

Singapore Management University

## Institutional Knowledge at Singapore Management University

---

COAR Asia OA Meeting 2021

COAR Asia OA Meeting 2021

---

Oct 26th, 1:00 PM

### Cashing the cheque of open access movement: Emerging tools built on open access data

Aaron TAY  
*Singapore Management University*

Follow this and additional works at: <https://ink.library.smu.edu.sg/asiaoa2021>



Part of the [Scholarly Communication Commons](#)

---

TAY, Aaron. Cashing the cheque of open access movement: Emerging tools built on open access data. (2021). COAR Asia OA Meeting 2021. .

Available at: <https://ink.library.smu.edu.sg/asiaoa2021/program/agenda/14>

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 COAR Asia OA Meeting 2021 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).

# Cashing the Cheque of Open Access Movement: Emerging Tools Built on Open Access Data

Aaron Tay (Lead, Data Services)

Singapore Management University Libraries

26 October 2021



[Aarontay@gmail.com](mailto:Aarontay@gmail.com)



[@aarontay](https://twitter.com/aarontay)



<https://musingsaboutlibrarianship.blogspot.com>

## What does the title of the talk mean?

***"finally cash the cheques written by the Open Access movement."***  
**-- OurResearch, July 2018**

We're building the "AI-powered support tools" now. What kind of tools? Well, let's go back to the Hamlet example...today, publishers solve the context problem for readers of Shakespeare by adding notes to the text that define and explain difficult words and phrases. We're gonna do the same thing for 20 million scholarly articles. And that's just the start...we're also working on concept maps, automated plain-language translations (think automatic [Simple Wikipedia](#)), structured abstracts, topic guides, and more. Thanks to recent progress in AI, all this can be automated, so we can do it at scale. That's new. And it's big. -- **OurResearch Blog, 2018**

## Imagine you live in a 100% OA world...

And you want to create a tool that can auto-suggest missing references based on text you have typed

### Things you may need

- Article information (Title, Abstract, Year of Publication, subject headings)
- References/Citations
- Citation context/sentiment

**Open Knowledge = Open metadata+full text**

Scholarly data



```
graph TD; SD[Scholarly data] --> OSM[Open Structured Metadata (title, abstract, reference sources etc)]; SD --> OFT[Open Full text]; OFT -- TDM --> SDATA[Structured data]; SDATA -- Make open --> OSM;
```

Open Structured Metadata (title, abstract, reference sources etc)

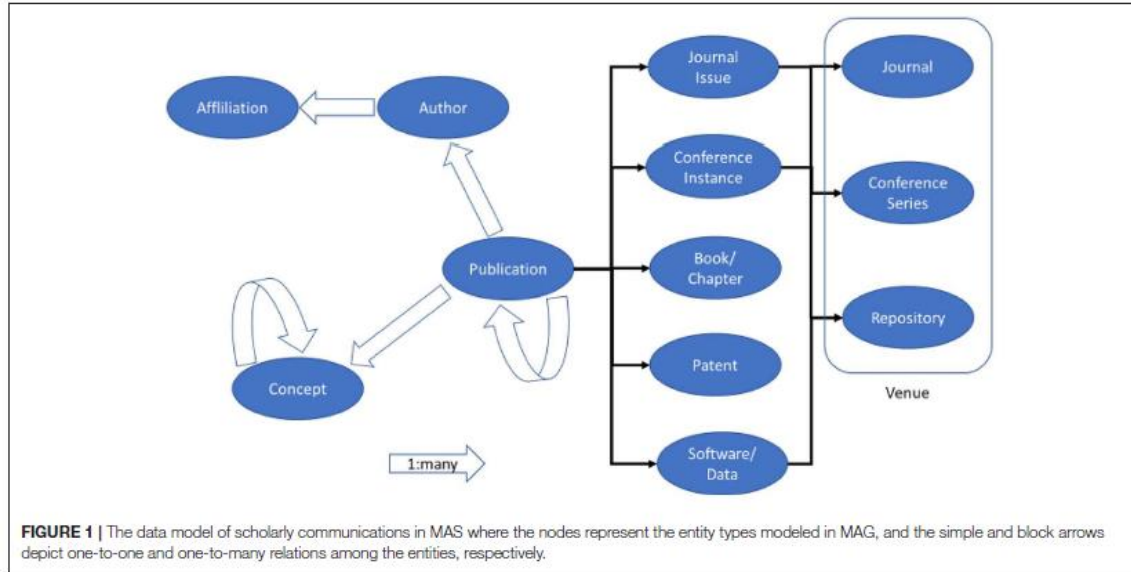
Open Full text

TDM

Make open

Structured data

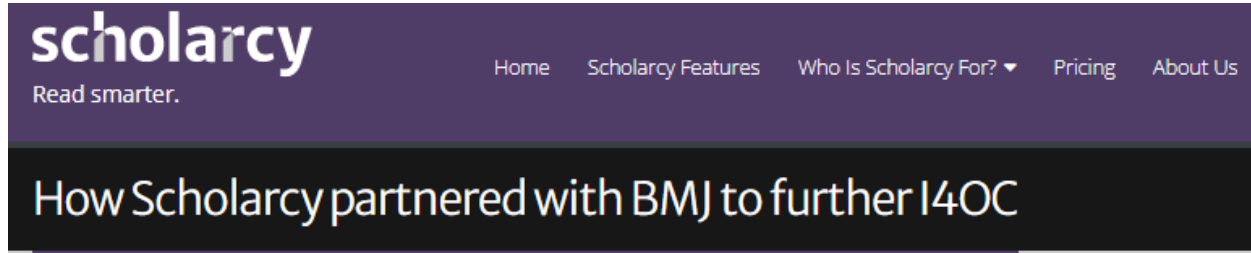
# Microsoft Academic Graph – Extracting open metadata from full-text



"These considerations precipitated the new focus of MAS from an academic search engine relying mainly on publisher feeds to a web-based approach that actively exploits the cutting edge cloud computing and AI technologies.....

<https://www.semanticscholar.org/paper/Microsoft-Academic-Graph%3A-When-experts-are-not-Wang-Shen/ea9a516d5cb0b298f0df50e82b3e0400b72fcdff>

## BMJ employs Scholarcy to extract references from back files



"[The BMJ](#) has an archive extending to hundreds of thousands of articles (some dating back to the 1840s) that exist only in PDF format. At the end of 2018, BMJ wanted to mine these PDFs for references and automatically structure them in [CrossRef](#) XML format, to make them widely available to the research community as part of the [Initiative for Open Citations](#) (I4OC)."

<https://www.scholarcy.com/unlocking-100-years-of-scientific-papers-how-scholarcy-partnered-with-bmj-to-further-i4oc/>

## Open Citation source/corpus\*

---

[Crossref/I4OC](#) (The Initiative for Open Citations)

---

[Microsoft Academic Graph](#)\*

---

[The Semantic Scholar Open Research Corpus \(S2ORC\)](#)

---

[OpenCitations Corpus \(OCC\)](#)

---

[Lens.org](#)

---

[NIH-OCC](#)

---

[Wikidata](#)

---

[FatCat](#) / [Refcat](#)

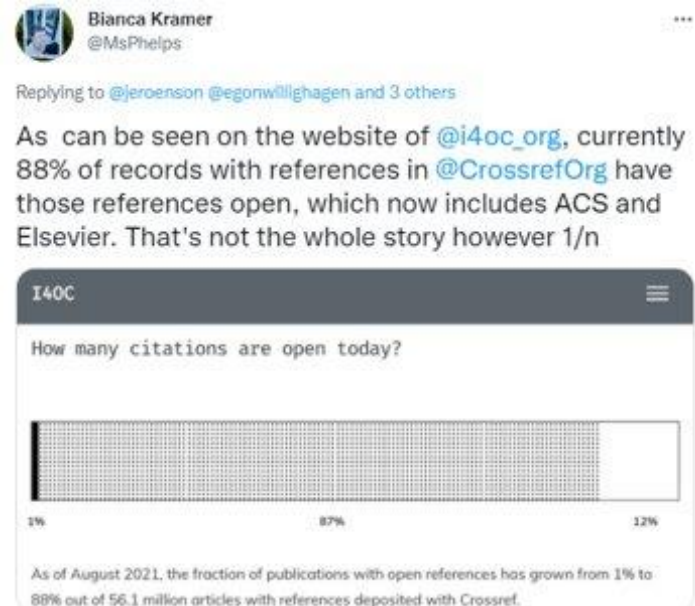
---



# I4OC (Initiative for Open Citations) – success – >1+ billion citations

All major publishers\* participating including

- \* Elsevier
- \* Springer-Nature
- \* Wiley
- \* Sage
- \* Taylor & Francis



Bianca Kramer  
@MsPhelps

Replying to @jeroenson @egonwillighagen and 3 others

As can be seen on the website of [@i4oc\\_org](#), currently 88% of records with references in [@CrossrefOrg](#) have those references open, which now includes ACS and Elsevier. That's not the whole story however 1/n

**I4OC**

How many citations are open today?

1%	88%	12%
----	-----	-----

As of August 2021, the fraction of publications with open references has grown from 1% to 88% out of 56.1 million articles with references deposited with Crossref.

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

Title/author/abstract/subject

References (relationships!)

Affiliations

Funding info

Altmetrics

Open access status

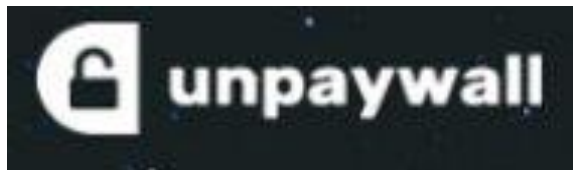
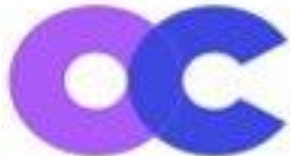
Others? - Open peer review data etc

## Some sources of Open Scholarly Metadata



[Crossref](#), [Datacite](#), [ORCID](#), [ROR](#), [NIH](#)

## Some sources of Open Scholarly Metadata (II)



[CORE](#), [BASE](#), [Wikidata](#), [OpenCitations](#), [EuroPMC](#), [Lens.org](#), [Unpaywall](#), [Internet Archive \(Fatcat\)](#)

# Innovations in Scholarly Communication – Workflow tools - categorization

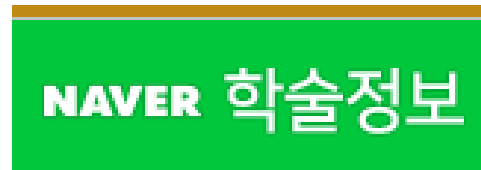


Q : How would the ready availability of open structured Scholarly metadata and full text strengthen innovation in research flow tools? Which existing tools become even better in a world of OA?

# Discovery

- New Discovery citation indexes (e.g. Lens.org)
- Bibliometric/Science mapping tools & Literature mapping tools
- Systematic review tools
- Knowledge extraction & Summarizer tools

## New Scholarly search citation indexes (Cross Disiplinary)



[See coverage of new citation indexes](#)

# New Scholarly search citation indexes





## Science mapping tools (for bibliometrics researchers)



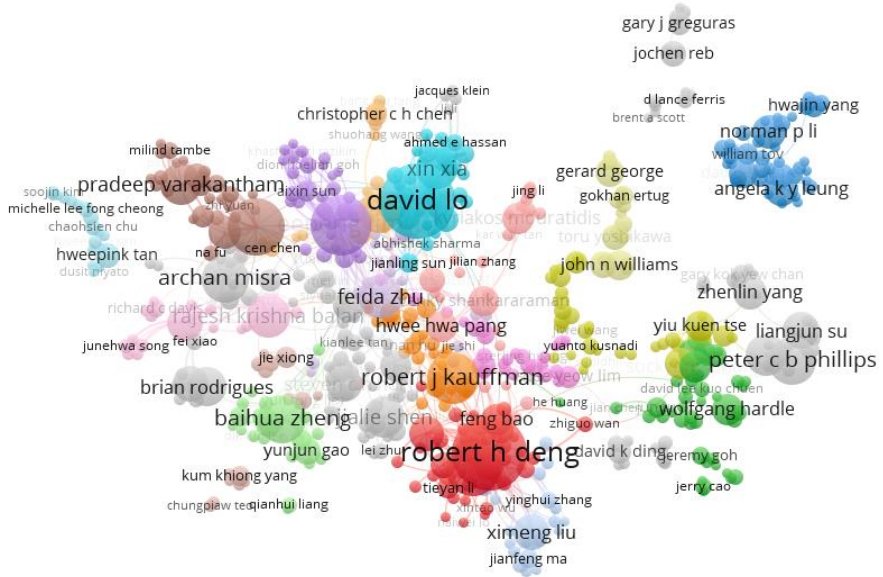
*CiteSpace*



# Harzing's Publish or Perish

[VOSviewer](#), [Citespace](#), [Bibliometrix](#), [CitNetExplorer](#), [Sci2](#), [HistCite](#),  
[Hazing Publish or Perish](#)

# Science mapping tools can now accept data from more inclusive sources e.g. MAG, COCI,



Create Map ✕

---

**Choose data source**

---

- Read data from bibliographic database files**  
Supported file types: Web of Science, Scopus, Dimensions, Lens, and PubMed.
- Read data from reference manager files**  
Supported file types: RIS, EndNote, and RefWorks.
- Download data through API**  
Supported APIs: Microsoft Academic, Crossref, Europe PMC, Semantic Scholar, OCC, COCI, and Wikidata.

Vosviewer (co-authorship network using-MAG)

Some options in Vosviewer

## Citation based Literature mapping services (for researchers)



[See list of emerging tools](#)





## Can Easy-To-Use Text Mining Applications Help With Information Retrieval Tasks? (CADTH, 2018)

term and phrase selection

search development for vague topics

concept identification

relevance ranking to assist with search refinement

Filter development

Autoscreening

**[CADTH Text Mining Opportunities: White Paper](#)**

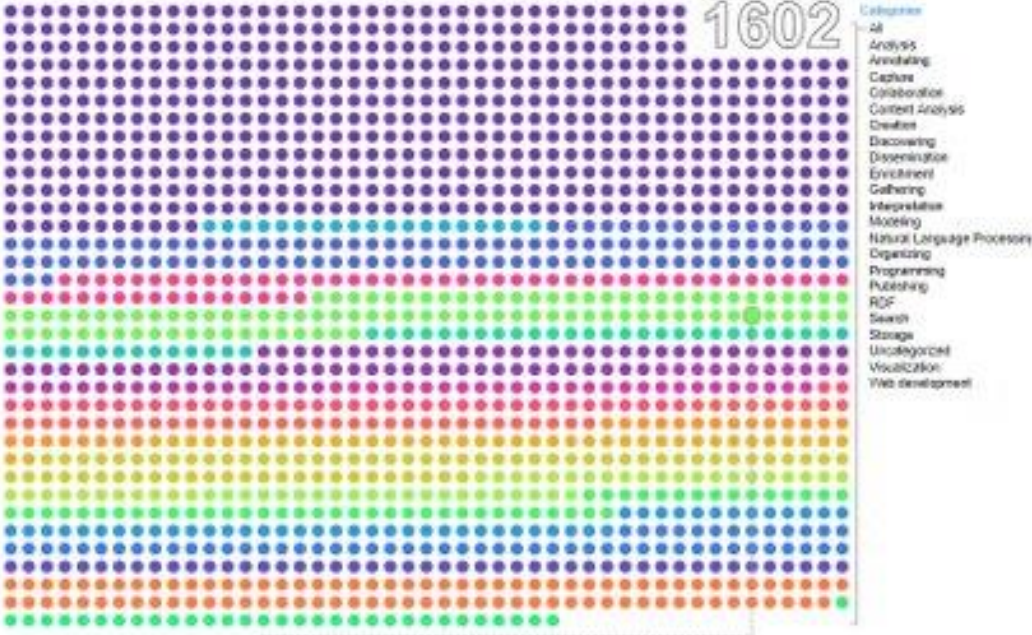
# Some text tool

- [Voyant](#)
- [AntConc](#)
- [TerMine](#)
- [ProAnt](#)
- [Many more](#)

TAPoR 3 Discover research tools for studying texts.

Browse the TAPoR collection:

1602



Categories:

- All
- Analysis
- Annotation
- Capture
- Collaboration
- Content Analysis
- Creation
- Discovering
- Dissemination
- Enrichment
- Gathering
- Interpretation
- Modeling
- Natural Language Processing
- Organizing
- Programming
- Publishing
- PDF
- Search
- Storage
- Uncategorized
- Visualization
- Web development

[TAPoR3](#)

## Going beyond - Citation sentiment/context

Citation Statements (sometimes known as citances) are the sentences surrounding a citation or reference within a publication.

For example, the following is a citation statement from *Neighborhood Change and Crime in the Modern Metropolis* Kirk and Laub (2010) to *Crack, Cocaine and Heroin: Drug Eras in Williamsburg, Brooklyn, 1960-2000* Curtis (2003)

“

Yet, this accessibility contributed to gentrification in the 1980s. As **Curtis (2003)** observes, residents displaced from Manhattan in the mid-1980s because of rising rents flocked to nearby Williamsburg. Many boarded-up buildings and abandoned industrial areas, some of which were used as "shooting galleries" and drug stashes, were transformed into lofts and condos for gentrifiers.

<https://help.scite.ai/en-us/article/how-does-citation-statement-search-work-13tnvx7/>

# Going beyond - Citation sentiment/context

## Scite ([Details](#))

- Mentioning cite
- Supporting cite
- Constrasting cite

## Semantic Scholar ([Details](#))

- Cites Background
- Cites Methods
- Cites Results
- influential cites

## Web of Science (beta New!)

- Background
- Basis
- Compare
- Discuss



# Others joining – Web of Science

## Enriched Cited References classified by use

Enriched cited references receives an exciting new enhancement—a classification of each in-text mention to indicate why the author may have cited the reference in each instance. Classifications are assigned by evaluating the author’s exact words in the sentence containing the mention as well as the sentence before and after. The mentions are classified as follows:

- **Background**—previously published research that orients the current study within a scholarly area.
- **Basis**—references that report the data sets, methods, concepts and ideas that the author is using for her work directly or on which the author bases her work.
- **Compare**—references that the current study’s results can be compared to; the author could be extending on a prior concept or method or confirming past findings or refuting past findings.
- **Discuss**—references mentioned because the current study is going into a more detailed discussion.

These classifications have been added to the existing enriched cited references visualization to help you find references that suit your needs but may appear outside the “expected” section. For example, if you wanted to look at references that are classified as “basis” but are not cited in the Materials and Methods section, you could review the references that appear along the “basis” classification line.

[New WoS July 22 Release Notes](#)

# Scite.ai - Open Scholarly metadata + OA full text + Publisher full-text

**scite\_**

Papers ▾

🔍

Product ▾
Resources ▾

*N Engl J Med* volume 382, issue 8, P727-733 2020 DOI: 10.1056/nejmoa2001017 [View full text](#) | 🔔 Set alert | Share [🐞](#) [📘](#) [🌐](#)

## A Novel Coronavirus from Patients with Pneumonia in China, 2019 [🔗](#)

Na Zhu<sup>1</sup>, Dingyu Zhang<sup>2</sup>, Wenling Wang<sup>3</sup> [et al.](#)

**Abstract:** The coronaviruses (CoVs) called the attention of the world for causing outbreaks of severe acute respiratory syndrome (SARS-CoV), in Asia in 2002-03, and respiratory disease in the Middle East (MERS-CoV), in 2012. In December 2019, yet again a new coronavirus (SARS-CoV-2) first identified in Wuhan, China, was associated with a severe respiratory infection, known today as COVID-19. This new virus is highly transmissible, and quickly spread throughout China and 30 additional countries. As result, the World Health Organization declared it a global health emergency.

Expand abstract ▾

**Search citation statements**

Context, author(s), title e... 🔍

Order By: Relevance ▾

**Citation Types** 🔗

<input checked="" type="checkbox"/>	Supporting	<input checked="" type="checkbox"/>	113
<input checked="" type="checkbox"/>	Mentioning	<input type="checkbox"/>	13,438
<input checked="" type="checkbox"/>	Contrasting	<input type="checkbox"/>	18
<input checked="" type="checkbox"/>	Unclassified	<input type="checkbox"/>	565

🗨️ Cited by 14,625 publications (14,134 citation statements)
📄 References 13 publications

🔗 **contrasting**

Confidence: 99%

[flag classification](#)

“...Interestingly, unlike previous literature that generally cited age as an important predictor of mortality in COVID-19, [1-3,32](#) our model did not find age to rank among the top predictors of mortality in our cohort, although age was significantly different between groups. None of the comorbidities and symptoms (except diarrhea) were among the top predictors in our cohort although some comorbidities and symptoms were significantly different between groups...”

Section: Discussion

🔗 **mentioning**

Confidence: 99%

[flag classification](#)

“...The most commonly used reverse-transcriptase polymerase chain reaction (RT-PCR) has poor sensitivity (high false-negative rate 7%) and long turnaround time (a few days to a week [8](#)) during which the patients are assumed COVID-19-positive, potentially holding up valuable resources. Moreover, there are heterogeneous symptom presentations, and many patients are asymptomatic but may deteriorate rapidly [1-3](#). Given these challenges, establishing a simplified risk stratification score system from studying the large array of clinical variables from large cohorts of patients could be helpful in COVID-19 disease management....”

Section: Introduction

[🔗 See 1 more Smart Citation](#)

32

# **KNOWLEDGE EXTRACTION & SUMMARIZATION**

# Knowledge extraction - Annotations in EuroPMC (SciLite)

The screenshot shows the EuroPMC search interface. The search query is "[ANNOTATION\_PROVIDER:'Europe PMC'] covid-19". The search results display an article titled "A class II MHC-targeted vaccine elicits immunity against SARS-CoV-2 and its variants." with a snippet: "A class II MHC-targeted vaccine elicits immunity against SARS-CoV-2 and its variants." The authors listed are Shihshih H<sup>1</sup>, Harmand J<sup>2</sup>, Rothlauf PW<sup>3</sup>, Liebeskind M<sup>3</sup>, Vakaki M<sup>3</sup>, McCaul N<sup>2</sup>, Woyet LM<sup>4</sup>, Poretti MC<sup>4</sup>, Legan M<sup>5</sup>, Jia Whelan SP<sup>2</sup>, ... (Show all 21) ... Ploegh HL<sup>7</sup>. The article is from the Proceedings of the National Academy of Sciences, 2021, 118(44), DOI: 10.1073/pnas.2116147118, PMID: 34191111. The SciLite annotations on the right include Gene Ontology (Gene Ontology, deaths (1), receptor-binding (1)), Genes/Proteins (SARS (5), RBD (1)), Diseases (infections (1)), and Organisms (severe acute respiratory syndrome, alpaca (1), mice (1)).

[New WoS July 22 Release Notes](#)

# Knowledge extraction from fulltext - Microsoft Academic - Field of study

The screenshot shows a search result for 'un sdg' on Microsoft Academic. The interface includes a filter sidebar on the left with options for 'Time' (set to 1916-2021) and 'Top Topics' (listing Sustainable development, Business, Political science, Sustainability, Computer science, Environmental economics, and Environmental resource mana...). The main content area shows a paper titled 'A Systematic Study of Sustainable Development Goal (SDG) Interactions' with 539 citations. The paper is from '2017 EARTH'S FUTURE' and lists authors: Prajal Pradhan, Luis Costa, Diego Rybski, Wolfgang Lucht, and Jürgen P. Kropp. A red box highlights a set of extracted keywords: Sustainable development, Prosperity, Poverty, Environmental economics, Set (psychology), Identification (information), Consumption (economics), Business, Significant negative correlation, and Significant positive correlation. Below the keywords, the abstract begins: 'Sustainable development goals (SDGs) have set the 2030 agenda to transform our world by tackling multiple challenges humankind is facing to ensure well-being, economic prosperity, and environmental protection. In contrast to conventional development agendas focusing on a restricted set of dimensions...'.

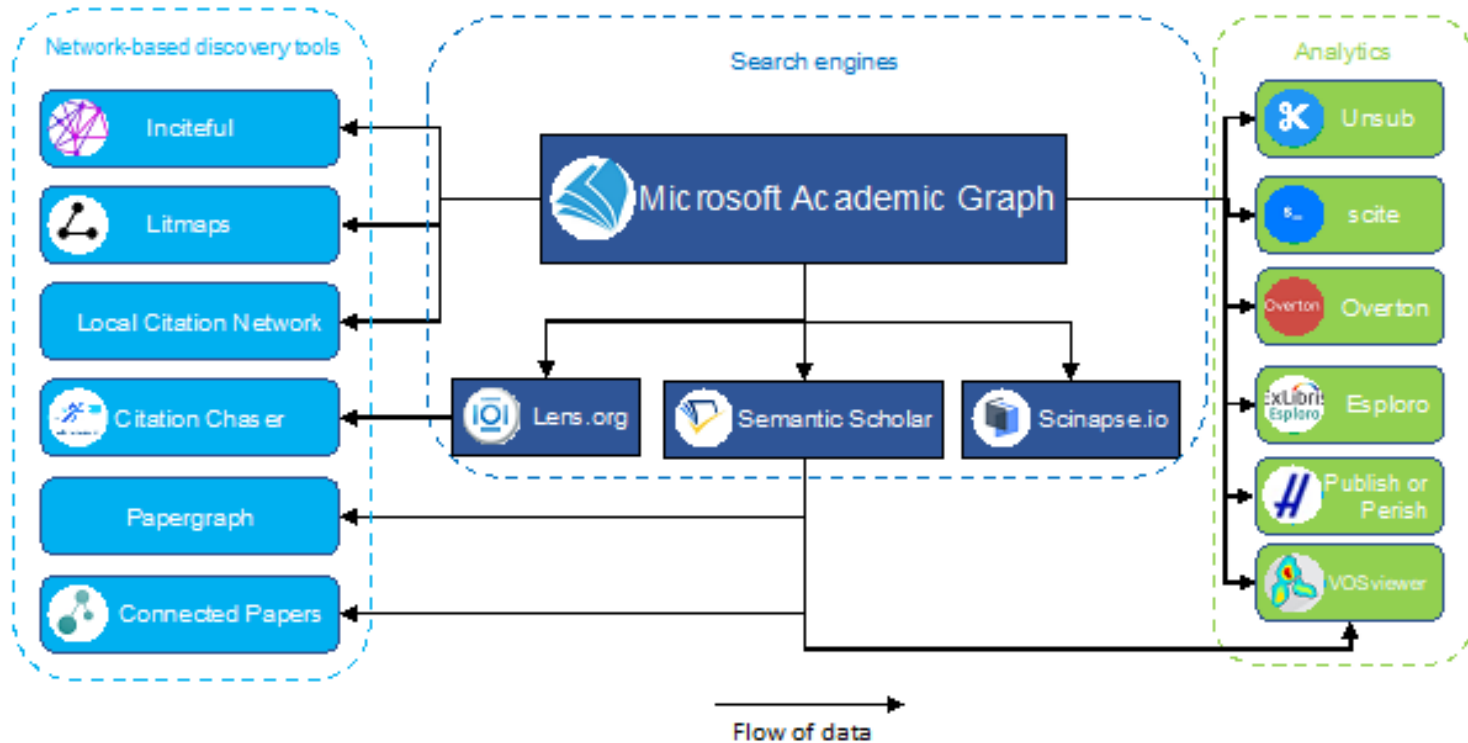
<https://europepmc.org/annotations>

# We can't duplicate this until most full-text is Open Access!

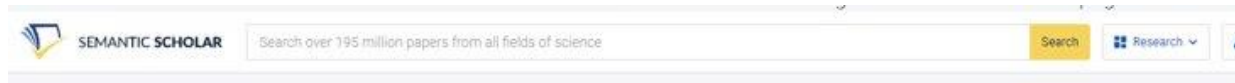


<https://academic.microsoft.com/topics/topicGraphExplorer/552854447>

# Loss of Microsoft Academic Graph



# Citation sentiments/context



## What Are TLDRs?

TLDRs (Too Long; Didn't Read) are super-short summaries of the main objective and results of a scientific paper generated using expert background knowledge and the latest GPT-3 style NLP techniques. This new feature is available in beta for more than 60 million papers on Semantic Scholar across computer science, biology, and medicine.

<https://academic.microsoft.com/topics/topicGraphExplorer/552854447>



# Citation sentiments/context



SEMANTIC SCHOLAR

"long covid" vaccination

**The Impact of COVID Vaccination on Symptoms of Long COVID. An International Survey of People with Lived Experience of Long COVID**

W. Strain, Online Sherwood, Amitava Banerjee, Vicky van der Toog, L. Hishmeh, J. Rossman · Medicine · SSRN Electronic Journal · 17 June 2021

**TLDR** The survey suggests COVID-19 vaccination may improve long COVID patients on average, and larger improvements in symptom severity scores were seen in those receiving the mRNA vaccines compared to adenoviral vector vaccines.

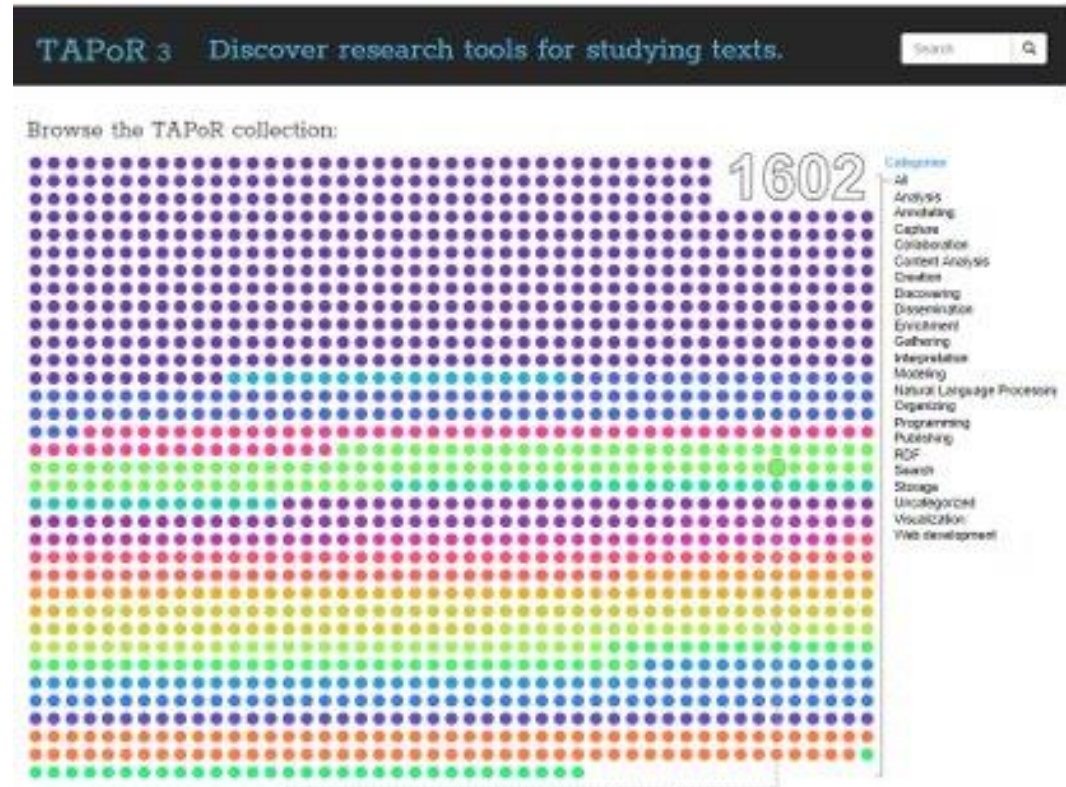
**Abstract**

**Long COVID** is a multi-system syndrome following SARS-COV-2 infection with persistent symptoms of at least 4 weeks, and frequently for several months. It has been suggested that there may be an autoimmune component. There has been an understandable caution amongst those experiencing long COVID given the potential impact of vaccination on boosting immune response. We aimed to survey people living with Long COVID evaluating the impact of their first dose vaccination on their symptoms. **Methods:** Patients with Long COVID were invited to complete a web-based questionnaire through postings on social media and direct mailing from support groups. Basic demographics, range and severity of Long COVID symptoms, which vaccine received and impact of vaccine were collated. **Results:** 900 people participated in the questionnaire, of whom 45 had pre-existing myalgic encephalomyelitis or chronic fatigue syndrome (ME/CFS) but no evidence of COVID infection and a further 43 did not complete the survey in full. The demographics and symptomology of the remaining 812 people were similar to Office of National Statistics. Following vaccination, 57.9% of participants reported improvements in symptoms, 17.9% reported deterioration and the remainder no change. There was considerable individual variation in responses. Larger improvements in symptom severity scores were seen in those receiving the mRNA vaccines compared to adenoviral vector vaccines. **Conclusions.** Our survey suggests COVID-19 vaccination may improve long COVID patients on average. The observational nature of the survey limits drawing direct causal inference, but demands validation with a randomised controlled trial. **Evolution:** No findings were reported for

<https://academic.microsoft.com/topics/topicGraphExplorer/552854447>

## Some text tool

- [Voyant](#)
- [AntConc](#)
- [TerMine](#)
- [ProAnt](#)
- [Many more](#)



[TAPoR3](#)

# Writing

## Suggest citations based on text

Keenious is a free tool that analyses your writing and shows you the most relevant research from millions of online publications, in seconds.



### Microsoft Word

Microsoft Word (2016, 2019, Online)

[GET FOR MICROSOFT WORD](#)

Then click **GET IT NOW** button on the left



### Google Docs

Requires a Google Account

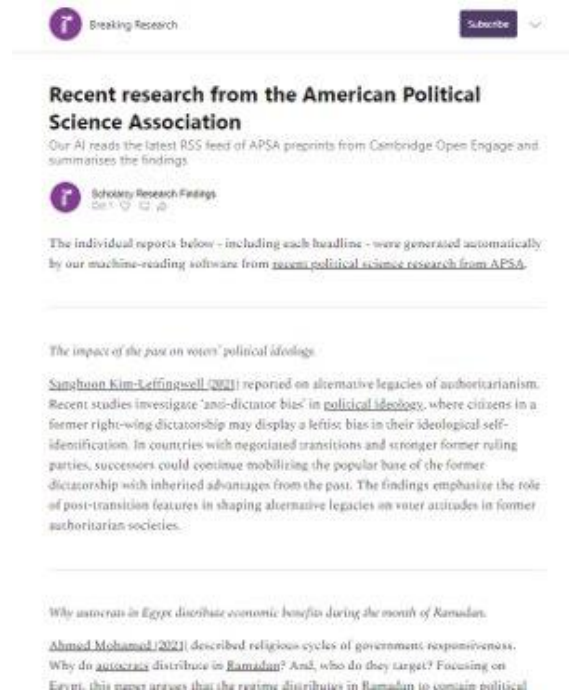
[GET FOR GOOGLE DOCS](#)

Then click **Install** button

[Keenius](#)

# Writing

## Auto-generate annotated bibliography (Scholarcy)



Breaking Research Subscribe

### Recent research from the American Political Science Association

Our AI reads the latest RSS feed of APSA preprints from Cambridge Open Engage and summarises the findings.

Scholarcy Research Findings

The individual reports below - including each headline - were generated automatically by our machine-reading software from [recent political science research from APSA](#).

---

*The impact of the past on voters' political ideology*

[Sungwon Kim-Lefringwell \(2021\)](#) reported on alternative legacies of authoritarianism. Recent studies investigate 'anti-dictator bias' in [political ideology](#), where citizens in a former right-wing dictatorship may display a leftist bias in their ideological self-identification. In countries with negotiated transitions and stronger former ruling parties, successors could continue mobilizing the popular base of the former dictatorship with inherited advantages from the past. The findings emphasize the role of post-transition features in shaping alternative legacies on voter attitudes in former authoritarian societies.

---

*Why autocrats in Egypt distribute economic benefits during the month of Ramadan*

[Ahmed Mohamed \(2021\)](#) described religious cycles of government responsiveness. Why do [autocrats](#) distribute in Ramadan? And, who do they target? Focusing on Egypt, this paper argues that the regime distributes in Ramadan to contain political

<https://scholarcy.substack.com/p/recent-research-from-the-american>

# Writing

## Auto-generate annotated bibliography (Scholarcy)

### Suggests background reading.

New to a field? Want to understand the main topics of the latest research? Scholarcy generates a background reading list helping you get up to speed. Scholarcy also highlights terms and abbreviations in the text so you can refer back to them while you are reading.

### Highlights important points.

Scholarcy's unique Robo-Highlighter™ automatically highlights important phrases and contributions made by the paper. No more printing off papers and manually going over them with a marker pen – Scholarcy's advanced AI has learnt how academic papers are written and can identify when an important point is being made.

### Creates a referenced summary.

Scholarcy summarises the whole paper with references, rewording statements in the third person, making it easier to cite the information correctly in your report, essay or thesis.

The summarisation process is fully customisable: choose the number of words, the level of highlighting and level of language variation.

### Finds the references.

No more trawling the web trying to find the papers in the references – Scholarcy does that for you, locating open-access PDFs from Google Scholar, arXiv and elsewhere. Scholarcy enlists the excellent UnPaywall API to help with this.

You can also download the entire bibliography in BibTex or .RIS format, so you can import each entry into your favourite reference management tool.

### Extracts tables and figures.

Need to check the numbers? Scholarcy finds the tables in a PDF or Word document and lets you download them in Excel format, so you can run your own calculations on the results.

Scholarcy can be configured to give you thumbnails of each figure in the PDF, cross-referenced in the text, so you can easily jump to the corresponding figure while you are reading.

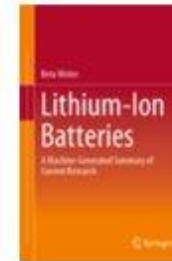
## Scholarcy features

## Writing

## Springer-Nature machine generated book

London | Heidelberg, 02 April 2019

Springer Nature published its first machine-generated book in chemistry. The book prototype provides an overview of the latest research in the rapidly growing field of lithium-ion batteries. The content is a cross-corpus auto-summarization of a large number of current research articles in this discipline. Serving as a structured excerpt from a huge set of papers, the innovative pipeline architecture aims at helping researchers to manage the information overload in this discipline efficiently.



In close collaboration between Springer Nature and researchers from Goethe University Frankfurt/Main, a state-of-the-art algorithm, the so-called *Beta Writer*, was developed to select, consume and process relevant publications in this field from Springer Nature's content platform *SpringerLink*. Based on this peer-reviewed and published content, the Beta Writer uses a similarity-based clustering routine to arrange the source documents into coherent chapters and sections. It then creates succinct summaries of the articles. The extracted quotes are referenced by hyperlinks which allow readers to further explore the original source documents. Automatically created introductions, table of contents and references facilitate the orientation within the book.

<https://www.springer.com/gp/about-springer/media/press-releases/corporate/springer-nature-machine-generated-book/16590126>


## Publication

# UNSILO Technical Checks of manuscripts

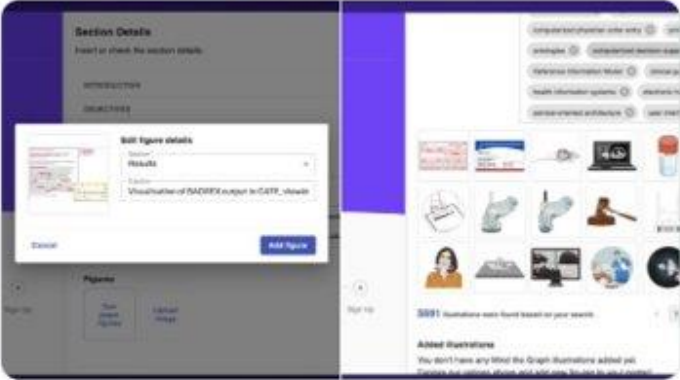
- UNSILO uses "use a combination of machine learning, rules, and natural language processing to provide editorial teams and authors with turnkey access to critical information on how well manuscripts adhere to author guidelines"
- Some checks
  - Conflicts of interest
  - Correct metadata
  - Correct use of citations and references
  - Acceptable language quality
- UNSILO Technical Checks integrated with ScholarOne, Editorial Manager, BenchPress and Manuscript Manager.

# Outreach

## Auto-generation of posters


**Scholarcy: read smarter**  
 @scholarcy

The brilliant Poster Maker from @mindthegraph uses the Scholarcy API to extract the key findings, references and figures from your paper. You can then select images from their extensive template library to generate the perfect poster for your conference.  
[mindthegraph.com/app/poster-maker...](https://mindthegraph.com/app/poster-maker)



### Other auto-generation possibilities

- Visual Abstracts
- Plan English abstracts
- Press Releases
- Video Abstracts

<https://www.scholarcy.com/combining-ai-and-visual-design-to-create-beautiful-scientific-posters/>



# **CORD-19 – A PREVIEW OF THE FUTURE? - ALLEN INSTITUTE FOR AI +WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY (OSTP) + CZI +MICROSOFT RESEARCH**

---

The Covid-19 Open Research Dataset (CORD-19) is a growing <sup>1</sup> resource of scientific papers on Covid-19 and related historical coronavirus research. CORD-19 is designed to facilitate the development of text mining and information retrieval systems over its rich collection of metadata and structured full text papers. Since its release, CORD-19 has been downloaded over 75K times and has served as the basis of many Covid-19 text mining and discovery systems. In this article, we describe the mechanics of dataset construction, highlighting challenges and key design decisions, provide an overview of how CORD-19 has been used, and preview tools and upcoming shared tasks built around the dataset. We hope this resource will continue to bring together the computing community, biomedical experts, and policy makers in the search for effective treatments and management policies for Covid-19.

## **[CORD-19: The Covid-19 Open Research Dataset](#)**

## Some observations of emerging tools

---

Diversity in origins – from startups, non-profits, individual researchers/hobbyist

---

May be wholly or partially based on open data

---

Various business models (open, free as in free beer, commercial)



## Conclusion

We are in the early days of cashing the cheque....

Open tools/services vs Tools/services based on open data

What roles should libraries and institutions play?

# Thank You!



Aarontay@gmail.com



@aarontay



<https://musingsaboutlibrarianship.blogspot.com>

Special Acknowledgement : Phil Gooch (Scholarcy) for advice on text mining applications