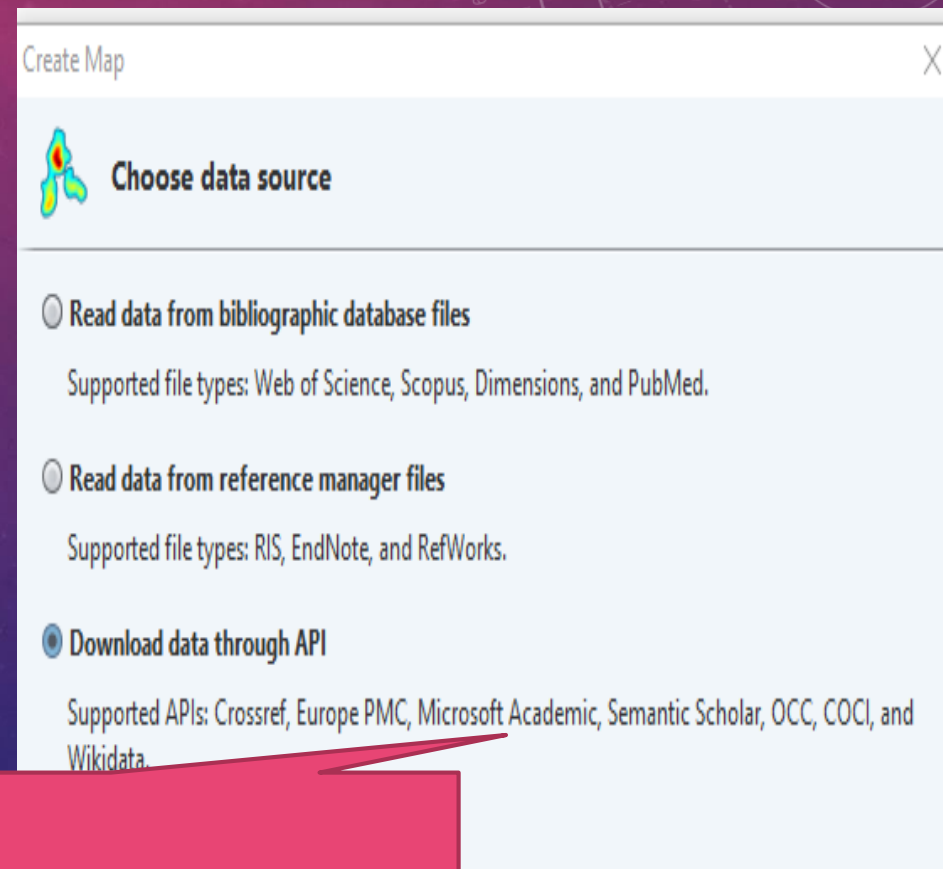
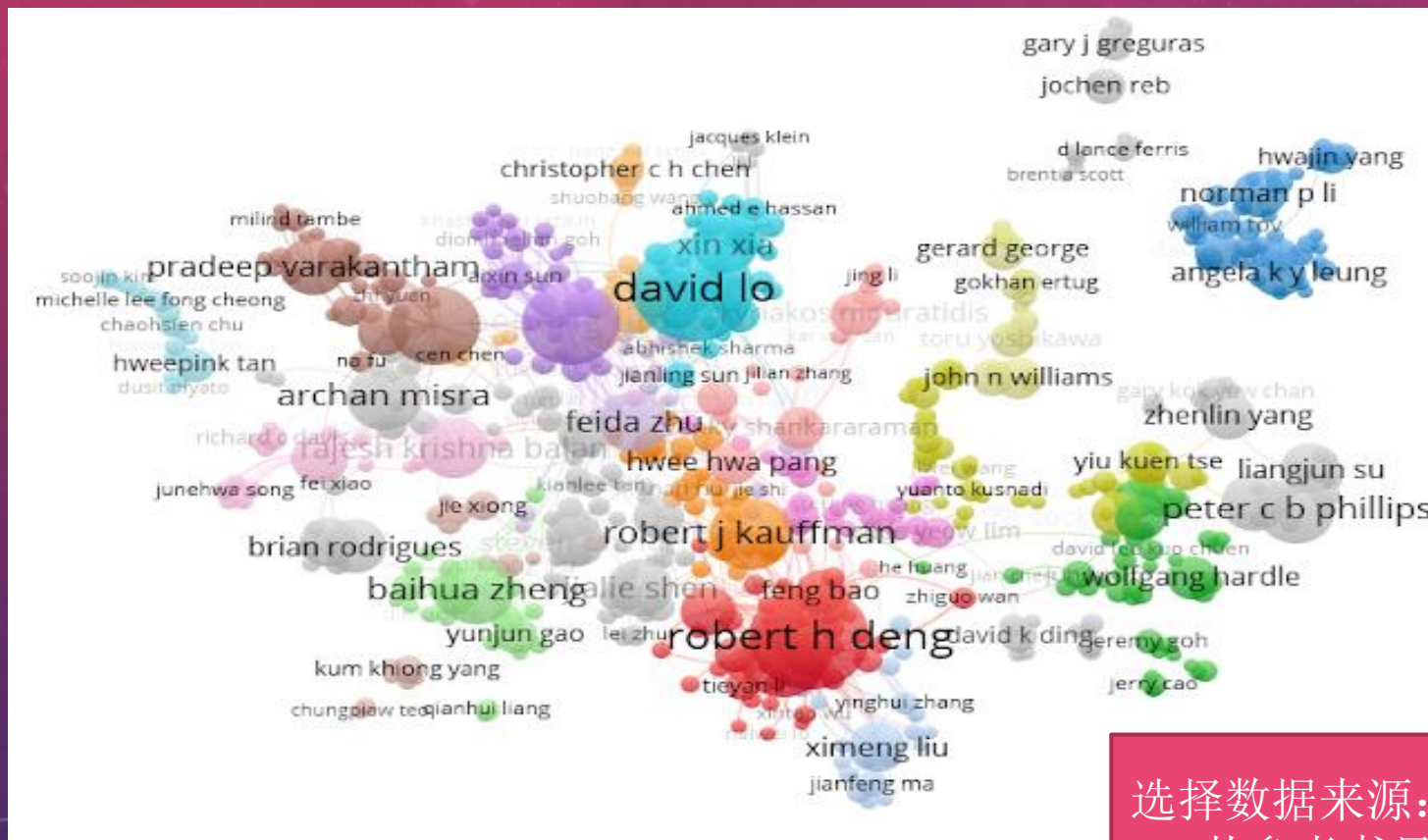
The screenshot displays the LENS interface with the following details:

- Navigation tabs: Scholarly Works (Cited), Citing Patents, List, Analysis.
- Tools: Checkmark, Eye, Cloud, Print, ON (Open Access), Scholarly Citations.
- Search Result:
  - EGFR mutation detection in ctDNA from NSCLC patient plasma: A cross-platform comparison of leading technologies to support the clinical development of AZD9291.
  - (Selected)
  - Authors: Kenneth S Thress, Roz Brant, T Hedley Carr, Simon Dearden, Suzanne Jenkins, Helen Brown, Tracey Hammett, Mireille Cantarini, J Carl Barrett
  - Journal: *Lung Cancer*, Issue: 3, Volume: 90, Pages: 509-515. | Oct 9, 2015
  - Additional Info: Open Access, Patents, Full Text, Substance, Funding, Clinical Trial, Affiliation.
  - Field of Study: [Icon]
  - 1 Patent Citations, 140 Scholarly Citations, Reference Count: 30
  - Journal Article: 045-401-560-949-469, 2097560495, 10.1016/j.lungcan.2015.10.004
  - 26494259, Find full-text at your institution

<https://musingsaboutlibrarianship.blogspot.com/2018/11/lensorg-detailed-review-of-new-open.html>



# VISUALIZATION OF BIBLIOMETRIC NETWORKS – BEYOND SCOPUS AND WEB OF SCIENCE

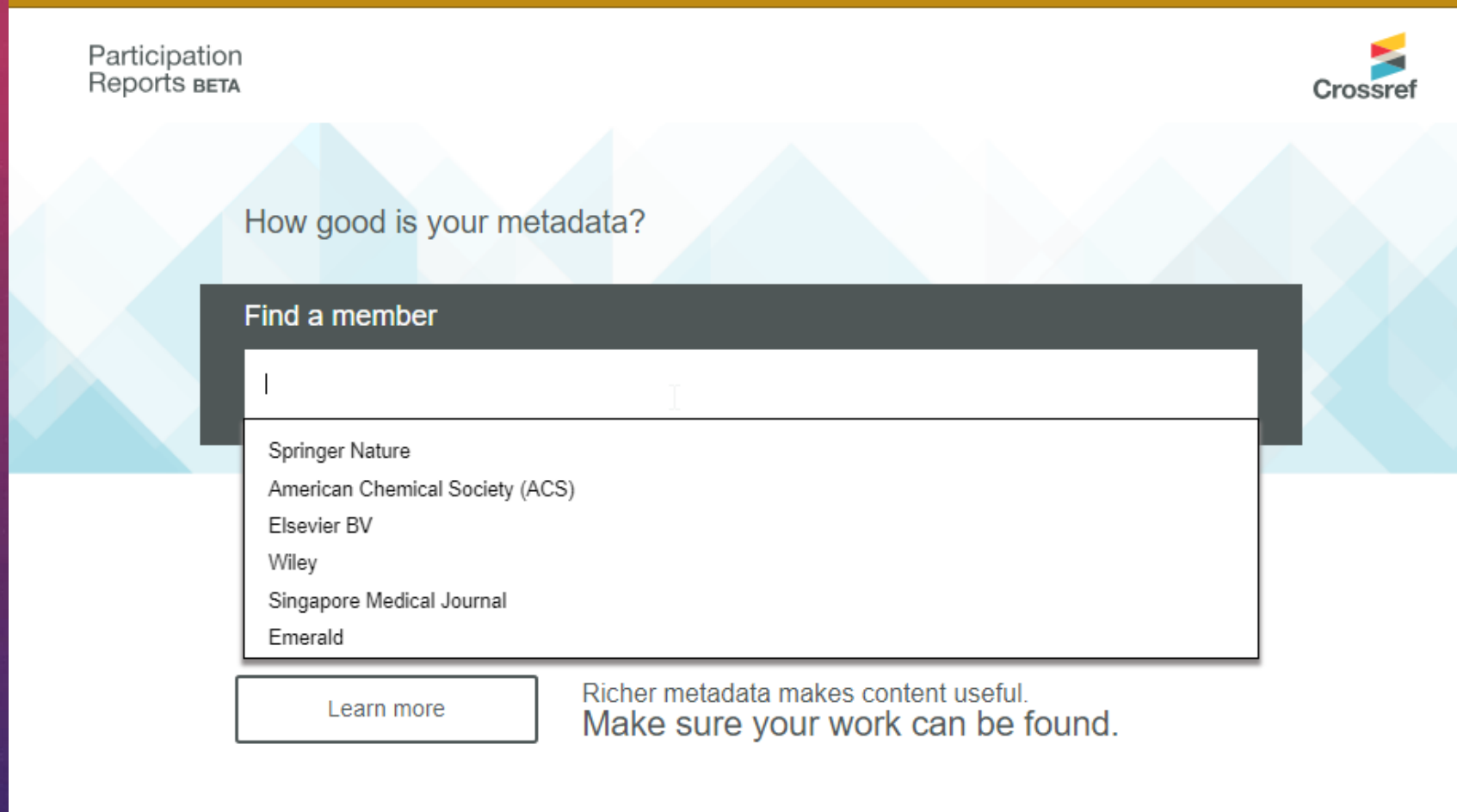


选择数据来源:

- 1、从参考书目数据库中读取数据
- 2、从引用管理文件中读取数据
- 3、通过网页插件下载数据

<http://www.vosviewer.com/>

# MONITORING THE PERFORMANCE OF PUBLISHERS



Participation Reports BETA

Crossref

How good is your metadata?

Find a member

|

- Springer Nature
- American Chemical Society (ACS)
- Elsevier BV
- Wiley
- Singapore Medical Journal
- Emerald

Learn more

Richer metadata makes content useful.  
Make sure your work can be found.

<https://www.crossref.org/members/prep/>

# An example of a "Good" Publisher

举例：元数据做得好的出版社

Richer metadata makes content useful.  
Make sure your work can be found.

Springer Nature

1,289,719  
Total registered  
content items

Content type: Journal articles

Journal articles 739,270   Datasets 2,566   Books 31,609   Book chapters 516,274

Journal articles

Search by title

Current content

References

92%

Open references

99%

ORCID IDs

33%

Funder Registry IDs

31%

Funding award numbers

25%

Crossmark enabled

89%

Text mining URLs

83%

License URLs

83%

Similarity Check URLs

99%

# An example of a "Bad" Publisher

举例：元数据做得不好的出版社

Richer metadata makes content useful.  
Make sure your work can be found.

Elsevier BV

1,738,153  
Total registered  
content items

Content type: Journal articles

Journal articles 1,639,076 Books 2,486 Book chapters 96,591

Journal articles

Search by title

Current content

References

75%

Open references

0%

ORCID IDs

20%

Funder Registry IDs

39%

Funding award numbers

31%

Crossmark enabled

75%

Text mining URLs

96%

License URLs

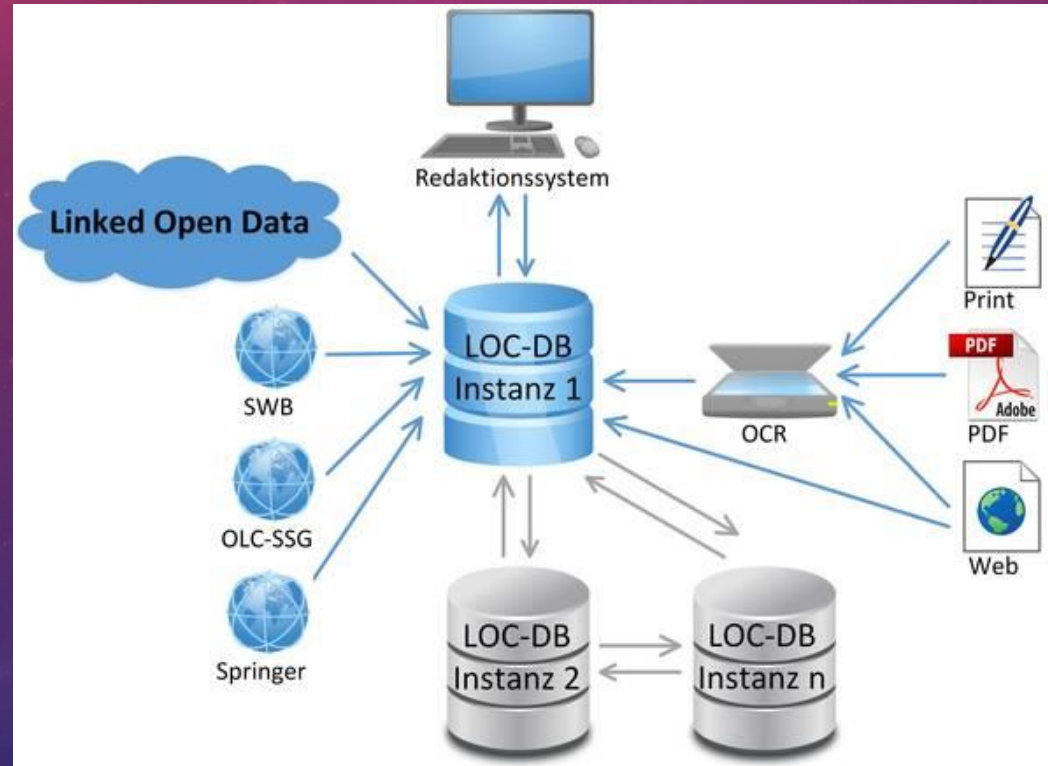
96%

Similarity Check URLs

0%

# CROWD SOURCING OF METADATA BY LIBRARIES

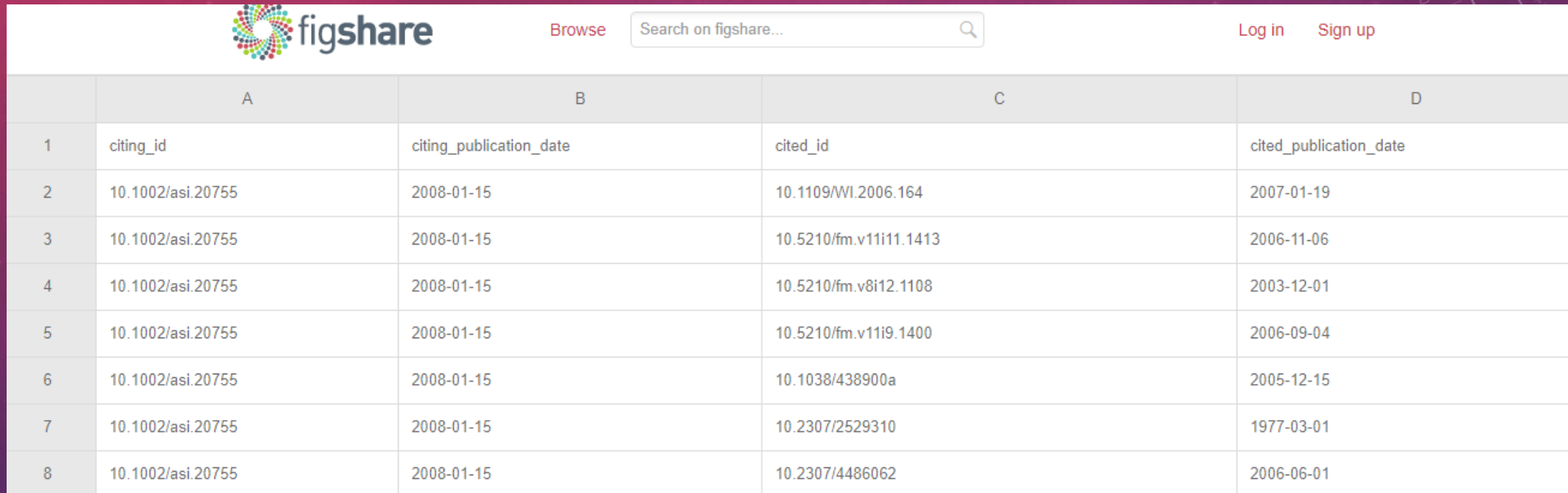
## 图书馆进行元数据众包



<https://locdb.bib.uni-mannheim.de/blog/en/>

# CROCI, THE CROWDSOURCED OPEN CITATIONS INDEX

## 众包的施引文献索引



The screenshot shows the Figshare website interface. At the top left is the Figshare logo. To its right is a 'Browse' button and a search bar with the placeholder text 'Search on figshare...'. Further right are 'Log in' and 'Sign up' links. Below the navigation bar is a table with four columns labeled A, B, C, and D. The table contains eight rows of data, each representing a citation record. The columns contain the following information: A (citing\_id), B (citing\_publication\_date), C (cited\_id), and D (cited\_publication\_date).

	A	B	C	D
1	citing_id	citing_publication_date	cited_id	cited_publication_date
2	10.1002/asi.20755	2008-01-15	10.1109/WI.2006.164	2007-01-19
3	10.1002/asi.20755	2008-01-15	10.5210/fm.v11i11.1413	2006-11-06
4	10.1002/asi.20755	2008-01-15	10.5210/fm.v8i12.1108	2003-12-01
5	10.1002/asi.20755	2008-01-15	10.5210/fm.v11i9.1400	2006-09-04
6	10.1002/asi.20755	2008-01-15	10.1038/438900a	2005-12-15
7	10.1002/asi.20755	2008-01-15	10.2307/2529310	1977-03-01
8	10.1002/asi.20755	2008-01-15	10.2307/4486062	2006-06-01

<https://opencitations.wordpress.com/2019/02/07/crowdsourcing-open-citations-with-croci/>

# DOES THE ARC OF THE SCHOLARLY COMMUNICATION UNIVERSE CURVE TOWARDS OPEN? 学术交流的穹顶是否已向打开？

- Open Access (Plan S etc)
- 开放访问
- Open Data
- 开放数据
- Open peer review
- 开放同行评审
- Open Syllabus (<http://opensyllabusproject.org/>)
- 开放字节





THANK YOU!  
感谢您!

[AARONTAY@GMAIL.COM](mailto:AARONTAY@GMAIL.COM)

[@AARONTAY ON TWITTER](https://twitter.com/AARONTAY)

[HTTPS://MUSINGSABOUTLIBRARIANSHIP.BLOGSPOT.SG](https://musingsaboutlibrarianship.blogspot.sg)