Current Strategies for Updating Systematic Reviews

A webcast hosted by AIR's Center on KTDRR

January 26, 2022

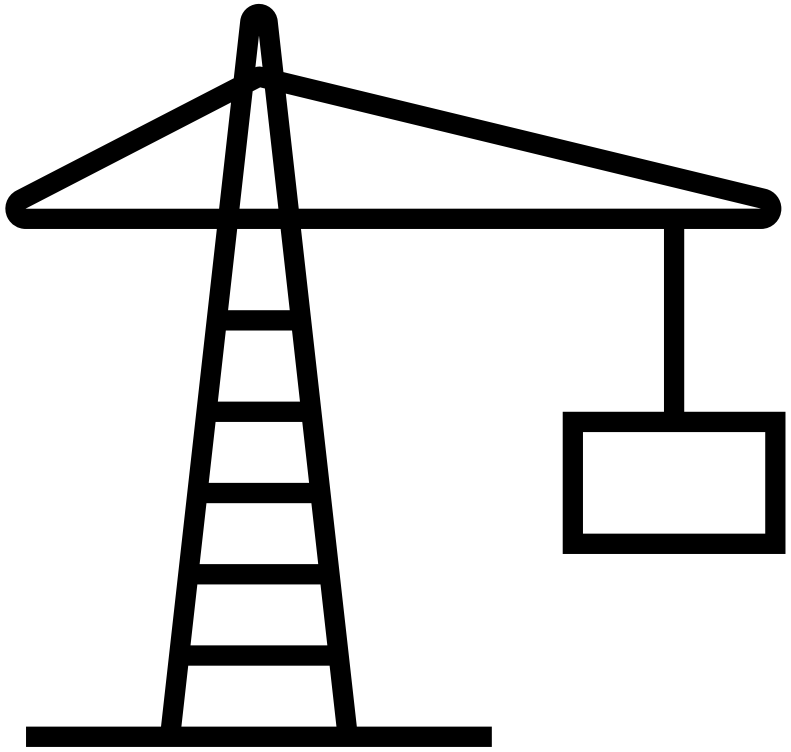
Center on
KNOWLEDGE TRANSLATION FOR
DISABILITY & REHABILITATION RESEARCH
at American Institutes for Research

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

James Thomas

EPPI Centre
Evidence for Policy & Practice

Under construction



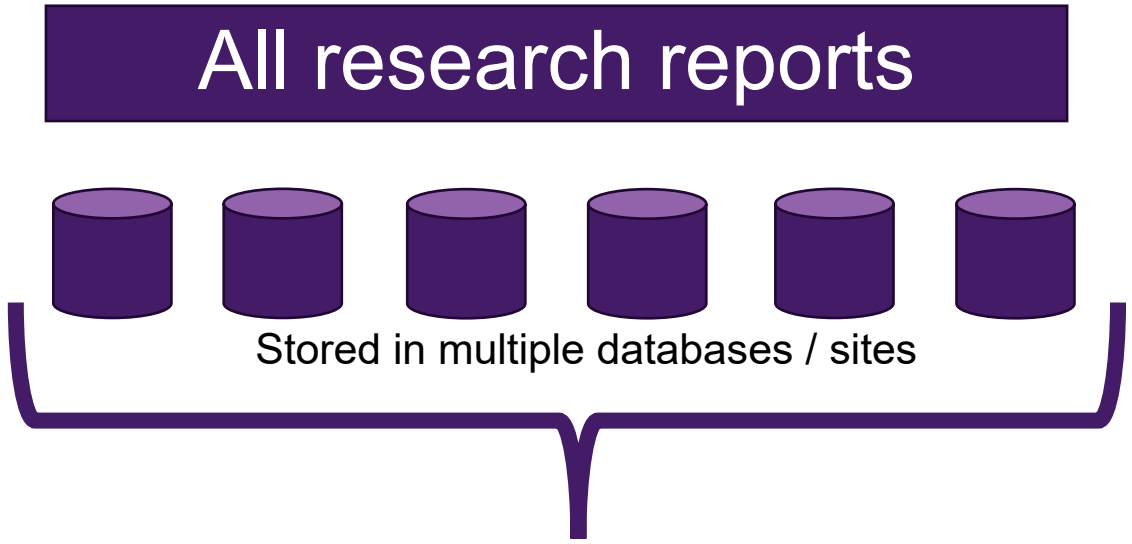
Outline

- 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

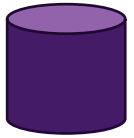
A new search paradigm?

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

Finding eligible studies



Boolean searches retrieve most likely relevant into local database

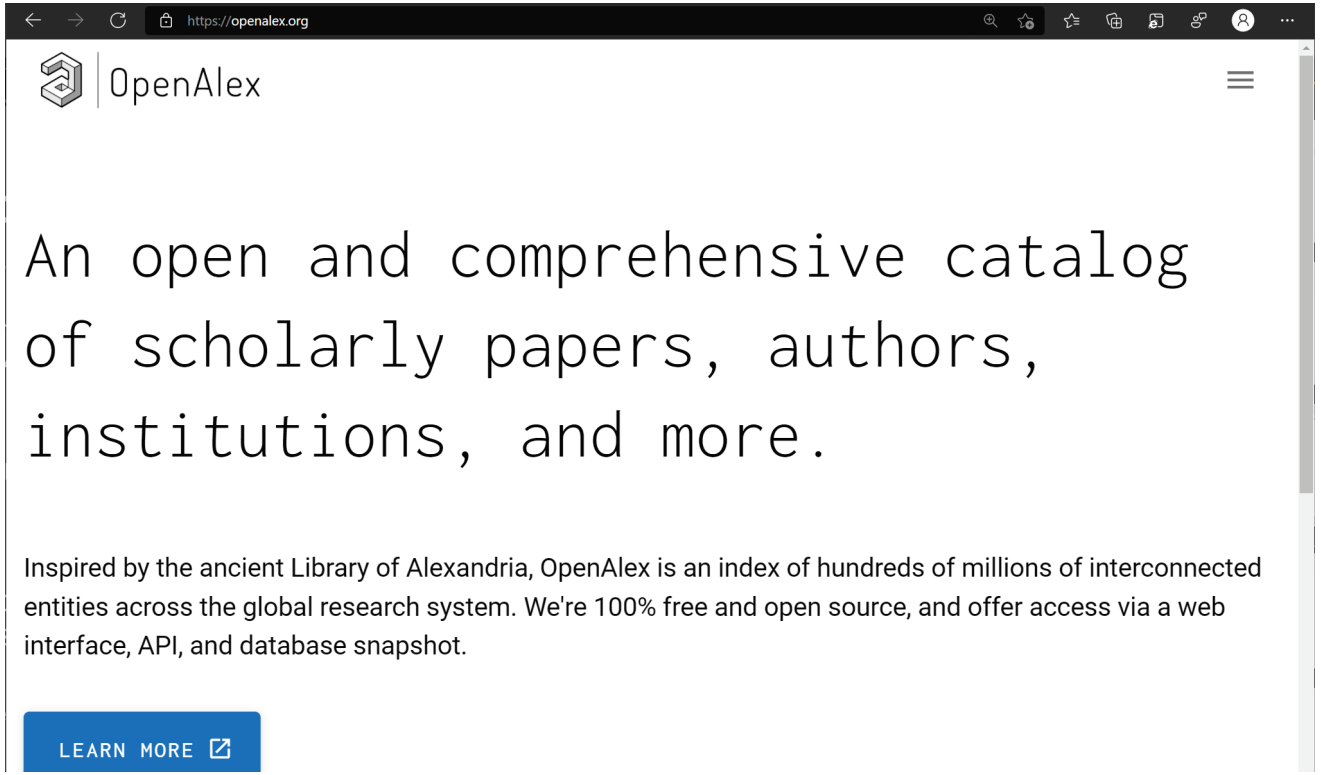


Records deduplicated and manually screened for eligibility

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).

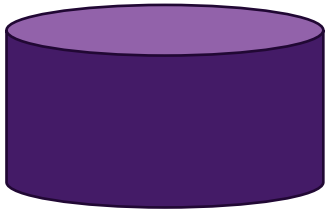
Movement towards open and more comprehensive sources



- 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

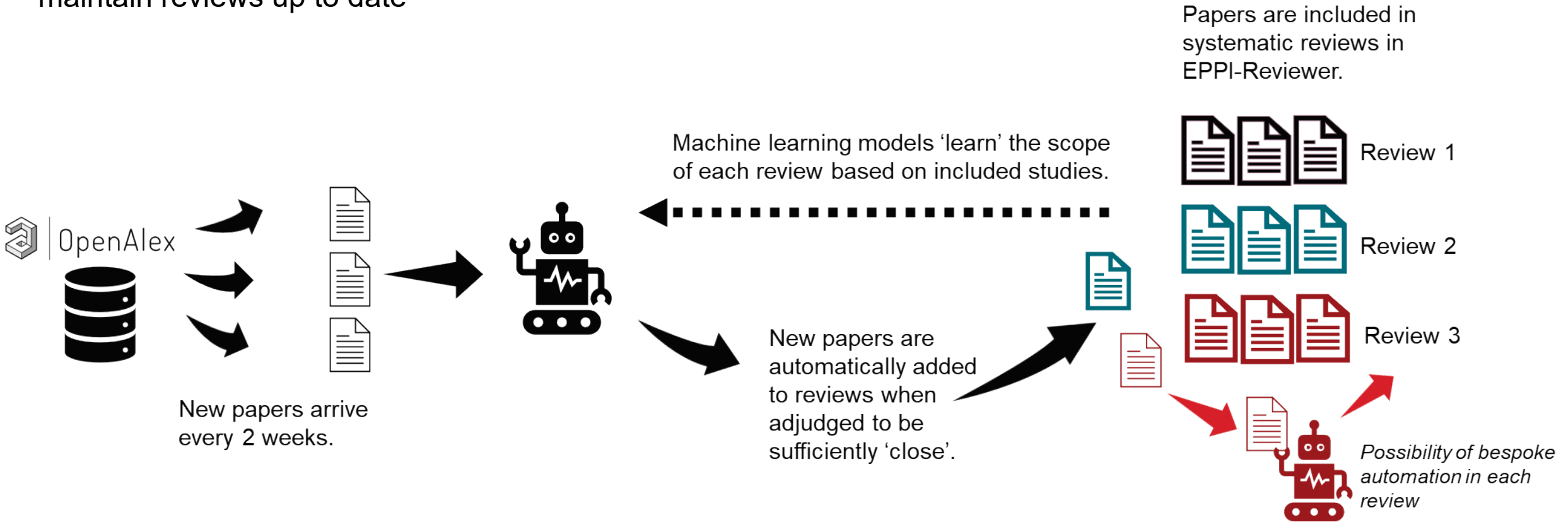


Stored in comprehensive repositories

What is the best way to find eligible studies?

Continuous update of reviews in EPPI-Reviewer

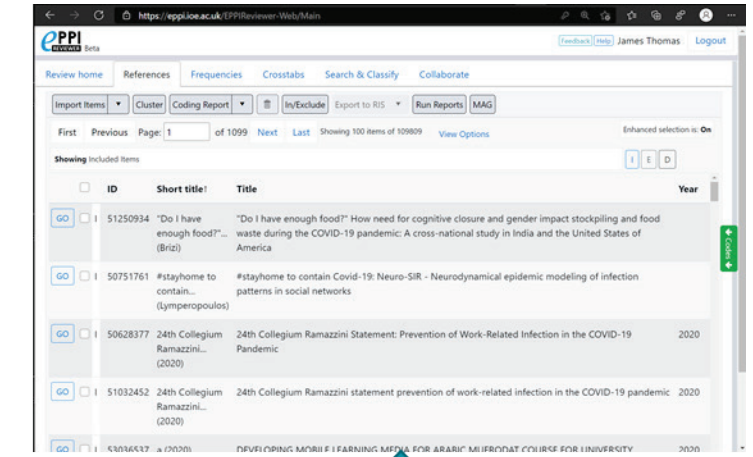
Main aim: to maintain a ‘surveillance’ of the literature as it emerges to maintain reviews up to date

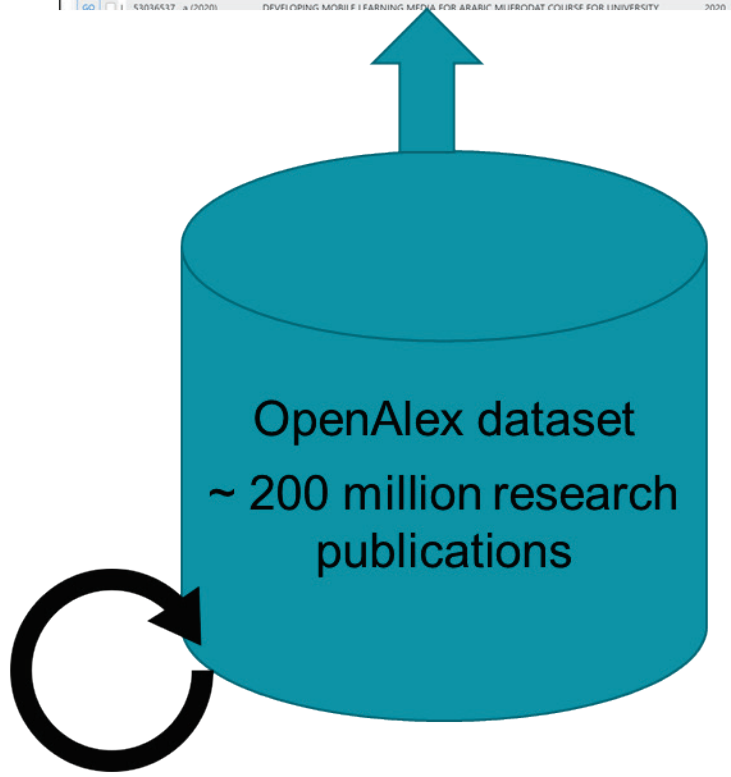


Our case study

- 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.

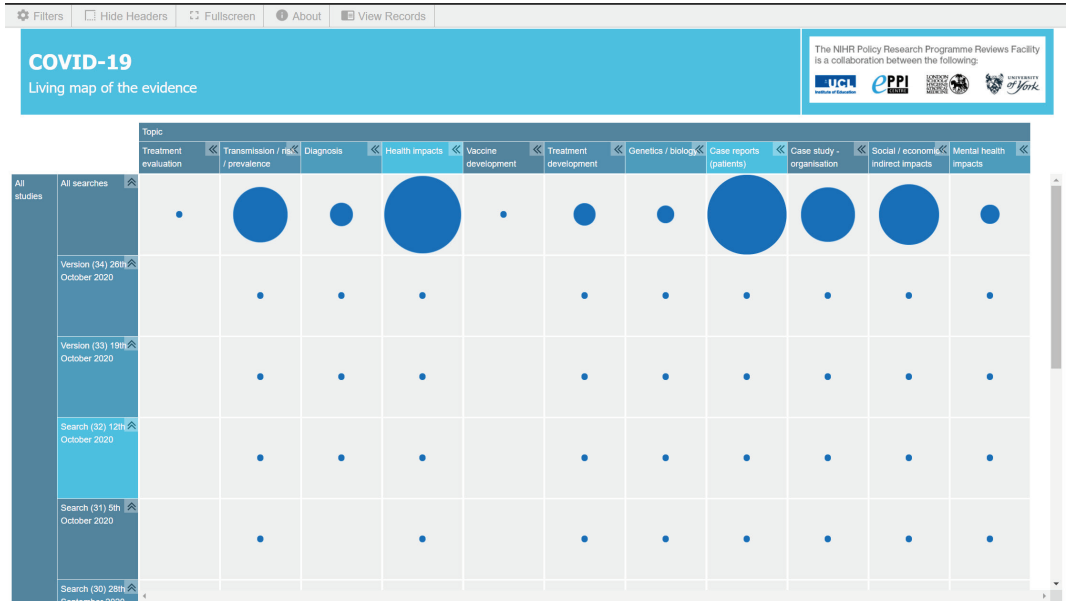
Overarching research questions

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)?

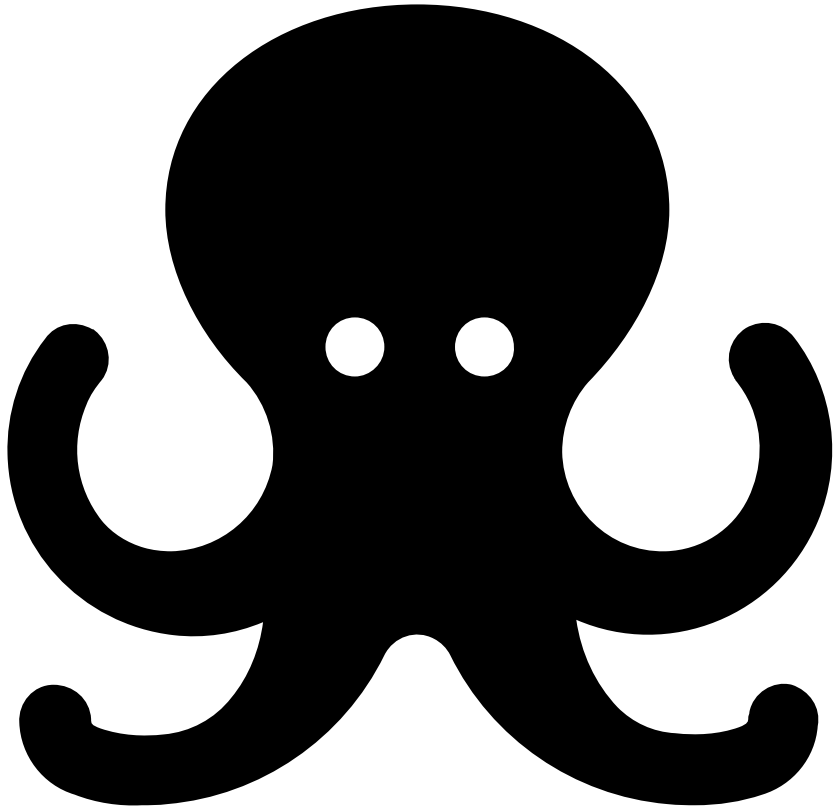
COVID-19 map of research activity

- Updated weekly
- To start with, was populated via searches of Embase and PubMed
- Comprehensive
- Large team
- A lot of work!



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

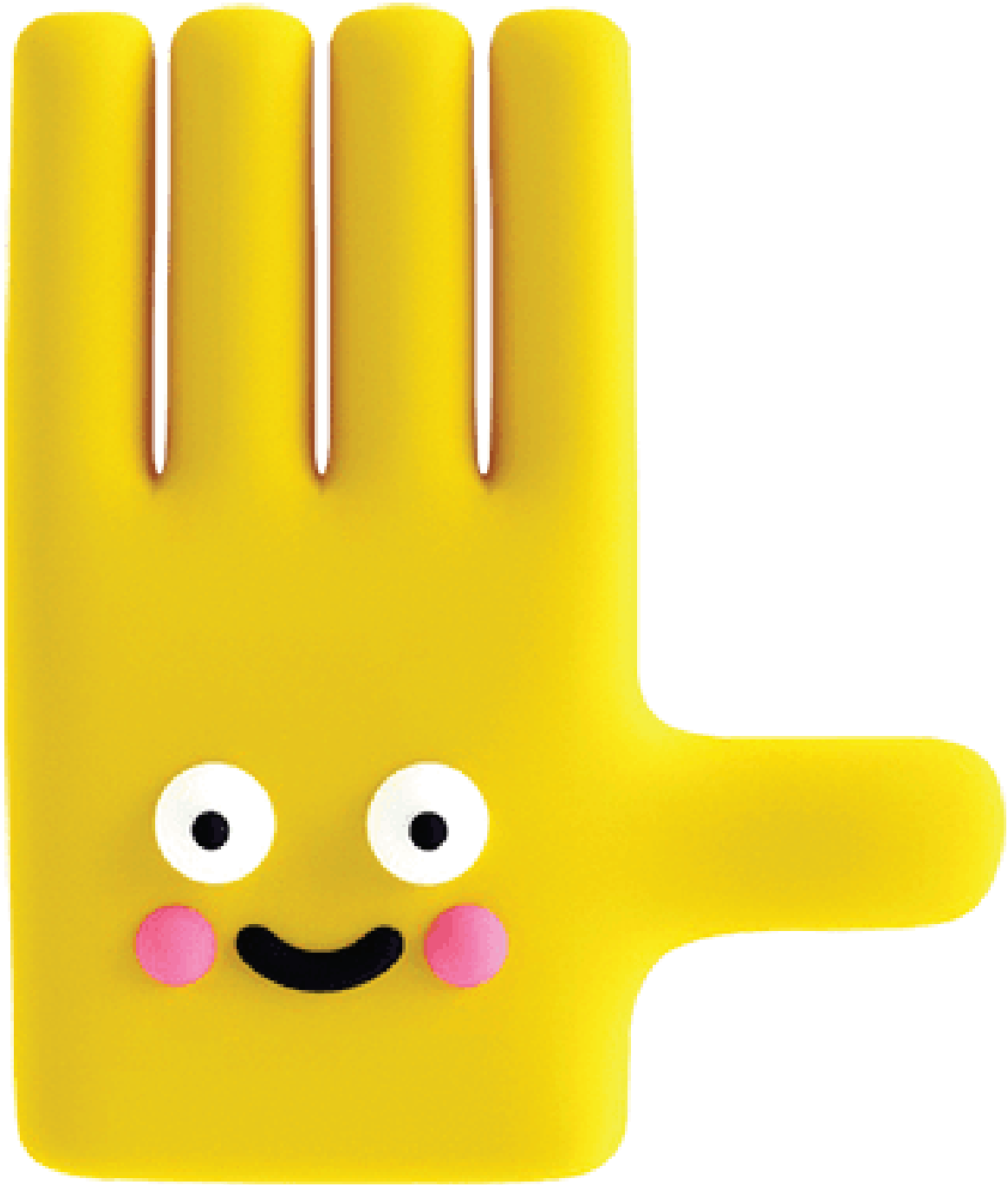
Octopus study: results

	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
Cumulative Recall vs Gold standard included	**0.83**	**0.79**	**0.55**	**0.79**	**0.79**	**0.99**	**0.94**	**0.94**
Cumulative Precision vs Gold standard included	**0.40**	**0.40**	**0.40**	**0.55**	**0.57**	**0.50**	**0.52**	**0.68**

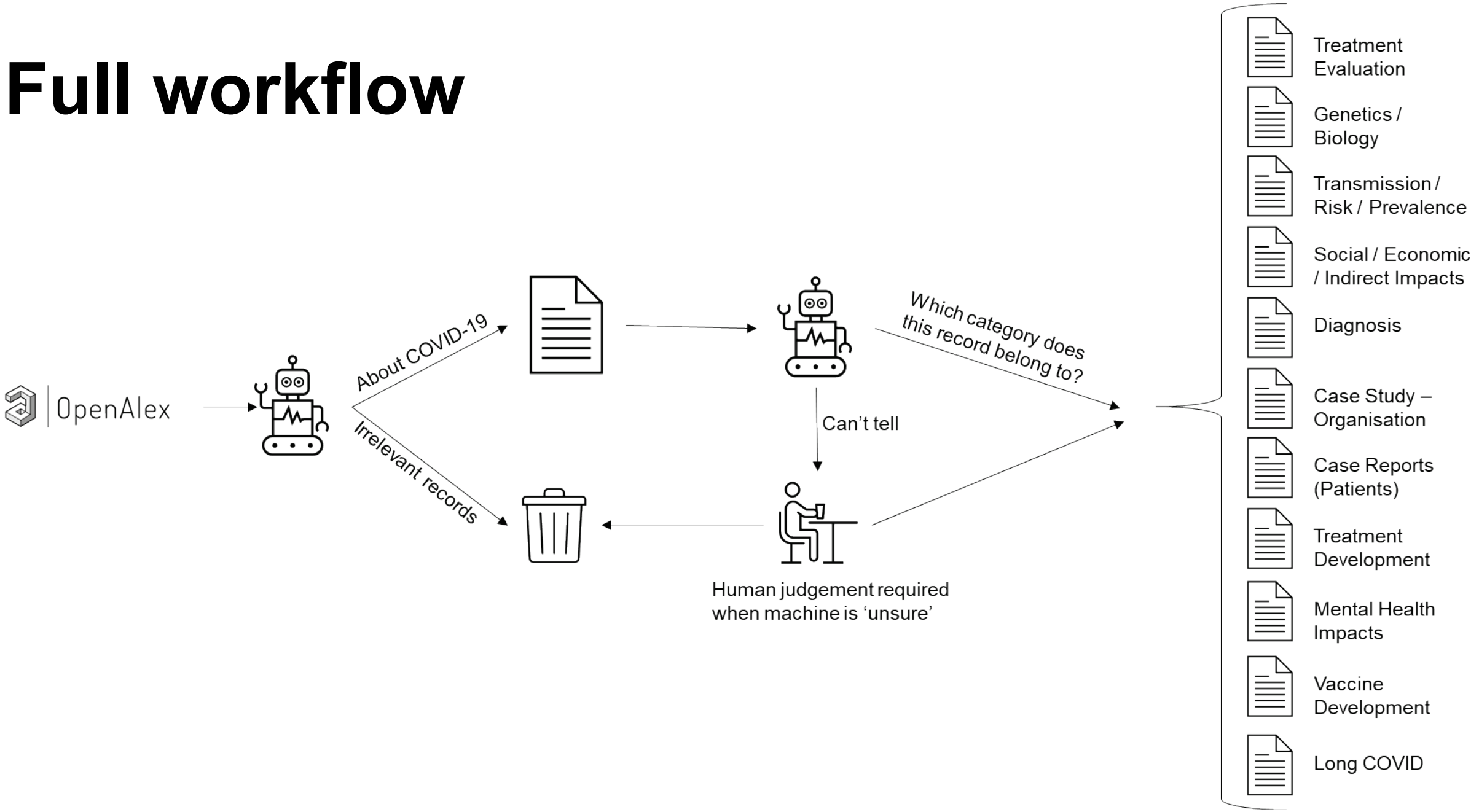
For full results, please see https://wellcomeopenresearch.org/articles/6-210/v1.

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

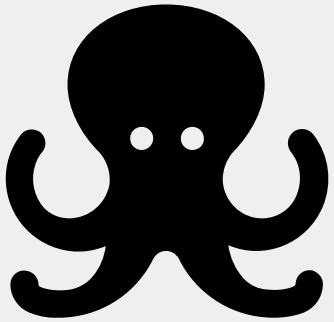


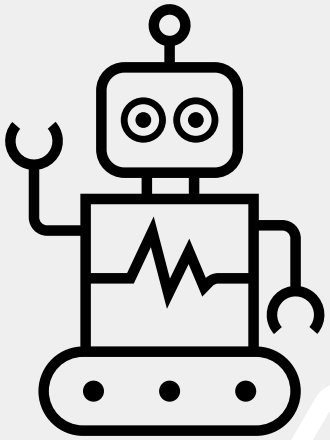
Full workflow



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

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
Web eppi.ioe.ac.uk/

Thank You!

Evaluation:

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

KTDRR Webcasts:

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

Center on
KNOWLEDGE TRANSLATION FOR
DISABILITY & REHABILITATION RESEARCH
at American Institutes for Research

www.ktdrr.org

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

The contents of this presentation were developed under grant number 90DPKT0001 from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS).
The contents of this presentation do not necessarily represent the policy of NIDILRR, ACL, HHS, and you should not assume endorsement by the Federal Government.