

# Current Strategies for Updating Systematic Reviews

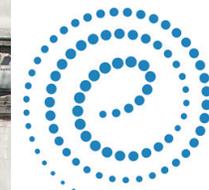
A webcast hosted by AIR's Center on KTDRR  
January 26, 2022

Center on  
**KNOWLEDGE TRANSLATION FOR  
DISABILITY & REHABILITATION RESEARCH**

at American Institutes for Research ■

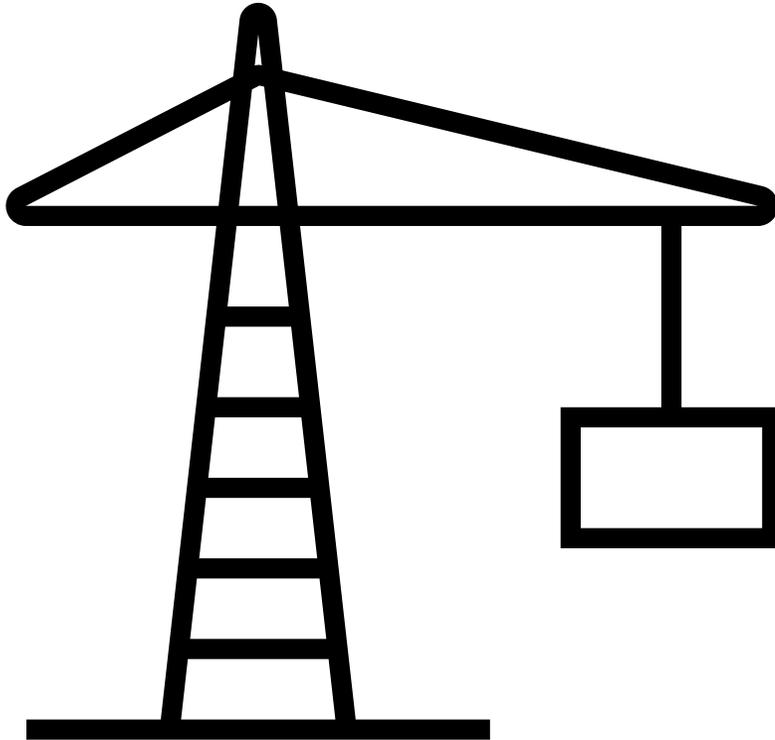
# Using New Tools and Technologies to Maintain Our Living Map of COVID-19 Research

James Thomas



**EPPi Centre**  
Evidence for  
Policy & Practice

# Under construction



# Outline

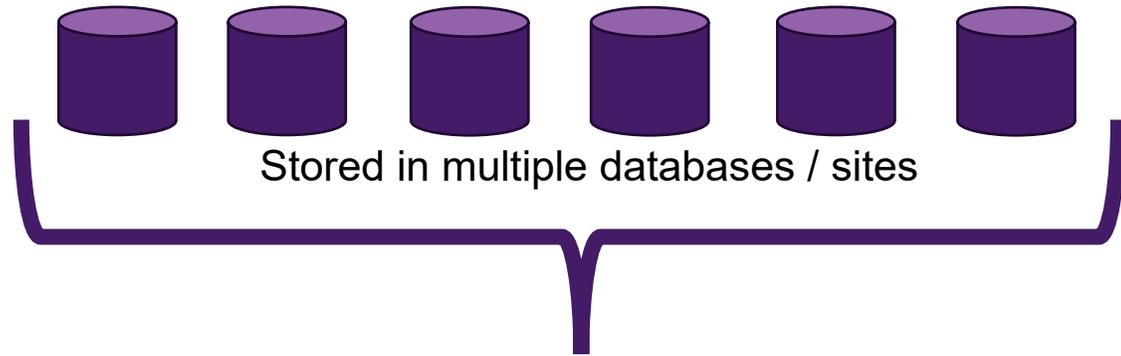
- A new search paradigm for updating systematic reviews and maps?
- Case study: the COVID-19 map of research maintained by EPPI-Centre, University College London and CRD, University of York

# A new search paradigm?

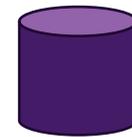
- Two enablers:
  - Increasingly open bibliographic data
  - New automation technologies and tools



All research reports



Boolean searches retrieve most likely relevant into local database



Records deduplicated and manually screened for eligibility



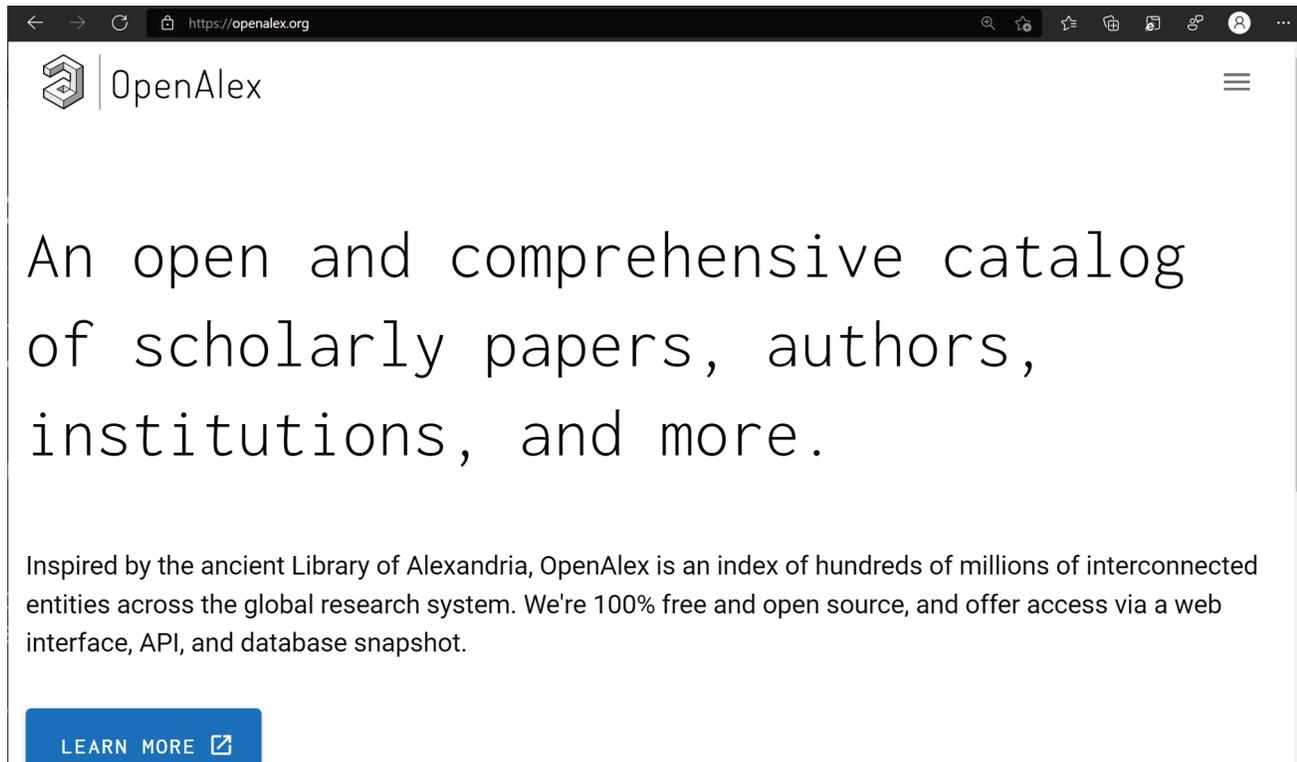
Finding eligible studies



# Why is it like this?

- Various organisations have indexed research in their area.
- It costs money to index research.
  - The business model (thus far) has been to charge for access to bibliographic databases.
  - There's been financial benefit to keeping a monopoly hold over the content of databases (i.e. you can't find a given set of records by searching elsewhere).
- This has led to the picture we see of numerous, subject-specific databases jealously guarding their content (and charging £££ for access).

# Movement towards open and more comprehensive sources

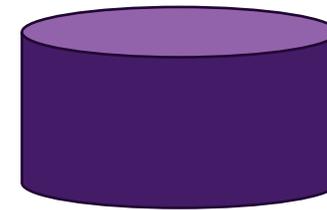


The screenshot shows the OpenAlex website homepage. The browser address bar displays "https://openalex.org". The page features the OpenAlex logo, a navigation menu icon, and a main heading: "An open and comprehensive catalog of scholarly papers, authors, institutions, and more." Below this is a paragraph: "Inspired by the ancient Library of Alexandria, OpenAlex is an index of hundreds of millions of interconnected entities across the global research system. We're 100% free and open source, and offer access via a web interface, API, and database snapshot." At the bottom left, there is a blue button labeled "LEARN MORE" with an external link icon.

- Open Alex / CrossRef / Dimensions... (Google Scholar to some extent) are publishing bibliographic data at scale.
- Microsoft Academic – now discontinued – played a major role.
- Bibliographic data are becoming *commoditized* (you can get the same content from multiple sources).
- You can ingest (essentially) ALL the world's published research into a local database.

**Commoditized  
bibliographic  
data can be  
aggregated /  
stored by  
anyone**

All research reports



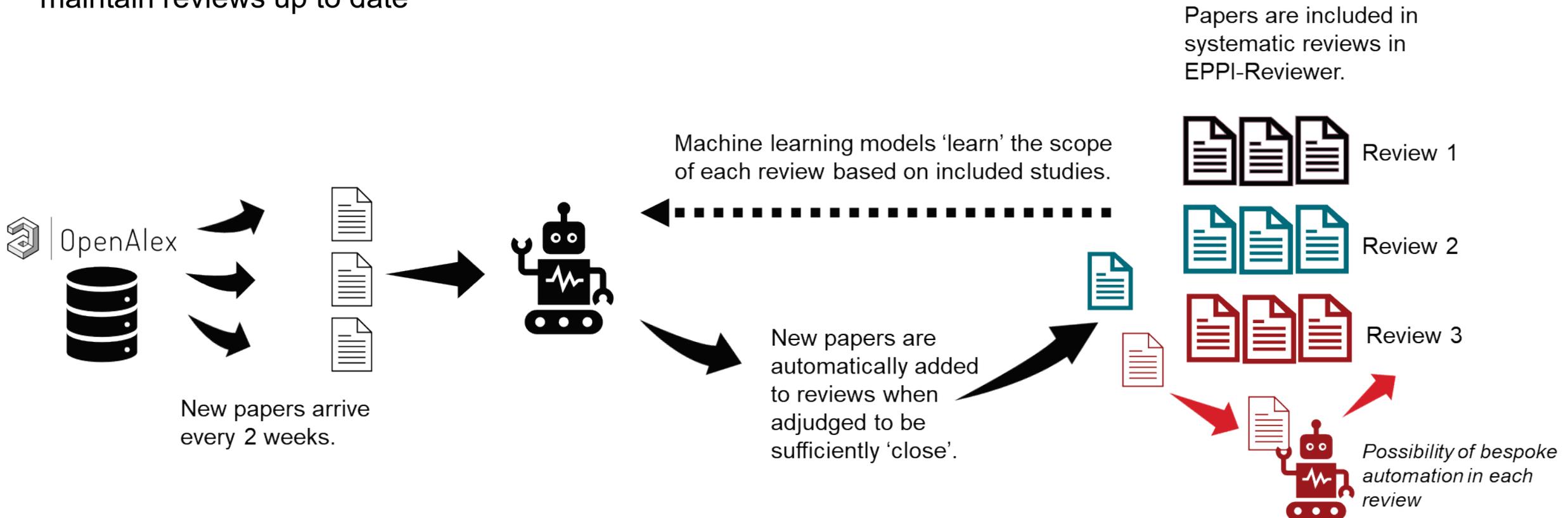
Stored in comprehensive repositories



What is the best way to find eligible studies?

# Continuous update of reviews in EPPI-Reviewer

Main aim: to maintain a 'surveillance' of the literature as it emerges to maintain reviews up to date



Papers are included in systematic reviews in EPPI-Reviewer.

Machine learning models 'learn' the scope of each review based on included studies.

New papers are automatically added to reviews when adjudged to be sufficiently 'close'.

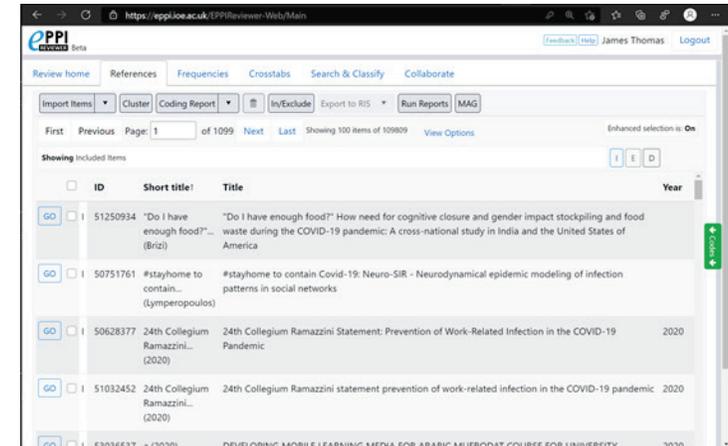
*Possibility of bespoke automation in each review*

# Our case study

- The EPPI-Reviewer platform contains a local copy of the OpenAlex\* dataset
- It's updated every 2 weeks
- Between 400k and 2m new records arrive each update
- We can find research using machine learning and Boolean searches

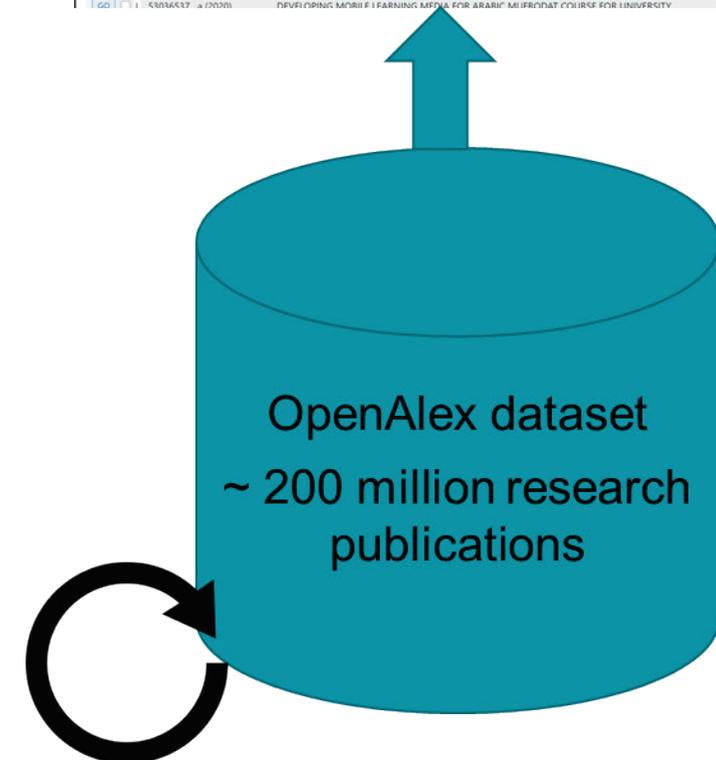
How can we use this new resource to maintain our COVID-19 map of research?

\*We originally started this work using Microsoft Academic which closed at the end of 2021. OpenAlex builds on Microsoft Academic and aims to be a 'drop-in' replacement.



The screenshot shows the EPPI-Reviewer web interface. The URL is https://eppi.ioe.ac.uk/EPPIReviewer-Web/Main. The interface includes a navigation menu with options like 'Review home', 'References', 'Frequencies', 'Crosstabs', 'Search & Classify', and 'Collaborate'. Below the menu, there are buttons for 'Import Items', 'Cluster', 'Coding Report', 'In/Exclude', 'Export to RIS', 'Run Reports', and 'MAG'. A pagination bar indicates 'Page: 1 of 1099' and 'Showing 100 items of 10989'. The main content area displays a table of 'Showing included items' with columns for 'ID', 'Short title', 'Title', and 'Year'. The table contains several rows of research entries, including one about 'Do I have enough food?' and another about '24th Collegium Ramazzini Statement: Prevention of Work-Related Infection in the COVID-19 Pandemic'.

ID	Short title	Title	Year
51250934	"Do I have enough food?..." (Briz)	"Do I have enough food? How need for cognitive closure and gender impact stockpiling and food waste during the COVID-19 pandemic: A cross-national study in India and the United States of America	
50751761	#stayhome to contain... (Lymperopoulos)	#stayhome to contain Covid-19: Neuro-SIR - Neurodynamical epidemic modeling of infection patterns in social networks	
50628377	24th Collegium Ramazzini... (2020)	24th Collegium Ramazzini Statement: Prevention of Work-Related Infection in the COVID-19 Pandemic	2020
51032452	24th Collegium Ramazzini... (2020)	24th Collegium Ramazzini statement prevention of work-related infection in the COVID-19 pandemic	2020
51356417	2020	DEVELOPING MORNING FARMING MEDIA FOR ARABIC MIFRODAT COURSE FOR UNIVERSITY	2020



Updated every two weeks

# Overarching research questions



Are all the studies we need for systematic reviews indexed in OpenAlex (MAG)?



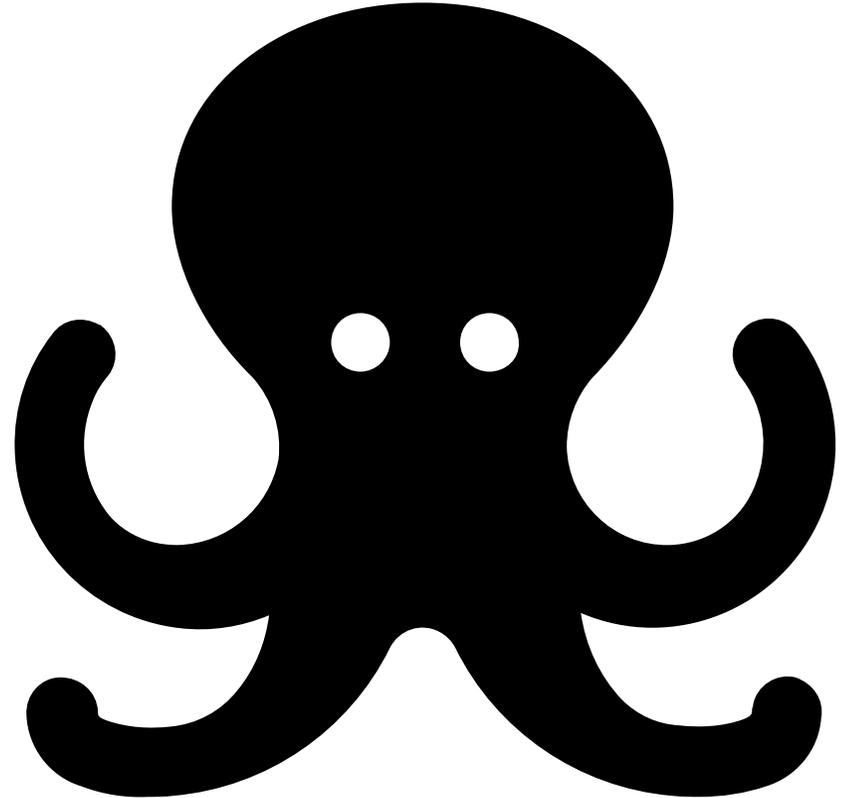
Can we efficiently identify the studies we need for systematic reviews using OpenAlex (MAG)?



# The Octopus Study

---

- Aim: to undertake a cost-effectiveness study of different options for maintaining the COVID-19 map
- Lots of data in map; added detailed time-on-task data for team
- Collected data for 4 weeks of map updates (June / July 2020)
  - Compared
    - Embase / PubMed search sources
    - MAG / OpenAlex as a single search source
    - Different machine learning options



# 8-arm CEA: Components of study identification workflows

Study Arm	Intervention / Comparator	Automation in Workflow	Boolean MEDLINE + Embase Searches	MEDLINE-Embase De-duplication	Automated MAG Update Search	Binary ML Classifier	Manual Screening	Priority Screening	Fixed Screening Target N = 1,500 records	Target Recall
1	Comparator A		●	●			●			1.0
2	Comparator B		●	●			●			0.95
3	Comparator C		●	●			●		●	0.95
4	Intervention A	●	●	●		●	●			0.95
5	Intervention B	●	●	●		●	●	●	●	0.95
6	Intervention C	●			●		●			1.0
7	Intervention D	●			●	●	●			0.95
8	Intervention E	●			●	●	●	●	●	0.95

# Octopus study: results

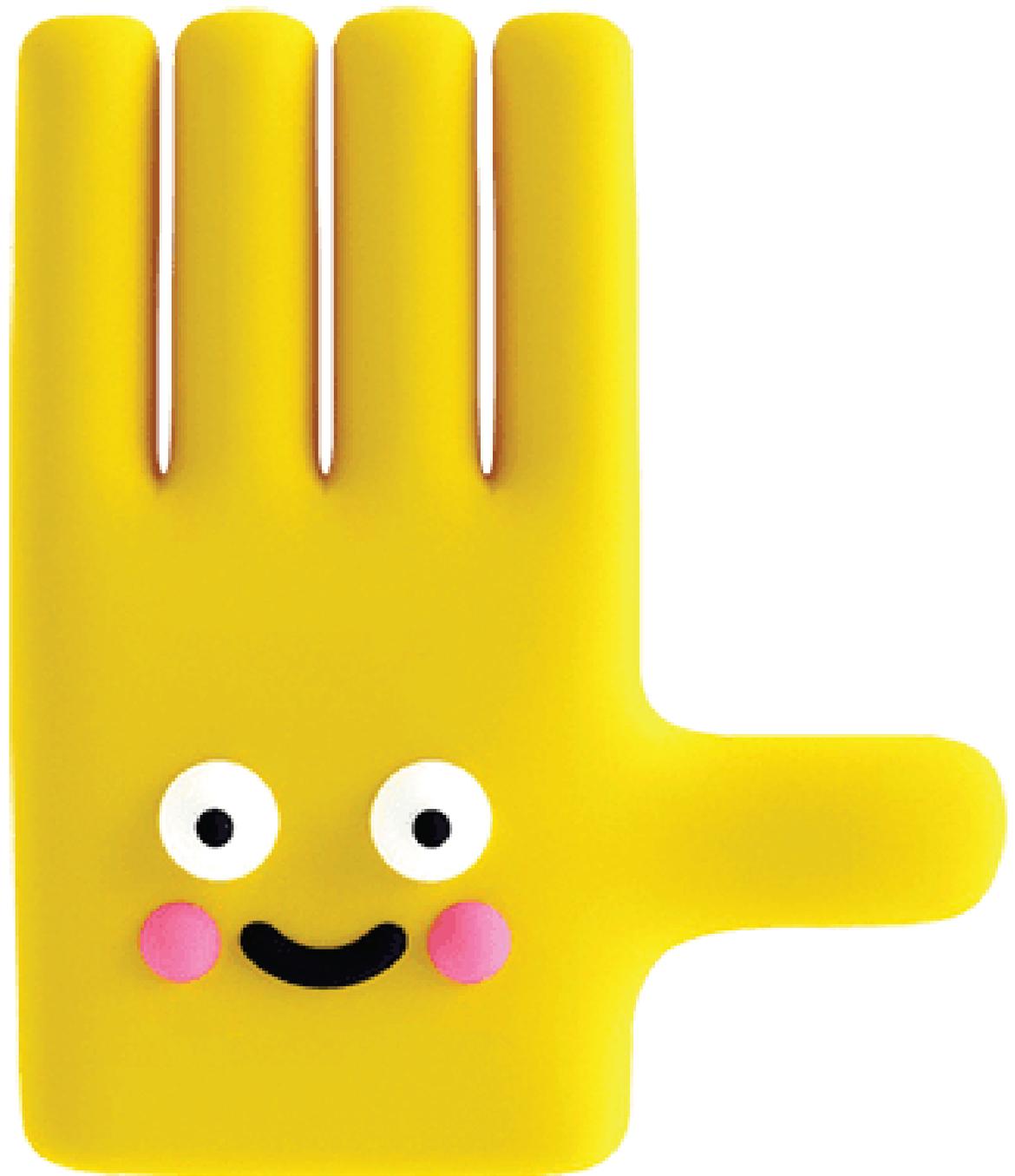
	Arm 1	Arm 2	Arm 3	Arm 4	Arm 5	Arm 6*	Arm 7*	Arm 8*
	Comparator A	Comparator B	Comparator C	Intervention A	Intervention B	Intervention C	Intervention D	Intervention E
Gold standard included	4378	4378	4378	4378	4378	4378	4378	4378
Medline-Embase unique included	65	65	65	65	65	65	65	65
MAG unique included	743	743	743	743	743	743	743	743
Medline-Embase + MAG common included	3570	3570	3570	3570	3570	3570	3570	3570
Screened-coded	9180	8722	6000	6315	6000	8639	7898	6000
Included	3635	3455	2392	3468	3441	4313	4104	4104
Target recall vs Source(s) included	1	0.95	0.95	0.95	0.95	1	0.95	0.95
Recall vs Source(s) included	1.00	0.95	0.66	0.95	0.95	1.00	0.95	0.95
<b>Cumulative Recall vs Gold standard included</b>	<b>0.83</b>	<b>0.79</b>	<b>0.55</b>	<b>0.79</b>	<b>0.79</b>	<b>0.99</b>	<b>0.94</b>	<b>0.94</b>
<b>Cumulative Precision vs Gold standard included</b>	<b>0.40</b>	<b>0.40</b>	<b>0.40</b>	<b>0.55</b>	<b>0.57</b>	<b>0.50</b>	<b>0.52</b>	<b>0.68</b>

For full results, please see <https://wellcomeopenresearch.org/articles/6-210/v1>.

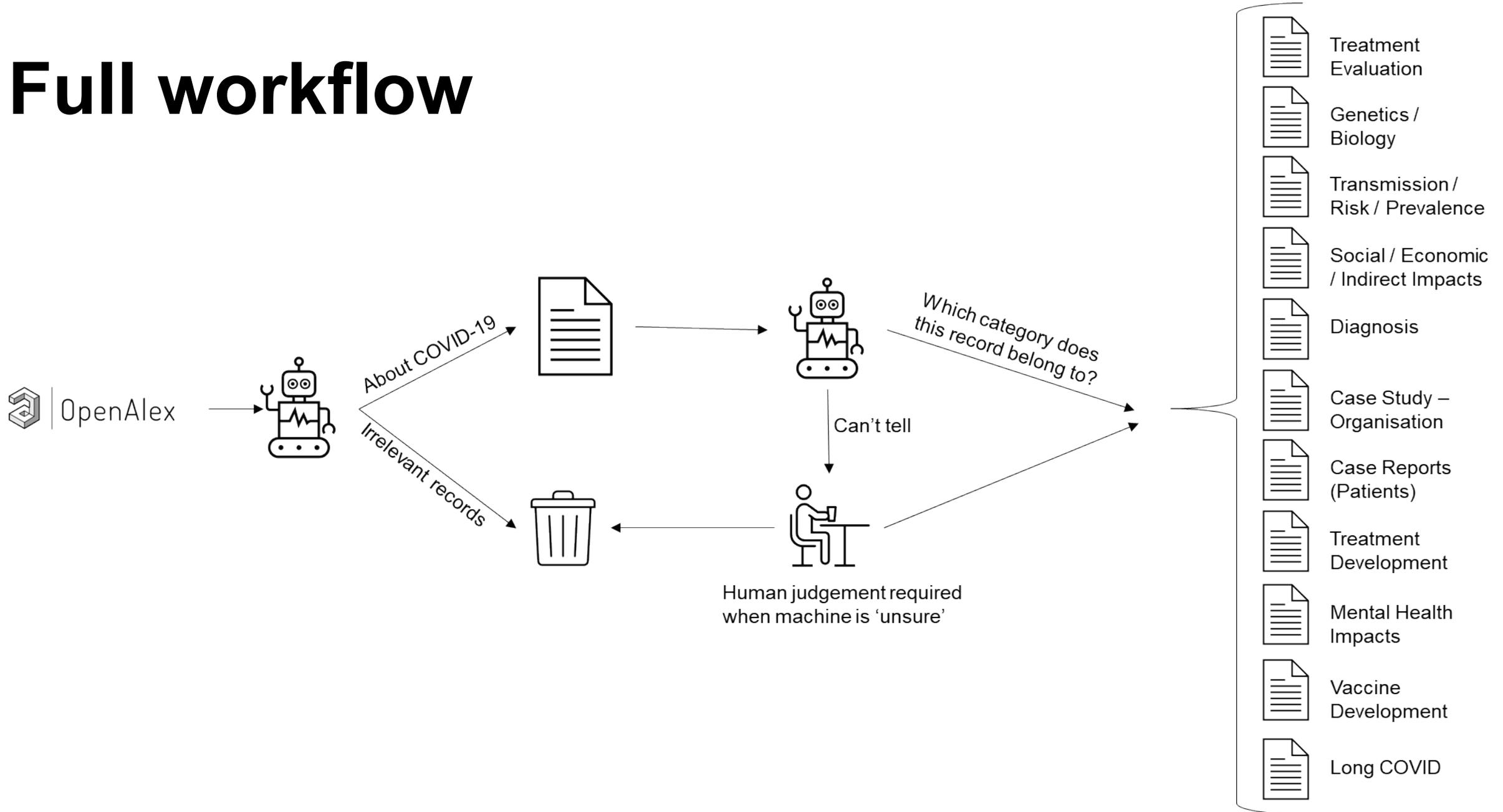
# Further automation ...

---

- The map contains ~100k records manually categorized into 12 domains (diagnosis, health impacts, treatment evaluation...)
- We have plenty of training data for more refined machine learning
- We now use a BERT model in addition to the previous machine learning to put research into categories automatically
- A second model identifies research on long COVID

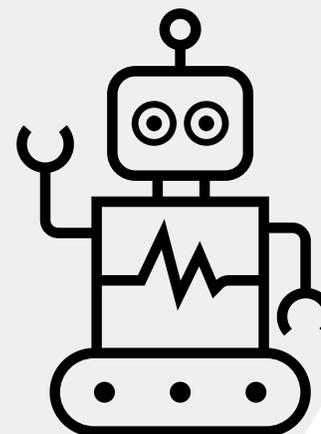
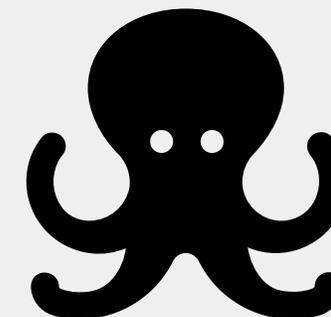


# Full workflow



# Conclusions

- We conducted a detailed study examining the costs and benefits of two different search sources together with automation options.
- Our traditional search sources had lower recall than expected.
- OpenAlex / MAG contains more research that is not indexed in Embase / PubMed than expected.
- Thanks to a lot of training data, we can also classify a large proportion of records automatically.
- With automation, we can obtain a higher recall at lower cost than our initial workflow.
- We are currently evolving the workflow with a view to maximising the efficiency gains from automation.



## Thank you

James Thomas

EPPI-Centre website: <http://eppi.ioe.ac.uk>

Email:

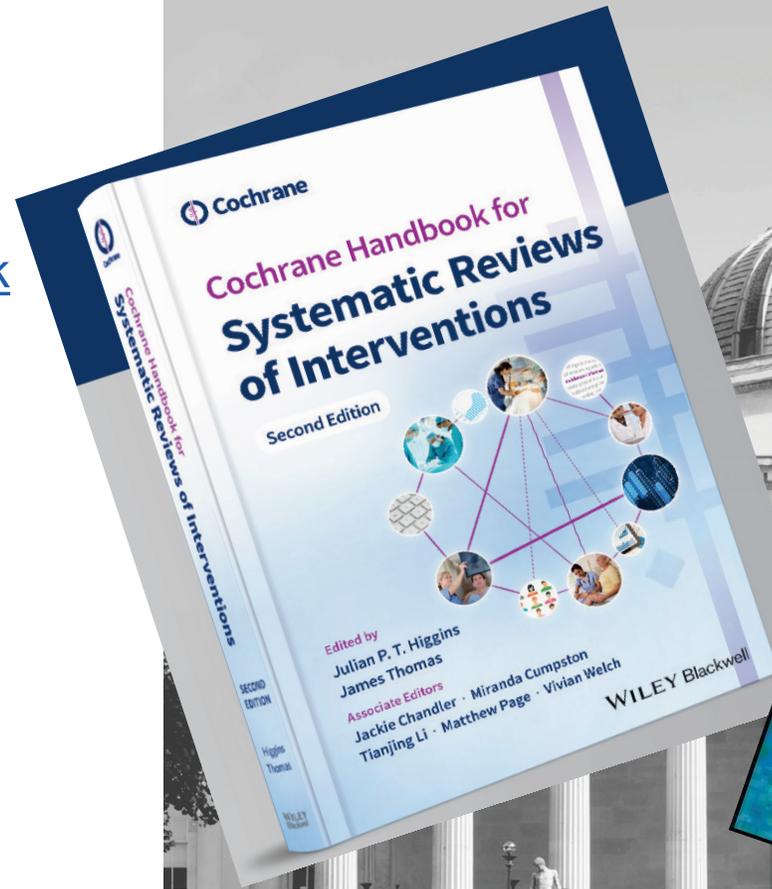
[james.thomas@ucl.ac.uk](mailto:james.thomas@ucl.ac.uk)

Twitter:

James\_M\_Thomas

**EPPI-Centre**  
Social Science Research Unit  
Institute of Education  
University of London  
18 Woburn Square  
London WC1H 0NR

Tel +44 (0)20 7612 6397  
Fax +44 (0)20 7612 6400  
Email [eppi@ioe.ac.uk](mailto:eppi@ioe.ac.uk)  
Web [eppi.ioe.ac.uk/](http://eppi.ioe.ac.uk/)



# Thank You!

## Evaluation:

<https://survey.alchemer.com/s3/6701407/Eval-UpdatingSR>

## KTDRR Webcasts:

<https://ktdrr.org/training/webcasts/index.html>

 [www.ktdrr.org](http://www.ktdrr.org)

 [ktdrr@air.org](mailto:ktdrr@air.org)

 4700 Mueller Blvd, Austin, TX 78723

 800.266.1832

Center on  
**KNOWLEDGE TRANSLATION FOR  
DISABILITY & REHABILITATION RESEARCH**

at American Institutes for Research ■