

Open Access enables new tools and features

Aaron Tay (Lead, Data Services)

Singapore Management University Libraries

18 May 2022

INCONECSS 2022



Aarontay@gmail.com



@aarontay



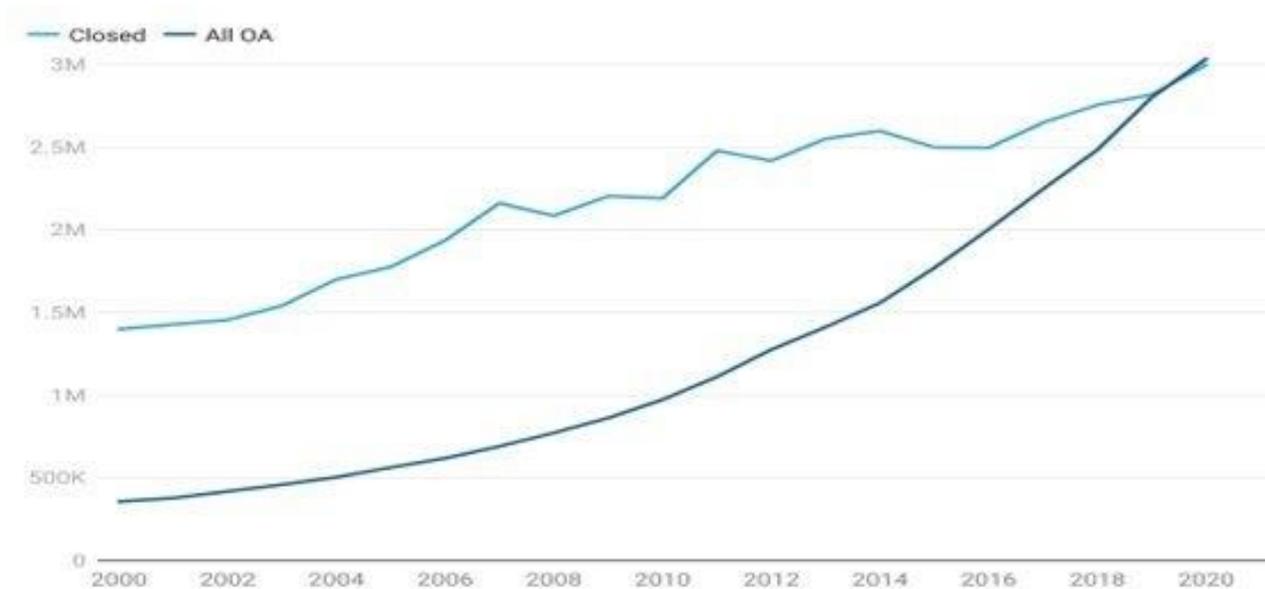
<https://musingsaboutlibrarianship.blogspot.com>

***"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**

Projections of Open Access - OA tipping point in 2020 (Dimensions) ?



Open Access surpasses subscription publication globally for the first time

Projections of Open Access (as at 2019)

In 2019:

- 31% of all journal articles are available as OA
- 52% of article views are to OA articles

Given existing trends, we estimate that by 2025:

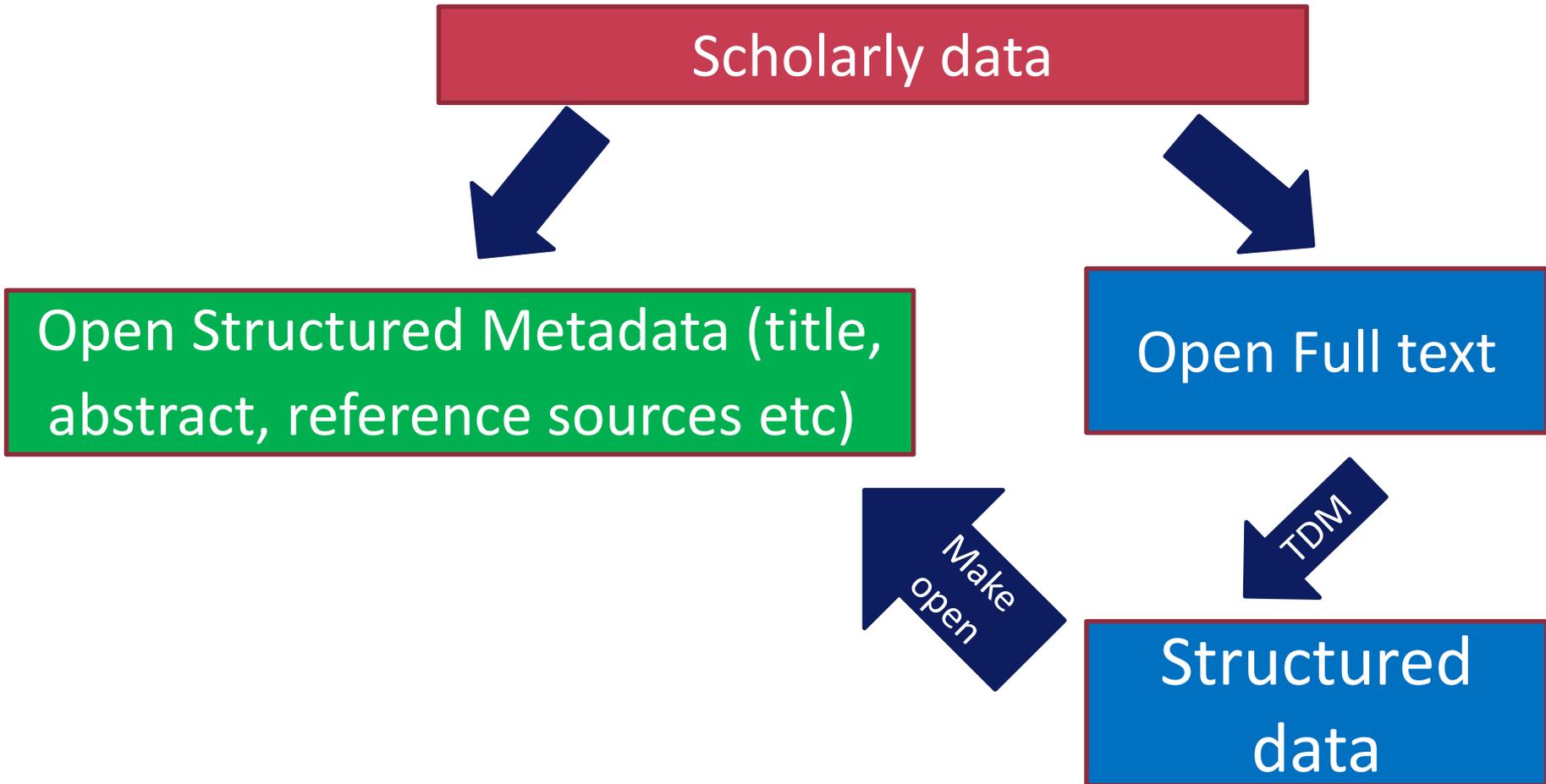
- 44% of all journal articles will be available as OA
- 70% of article views will be to OA articles

Piowar, H., Priem, J., & Orr, R. (2019). The Future of OA: A large-scale analysis projecting Open Access publication and readership. *BioRxiv*, 795310.

Other developments

- Exception for Text data mining passed into Singapore Copyright Act (Nov 2021) - Similar law in UK, Japan.
- A mandatory and non-overridable copyright exception for text and data mining (TDM) for both commercial and non-commercial activity

Open Knowledge = Open metadata+full text

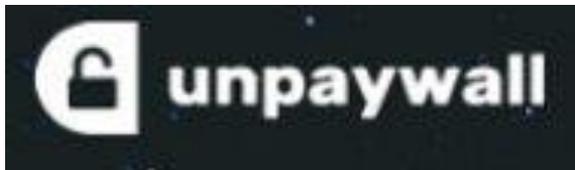
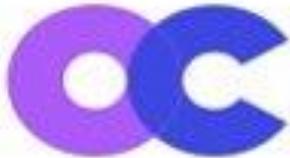


Some sources of Open Scholarly Metadata



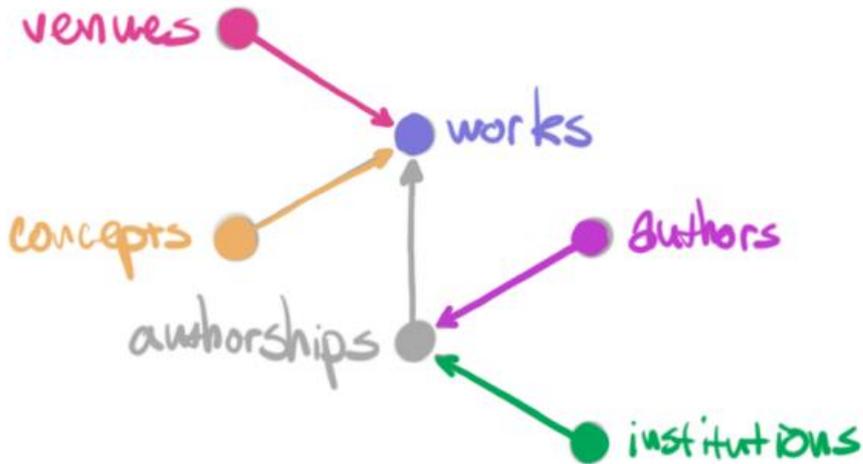
[Crossref](#), [Datacite](#), [ORCID](#), [ROR](#), [NIH](#)

Some sources of Open Scholarly Metadata (II)



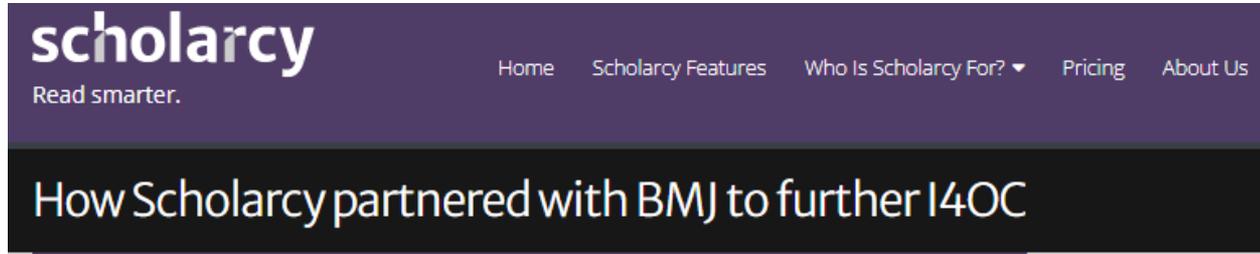
[CORE](#), [BASE](#), [Wikidata](#), [OpenCitations](#), [EuroPMC](#), [Lens.org](#), [Unpaywall](#), [Internet Archive \(Fatcat\)](#)

OpenAlex (successor to MAG) – Extracting open metadata from full-text



<https://arxiv.org/abs/2205.01833>

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-w-bmj-to-further-i4oc/>

Open Citation source/corpus*

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

[OpenAlex](#)

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

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

[Lens.org](#)

[NIH-OCC](#)

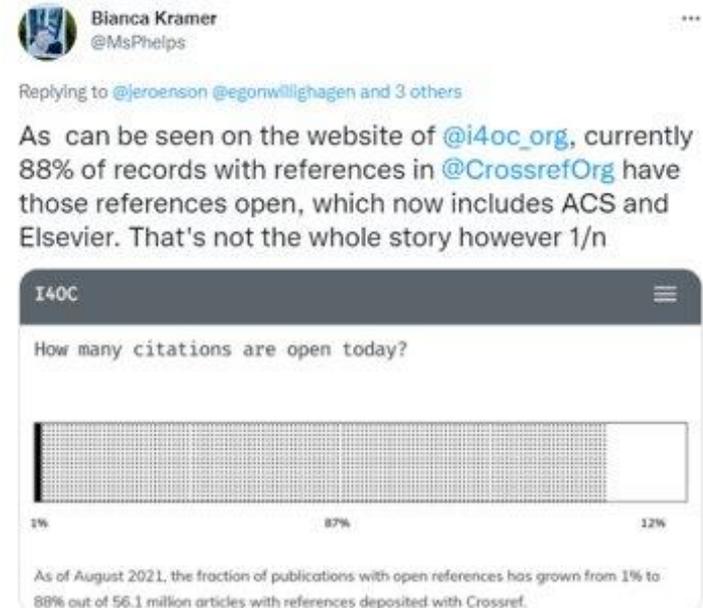
[Wikidata](#)

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

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

All major publishers* participating including

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



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



Full Text

Innovations in Scholarly Communication – Workflow tools - categorization

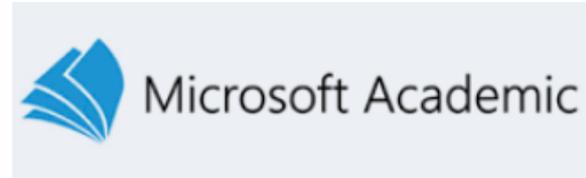


Q : How would the ready availability of open structured Scholarly metadata and full text strengthen innovation in research flow tools? Which existing tools become even better in a world of OA?

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 Disciplinary)



[See coverage of new citation indexes](#)

New Scholarly search citation indexes



Science mapping tools (for bibliometrics researchers)



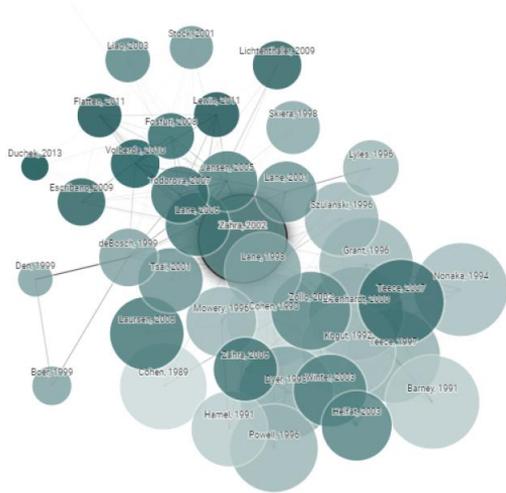
[VOSviewer](#), [Citespace](#), [Bibliometrix](#), [CitNetExplorer](#), [Sci2](#), HistCite,
 Hazing Publish or Perish

Citation based Literature mapping services (for researchers)



[See list of emerging tools](#)

New innovative tools by startups, hobbyist etc



7437 Citations, 86 References

Open in:

Origin paper

Researchers have used the absorptive capacity construct to explain various organizational phenomena. In this article we review the literature to identify key dimensions of absorptive capacity and offer a reconceptualization of this construct. Building upon the dynamic capabilities view of the firm, we distinguish between a firm's potential and realized capacity. We then advance a model outlining the conditions when the firm's potential and realized capacities can differentially influence the creation and sustenance of its competitive advantage.

[ConnectedPapers](#)

Beel & Gipp 2009 81

Google Scholar's Ranking Algorithm : An Introductory Overview

Ortega 2014 40

Academic search engines : a quantitative outlook

Wickham 2009 16248

Ggplot2: Elegant Graphics for Data Analysis

Orduña-Malea & López-Cózar 2014 29

The silent fading of an academic search engine: the case of Microsoft Academic Search
Online Information Review

These Authors

[Research Rabbit](#)

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](#)

Going beyond - Citation sentiment/context - scite example

Citation Types ⓘ

<input type="checkbox"/>	Supporting	✔	102
<input type="checkbox"/>	Mentioning	⌚	1,232
<input checked="" type="checkbox"/>	Contrasting	?	28
<input type="checkbox"/>	Unclassified	○	58

“...A cross-sectional online survey with the general population in China on mental health in the context of COVID-19 reports the prevalence of anxiety and depression to be around 32% and 34% respectively [3]. Some studies report women to be at greater risk of worse mental health outcomes in the pandemic [24 , 33], a finding which we could not confirm in our study with the exception of Adjustment Disorder. Vulnerable groups in our survey such as persons living in poorer economic circumstances and those with pre-existing physical and mental health problems showed greater concern about COVID-19, and it is likely that those population groups may also suffer from heightened psychological distress during the pandemic [5 , 34]....”

? contrasting (Confidence: 99%) [flag classification](#)

The Influence of Concern About COVID-19 on Mental Health in the Republic of Georgia: A Cross-Sectional Study

Makhashvili, Javakhishvili, Sturua *et al.* 2020

[Global Health](#)

📖 12 | ✔ 1 | ⌚ 15 | ? 0

[View full text](#) [Add to dashboard](#)

Citation context/Citation sentiment tools – What is a citation for?

Tool	Size (as of March 2022)	Citation types	Others	Further background
Semantic Scholar	200 Million papers, citation statements - unknown	<ol style="list-style-type: none"> 1. Cites Background 2. Cites Method 3. Cites Results 	Only identifies Highly Influential Citations	Help Research paper
scite	127 Million papers, citation statements 900 million	<ol style="list-style-type: none"> 1. Mentioning cite 2. Supporting cite 3. Constrasting cite 	Also detects retracted papers and allow citation statement searches	Help Research Paper
Elicit.org	Unclear data comes from Semantic Scholar	<ol style="list-style-type: none"> 1. "Possible critiques" 	This uses GPT-3 to look for possible critiques of papers from Semantic Scholar citation intents	
Web of Science	179 million papers, citation statements - unknown, is beta	<ol style="list-style-type: none"> 1. Background 2. Basis 3. Discussion 4. Support 5. Differ 	Unlike the other tools, this does citation context of references (backwards citations) rather than forward citations Pilot programme - first started in July 22 2021 notes but had the "compare" type. Feb 18 2022 release split it into "Support" & "Differ"	Details
Scholarcy	No index (as of now), users extract from full-text uploaded	<ol style="list-style-type: none"> 1. Builds on previous work 2. Differs from previous work 3. Confirms previous work 4. Counterpoints previous work 	Extractions from pdf uploaded. Unlike the other tools, this does citation context of references (backwards citations) rather than forwards citations	

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 .RS 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

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.

Elicit.org - GPT3+Semantic Scholar data- Large scale Language models are coming!

Elicit

FAQ Tasks Starred

does masks reduce covid infection?



Add a filter

Add a column (5)

.bib

.csv

Clear unstarred

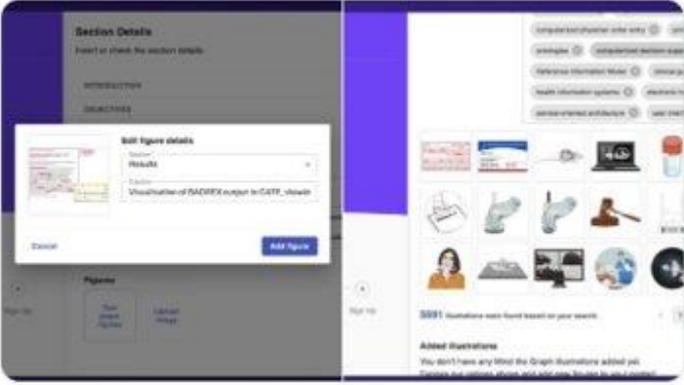
	Paper title	Year	Takeaway suggests yes/no	Effect	Outcome measured
	Face Masks Considerably Reduce Covid-19 Cases in Germany 	2020	Yes	reduced the cumulative number of registered Covid-19 cases between 2.3% and 13% over 1 period of 10 days after they became compulsory	daily growth rate of reported i
	High-Quality Masks Reduce COVID-19 Infections and Deaths in the US	2020	Yes	decreased the cumulative infection attack rate, peak prevalence, and population mortality	infections
	High-Quality Masks Reduce Covid-19 Infections and Death in the US	2021	Yes	decreased the cumulative infection attack rate, the peak prevalence, and the population mortality	COVID-19
	Face Masks Considerably Reduce Covid-19 Cases in Germany: A Synthetic Control Method Approach	2020	Yes	reduced the cumulative number of registered Covid-19 cases between 2.3% and 13% over 1 period of 10 days after they became compulsory	COVID-19 cases

Outreach

Auto-generation of posters (DELL-E2)


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/>

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)

The impact of AI/Machine learning on Scholarly research

Lit review: Explore research questions ▾

Does creatine improve cognition?

☆ Creatine may improve cognitive function and slow or prevent cognitive decline. ▾

Metabolic Agents that Enhance ATP can Improve Cognitive Functioning. A Review of the Evidence for Glucose, Oxygen, Pyruvate, Creatine, and L-Carnitine

101 citations (3 highly influential) · 2011 · Review · Yes

L. Owen, Sande-Rona Sunram-Lea · Nutrients

PDF · DOI · Semantic Scholar

Over the past four or five decades, there has been increasing interest in the neurochemical regulation of cognition. This field received considerable attention in the 1980s, with the identification of possible cognition enhancing agents or “smart drugs”. Even though many of the optimistic claims for some agents have proven premature, evidence suggests that several metabolic agents may prove to be effective in improving and preserving cognitive performance and may lead to better cognitive aging through the lifespan. Aging is characterized by a progressive deterioration in physiological functions and metabolic processes. There are a number of agents with the potential to improve metabolic activity. Research is now beginning to identify these various agents and delineate their potential usefulness for improving cognition in health and disease. This review provides a brief overview of the metabolic agents glucose, oxygen, pyruvate, creatine, and L-carnitine and their beneficial effects on cognitive function. These agents are directly responsible for generating ATP (adenosine triphosphate) the main cellular currency of energy. The brain is the most metabolically active organ in the body and as such is particularly vulnerable to disruption of energy resources. Therefore interventions that sustain adenosine triphosphate (ATP) levels may have importance for improving neuronal dysfunction and loss. Moreover, recently, it has been observed that environmental conditions and diet can affect transgenerational gene expression via epigenetic mechanisms. Metabolic agents might play a role in regulation of nutritional epigenetic effects. **In summary, the reviewed metabolic agents represent a promising strategy for improving cognitive function and possibly slowing or preventing cognitive decline.**

GPT-3 wrote this claim

After “reading” this abstract

Jungwon

@jungofthewon

⋮

1/ Lots of interest lately in making language models “truthful”. How can we prevent GPT-3 from “lying”?

We’ve worked on this in the context of [@elicitorg](#). In Elicit, GPT-3 tries to answer your research question given abstracts from papers.

(Can try at [elicit.org](#))

4:22 AM · Dec 10, 2021 · Twitter Web App

7 Retweets
2 Quote Tweets
29 Likes

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>

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