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

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

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.

THE RISE OF OPEN SCHOLARLY DATA & POSSIBLE IMPLICATIONS

9TH 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

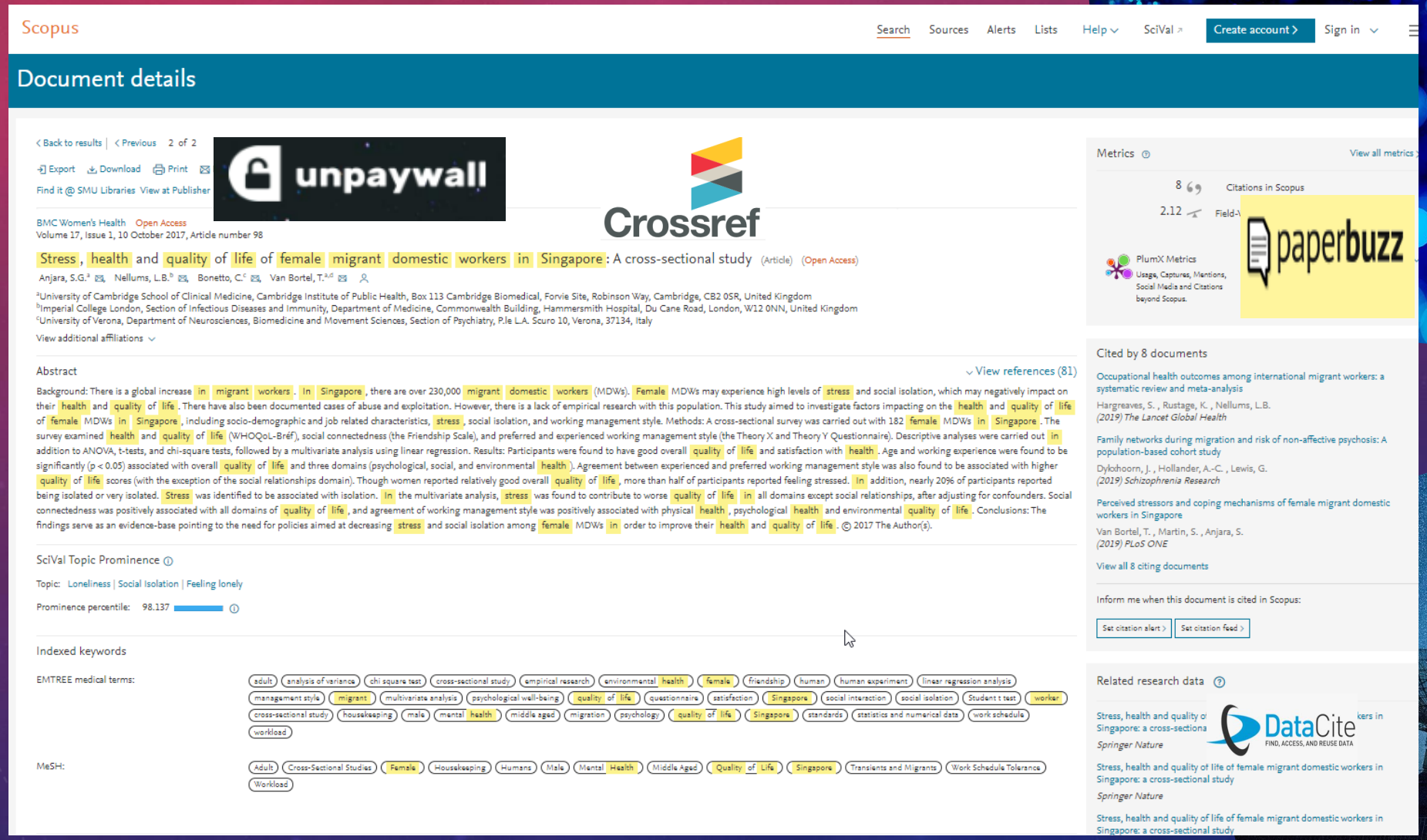
就现在的趋势看，我们预测在2025年来临前，44%的期刊文章将被开放发表、70%的访问量将来自开放访问文章

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**

//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



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，可而通过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

WIKIDATA

https://www.wikidata.org

CORE

https://core.ac.uk/

BASE

https://www.base-search.net

Microsoft Academic

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

iCite

http://unpaywall.org/

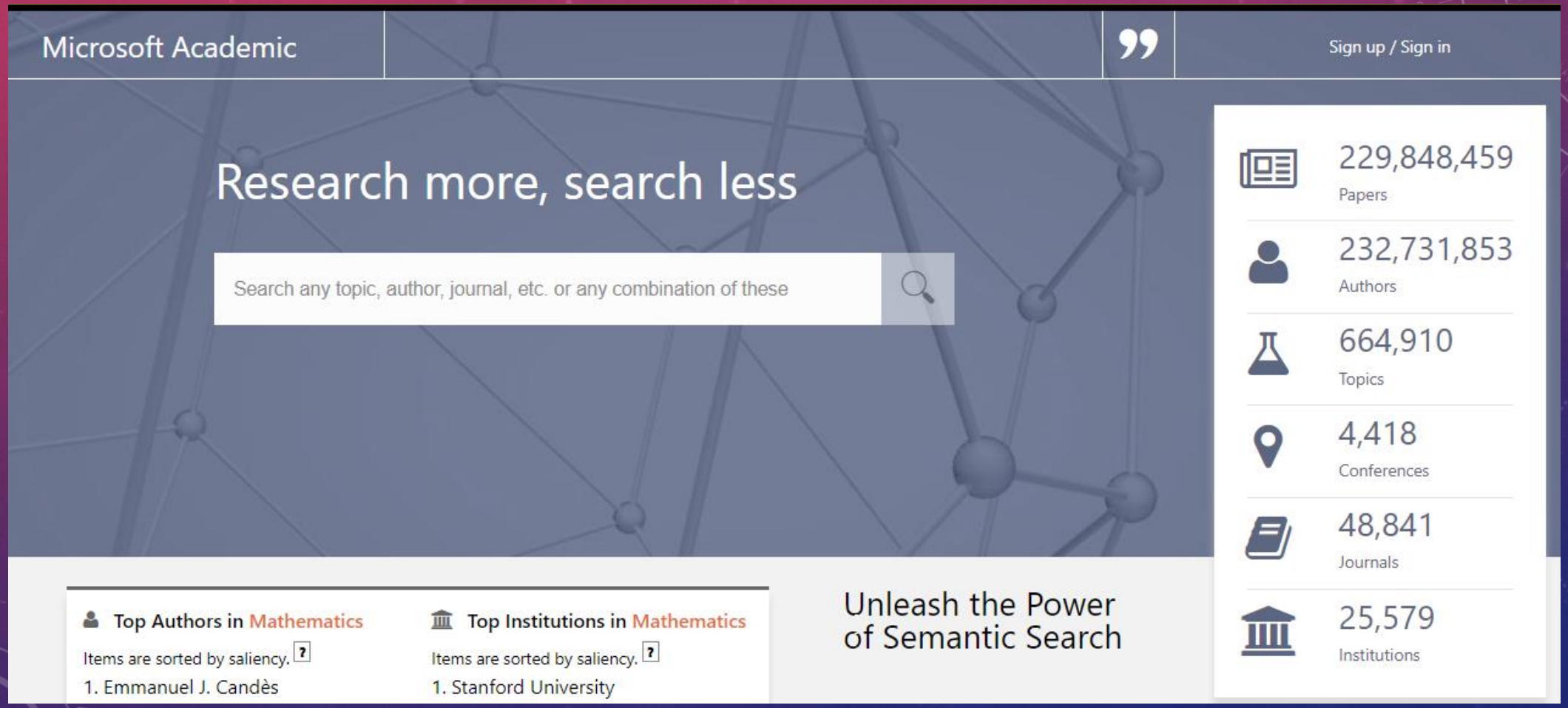
unpaywall

http://unpaywall.org/

paperbuzz

https://paperbuzz.org/

MICROSOFT (微软) ACADEMIC

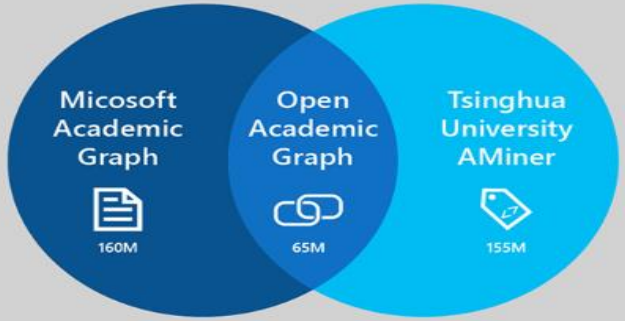


https://academic.microsoft.com/home

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

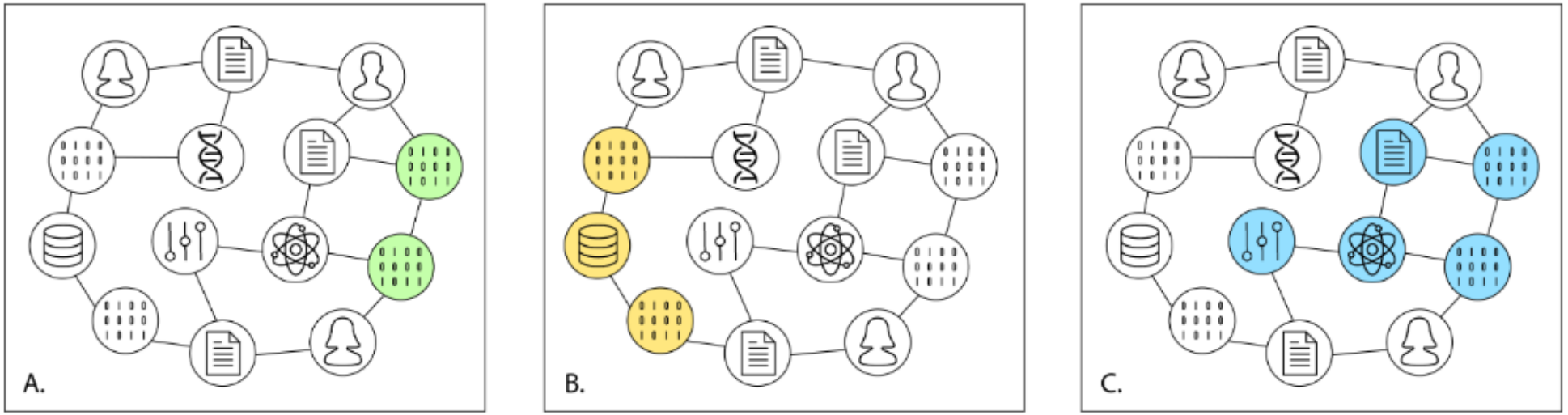


Fig 1. A schematic representation of the PID graph with digital objects connected by PIDs, showing three use cases: A: Different versions of software code, B: Datasets hosted by a particular repository, C: All digital objects connected to a research object.

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

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

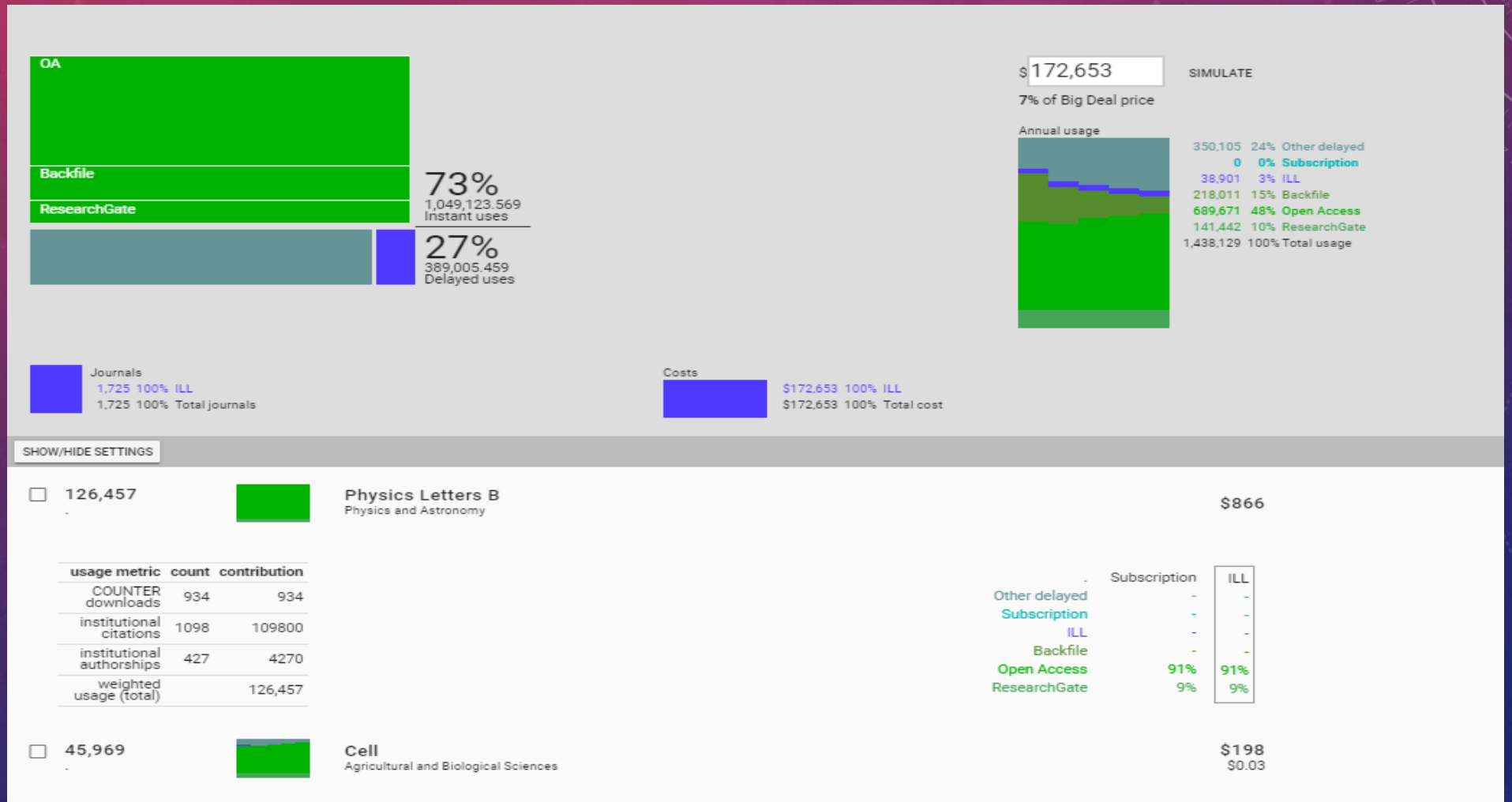
COMBINING OPEN DATA + INSTITUTIONAL DATA FOR DECISION MAKING – UNPAYWALL JOURNALS (为决策而合并的开放数据+ 机构数据)

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

Crossref ORCID PubMed.gov Microsoft Academic CORE unpaywall

Microsoft Academic Semantic Scholar SN SciGraph PubMed.gov

LENS.ORG
Solving The Problem Of Problem Solving™

The Lens serves global **patent** and **scholarly** knowledge as a public good to inform science and technology-enabled problem solving.

Patent Search | Scholar Search

Semantic Scholar

NEW!

scinapse
Academic Search Engine

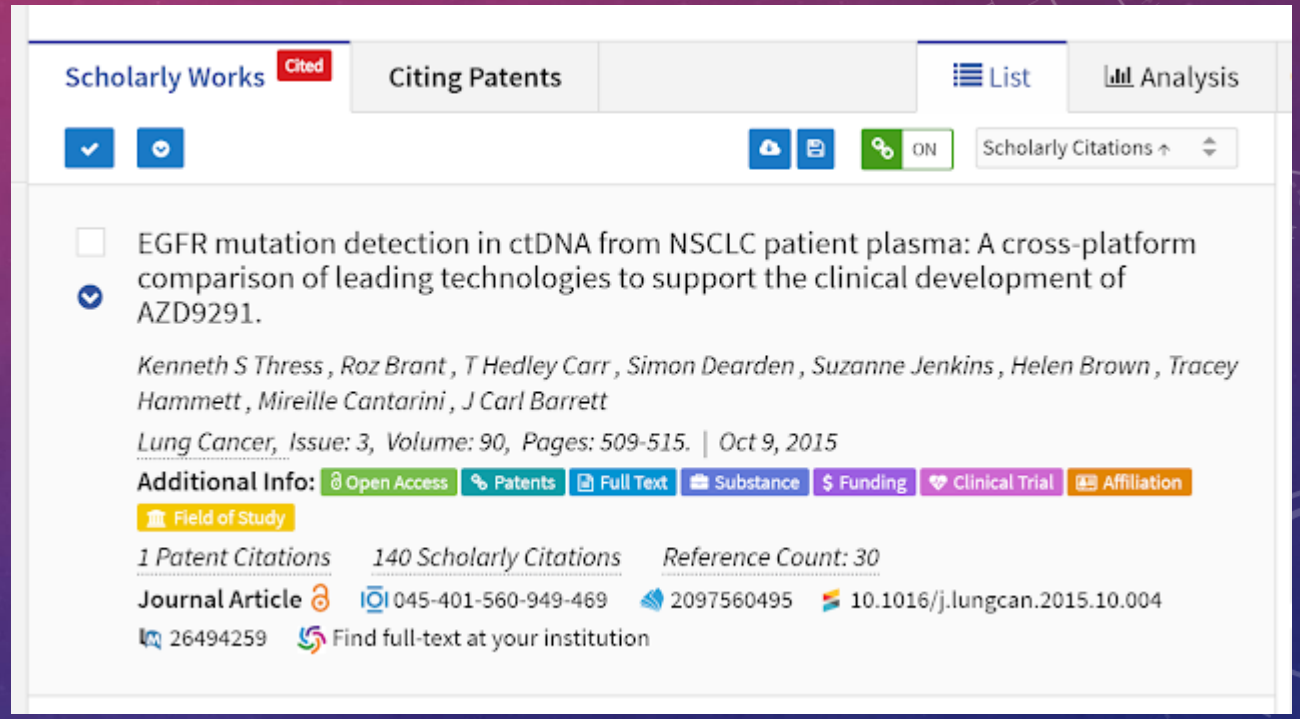
LENS – BLENDING OF MULTIPLE OPEN SOURCES
融合多种开放资源

Scholarly records from PubMed (28.6M)

Scholarly records from Crossref (98.3M)

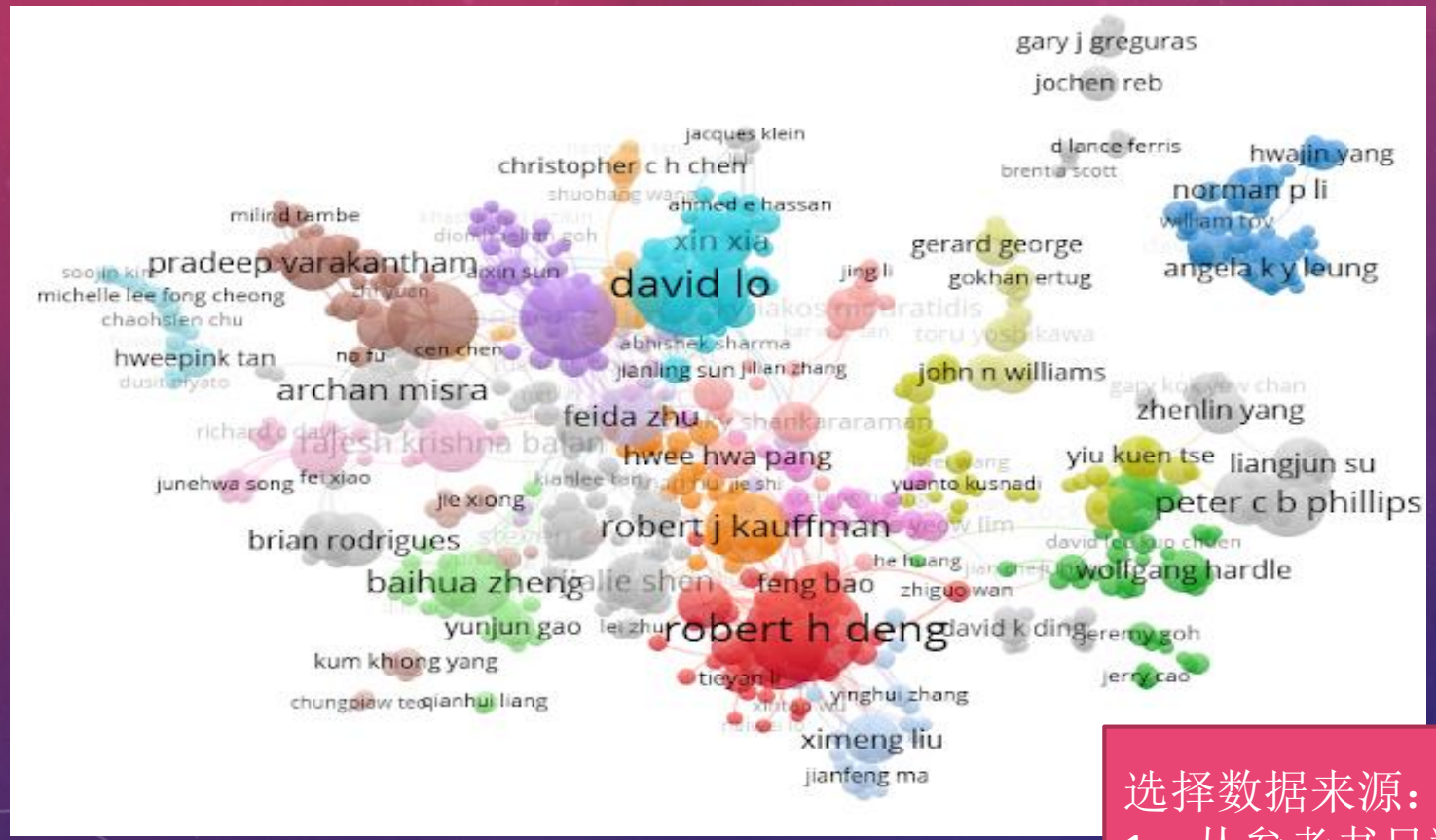
Scholarly records from Microsoft Academic(164.5M)

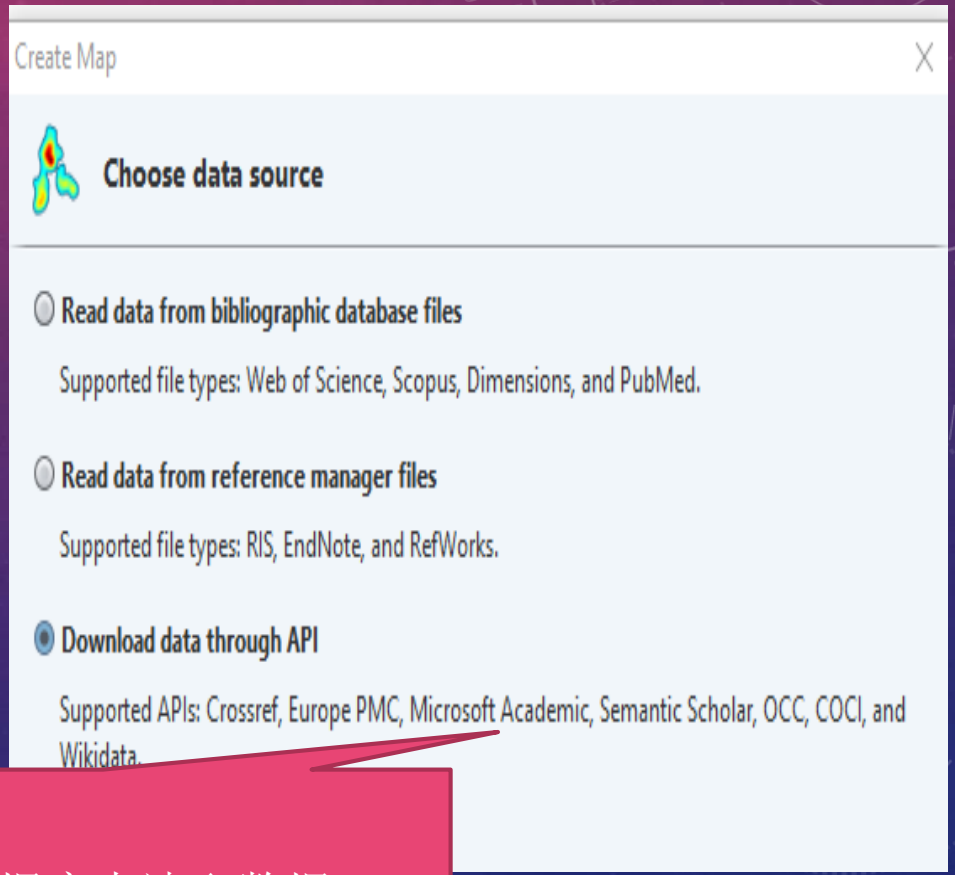
PubMed、Crossref和微软学术分别能检索到多少资源



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

VISUALIZATION OF BIBLIOMETRIC NETWORKS – BEYOND SCOPUS AND WEB OF SCIENCE

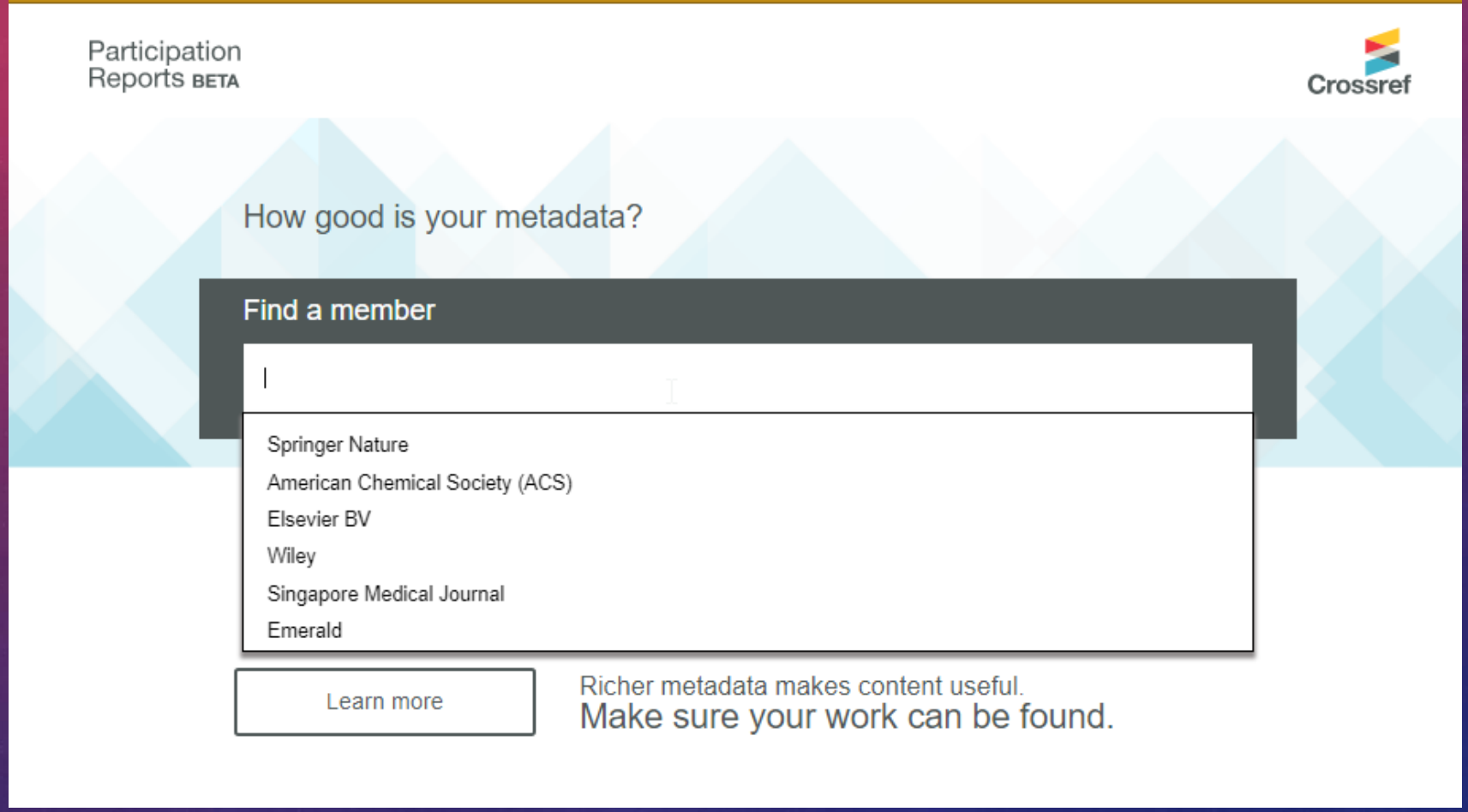




http://www.vosviewer.com/

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

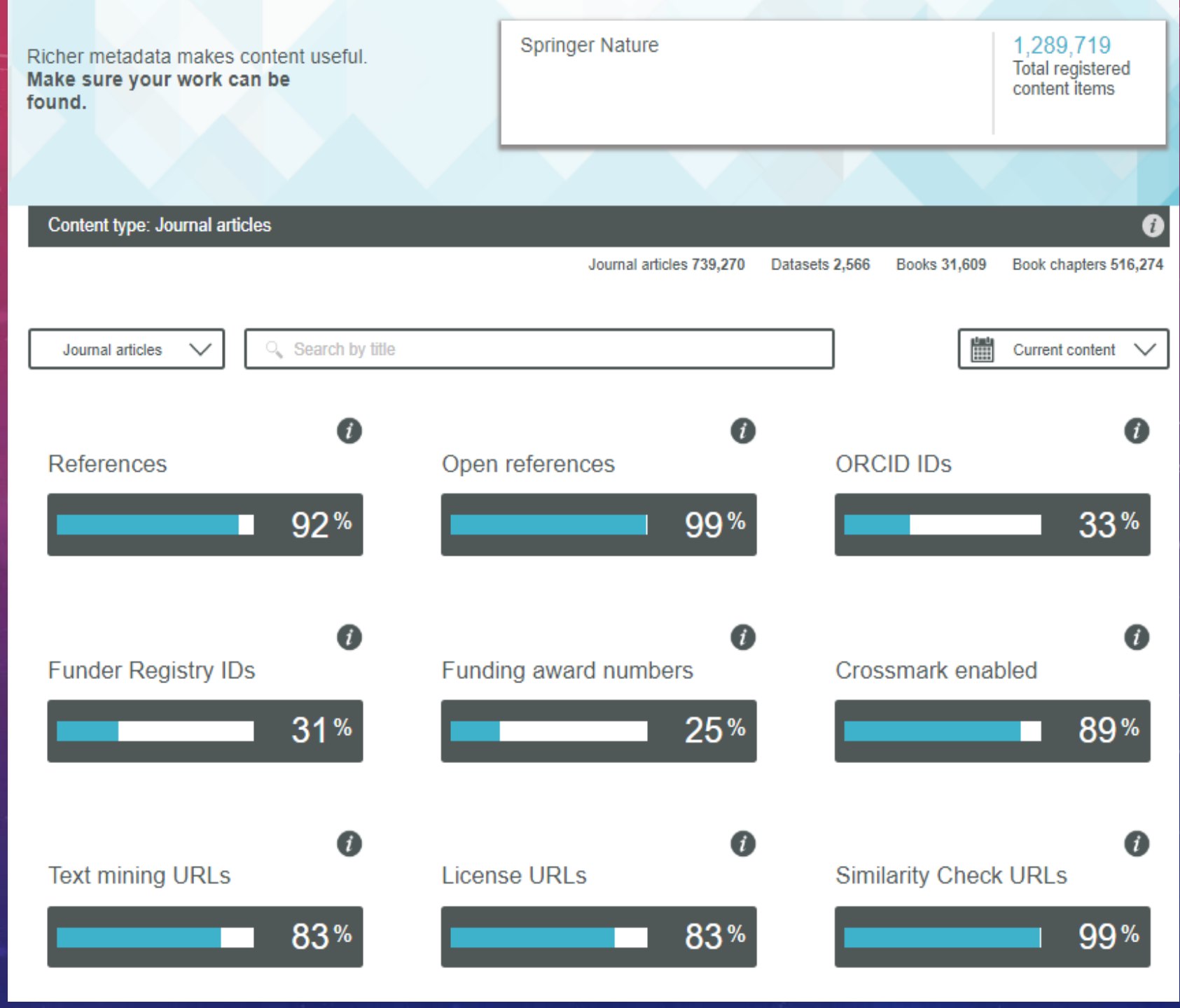
MONITORING THE PERFORMANCE OF PUBLISHERS



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

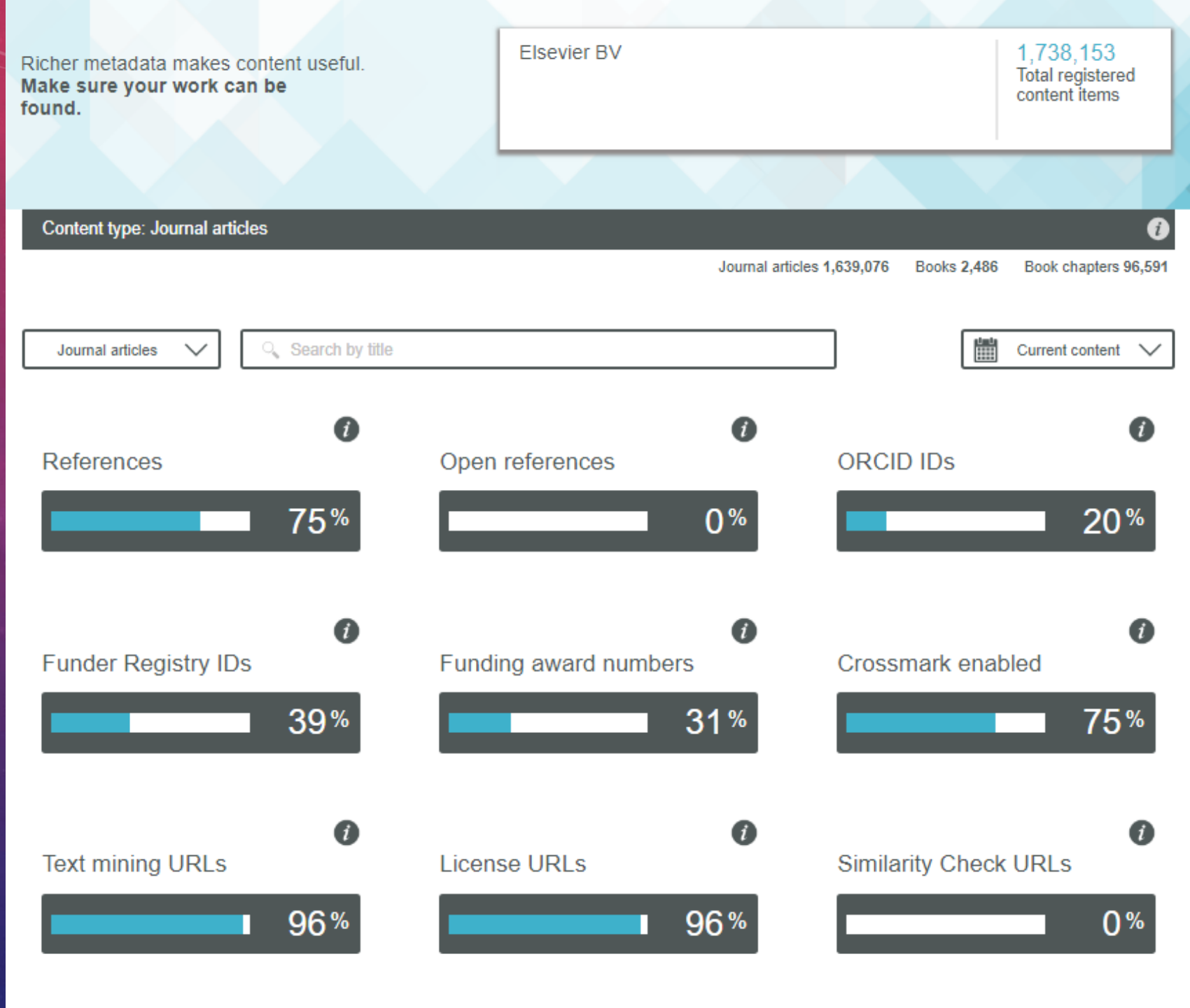
An example of a "Good" Publisher

举例：元数据做得好的出版社

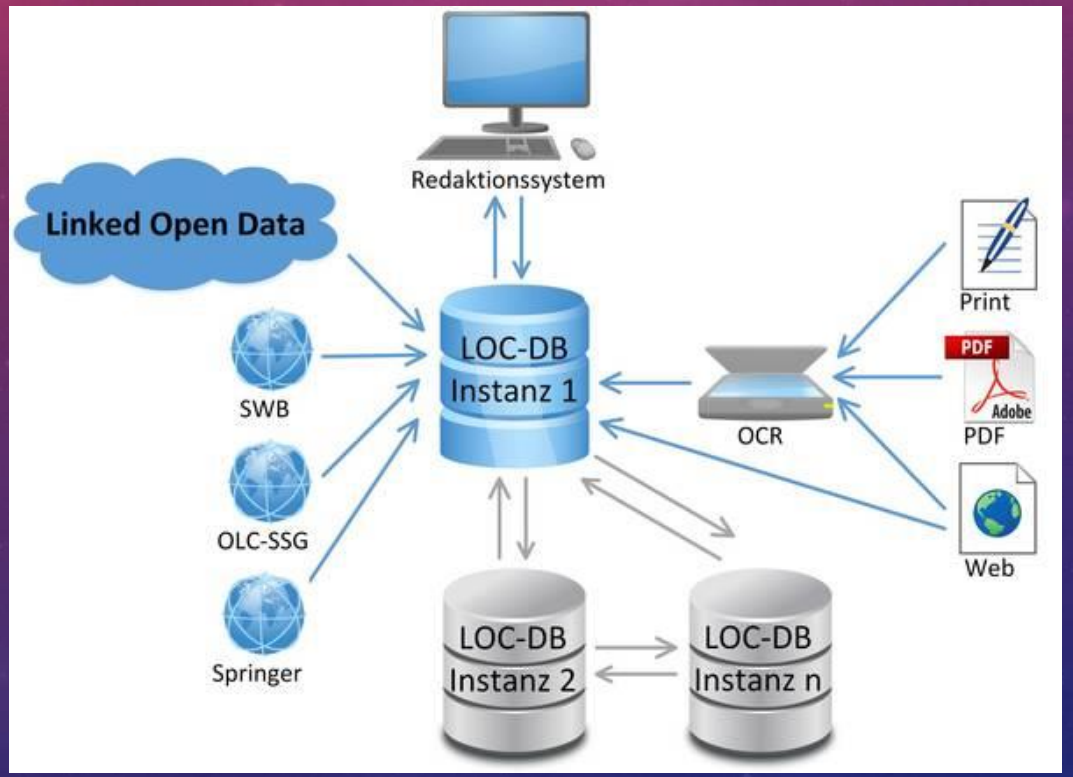


An example of a "Bad" Publisher

举例：元数据做得不好的出版社



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

figshare Browse Search on figshare... Log in Sign up

	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
- 开放的同行评审
- Open Syllabus（http://opensyllabusproject.org/）
- 开放字节

THANK YOU!
感谢您!

AARONTAY@GMAIL.COM

@AARONTAY ON TWITTER

HTTPS://MUSINGSABOUTLIBRARIANSHIP.BLOGSPOT.SG