



AMU and AMR, searching for a signal in noise

24TH May 2023



Data challenges and opportunities

Unstructured data

80-90% of data exists in unstructured formats

Data use

Less than **1%** of the world's data is analysed

AMR literature

Google Scholar search for AMR yielded **1.3 million** (May 2023)

Systematic review

Most reviews take between **6 to 18** months on average to complete



Searching for a signal in noise

- Evidence synthesis
- Systematic map of Ethiopian livestock disease use case
- LitXpress a Machine Learning tool to accelerate evidence synthesis
- LitXpress application to engineering and law

Louise Donnison, Isla MacVicar, Carys Redman-White (SEBI-L and GAAFS)

SEBI-L mobilises data and generates insights to inform livestock development investment decisions

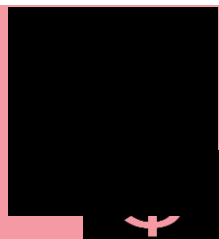
Andrew Horne, Mattia Opper and Colin Gormley (Edina)

Edina transforming innovative ideas into scalable digital solutions using emerging technologies

Application to the field of AMR and AMU your feedback

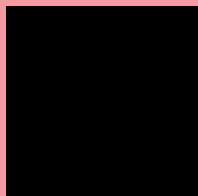


Evidence synthesis and maps



Combines information from multiple studies investigating the same thing providing a comprehensive view of the totality of the evidence and not just a sample of studies

Documented methodologies ensure all actions could be replicated and verified, reduces bias and subjectivity to provide more reliable findings



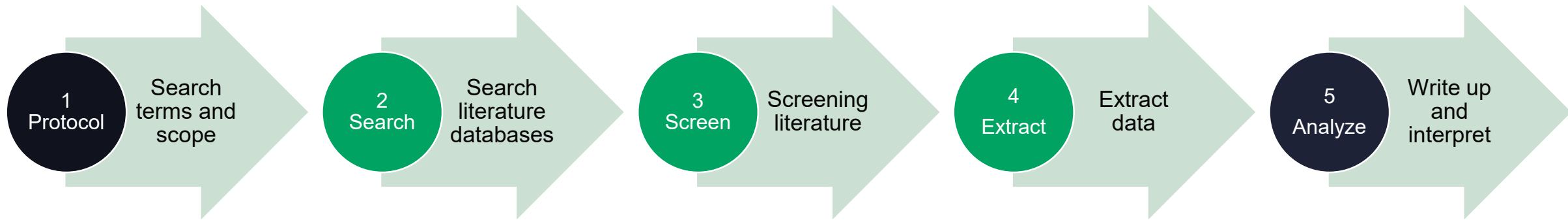
A good review meets the needs of decision makers, is current and timely, readily and widely available



Systematic maps show the big picture, identify gaps in research, help reduce waste and guide future research, hypothesis generators



Evidence Synthesis pathway



Define research question, search strategy including search strings and inclusion criteria

Conduct the searches and collate references using Endnotes

Screen literature for relevance to research question, Kappa test

Unstructured data from literature input into template

Analyse results, describe gaps and overall findings

6-18 months time consuming



Evidence synthesis case study



Systematic mapping of ruminant infectious disease frequency

Systematic map research question

What is the most recent available evidence on ruminant infectious disease frequency and disease-associated mortality in Ethiopia?

Method described in published protocol to reduce bias

Tsouloufi, T., Donnison, L., Smyth, K., & Peters, A. (2020). Development of a systematic mapping review protocol for the most recent evidence on ruminant infectious disease frequency and disease-associated mortality: Ethiopia as a case study. *Animal Health Research Reviews*, 21(1), 96-102.

A maps describe the evidence

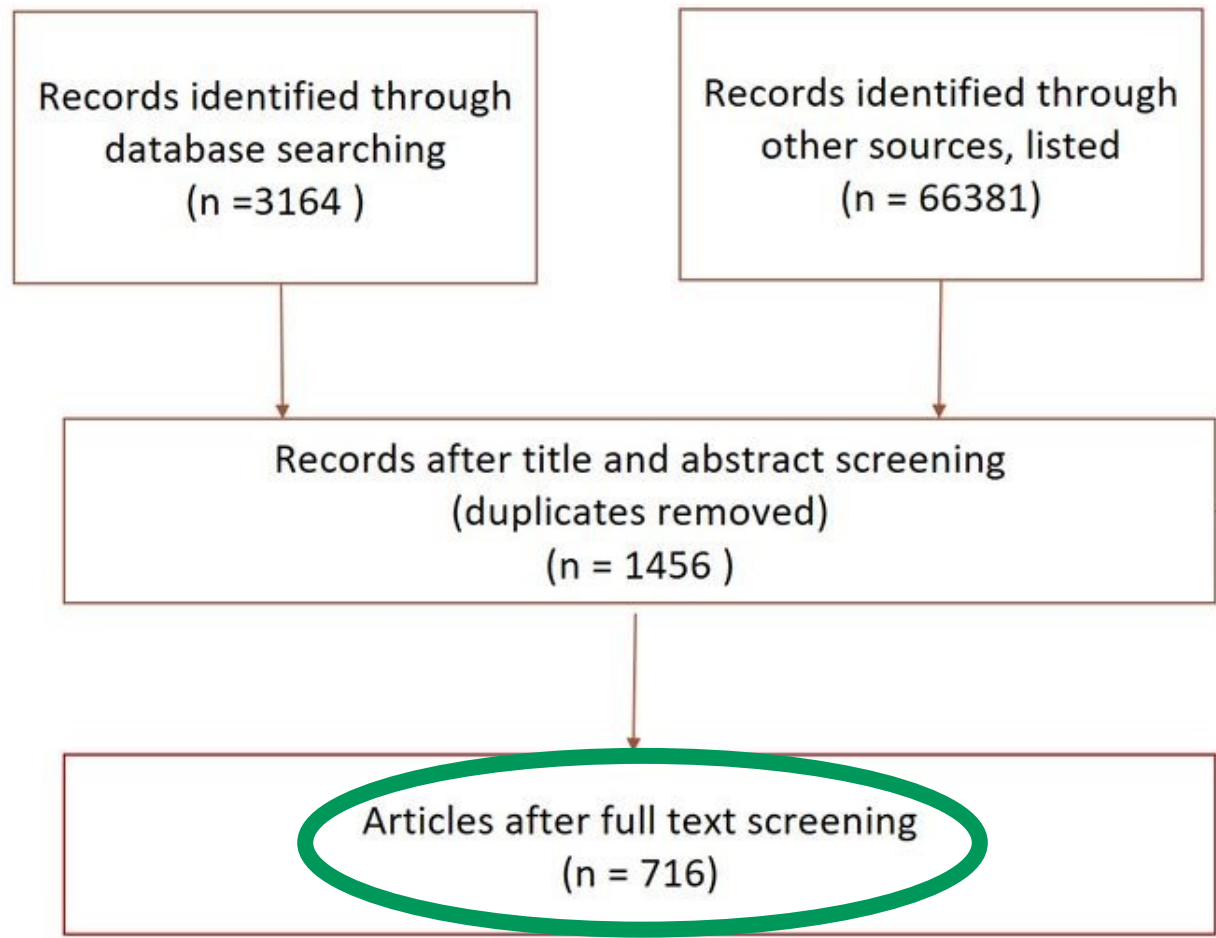
How much evidence is there? Where is the evidence? How have the studies been undertaken?



Searching and Screening literature

Searching

Screening



Bibliographic Databases Organisational Databases

PubMed
 Scopus®
 Google scholar
 FAO
 FIAT PANIS
 CGIAR
 ADDIS ABABA UNIVERSITY
 CABI
 USAID
 FROM THE AMERICAN PEOPLE
 ILRI
 INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE
 OIE
 World Organisation for Animal Health

Ethiopian ruminant disease landscape



Extraction

- Location
- Study population
- Publications
- Diseases
- Testing

Diagnostic test

Microscopy

PM examination

Parasitological examination

CFT

RBT

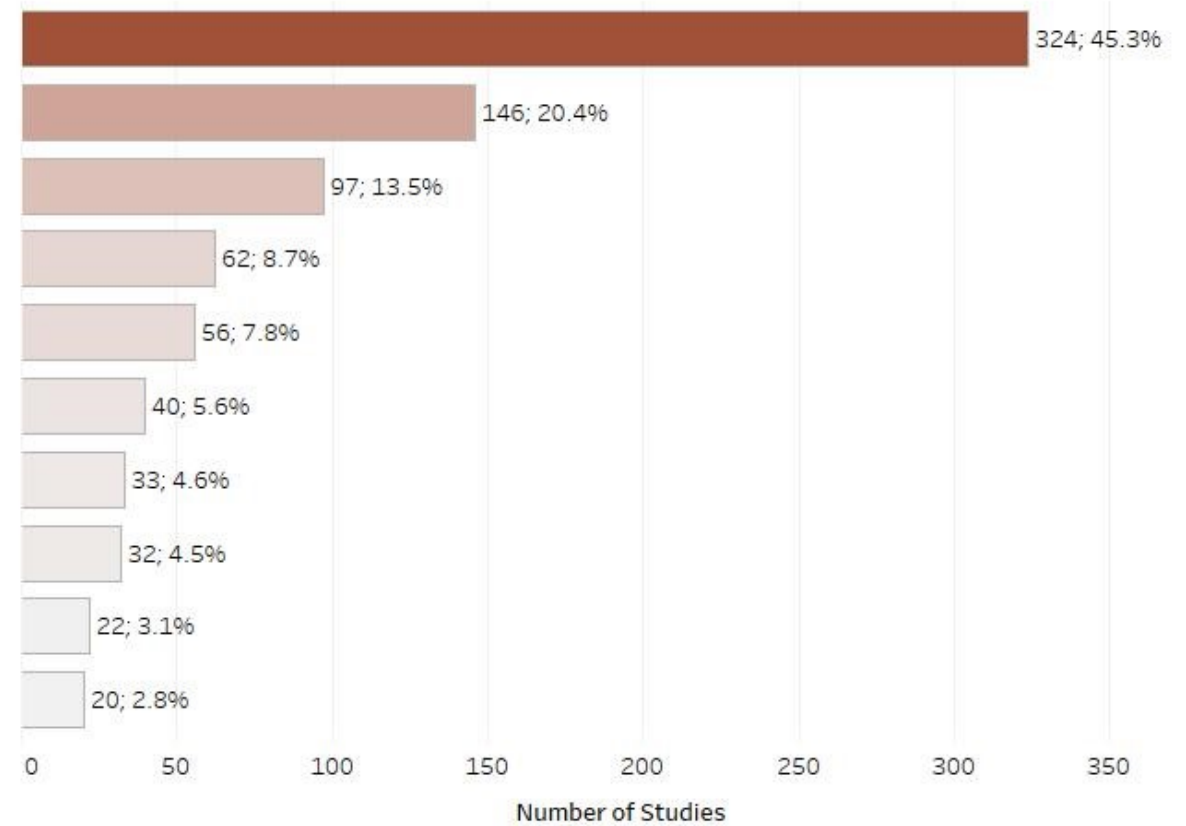
CIDT

Stereomicroscopy

c-ELISA

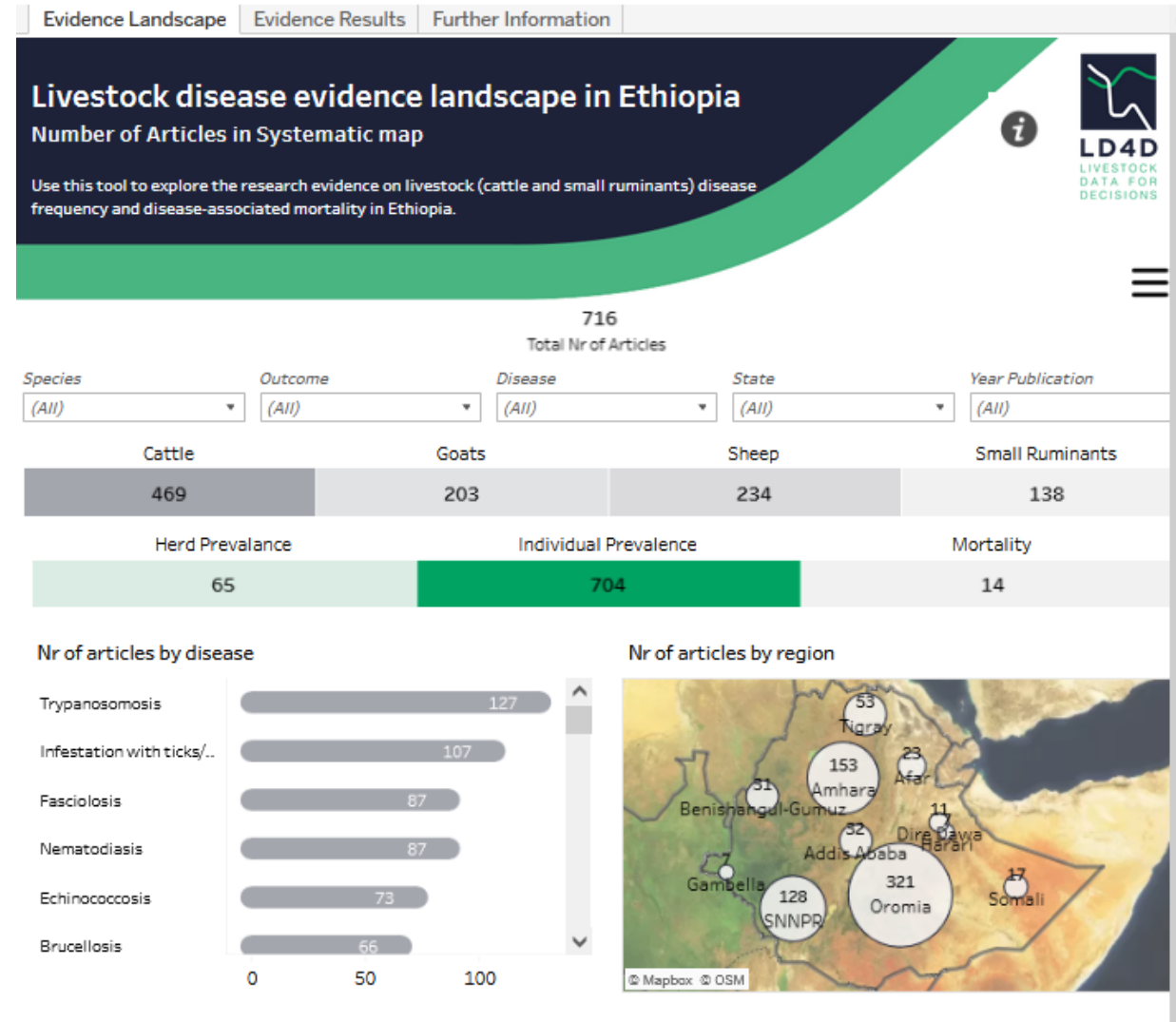
ELISA (3ABC)

Physical examination





SEBI-L using EAST to mobile evidence products



Livestock disease evidence results in Ethiopia

Prevalence, and mortality ranges in Systematic map

Use this tool to explore the research evidence on livestock (cattle and small ruminants) disease frequency and disease-associated mortality in Ethiopia.



LD4D
LIVESTOCK
DATA FOR
DECISIONS

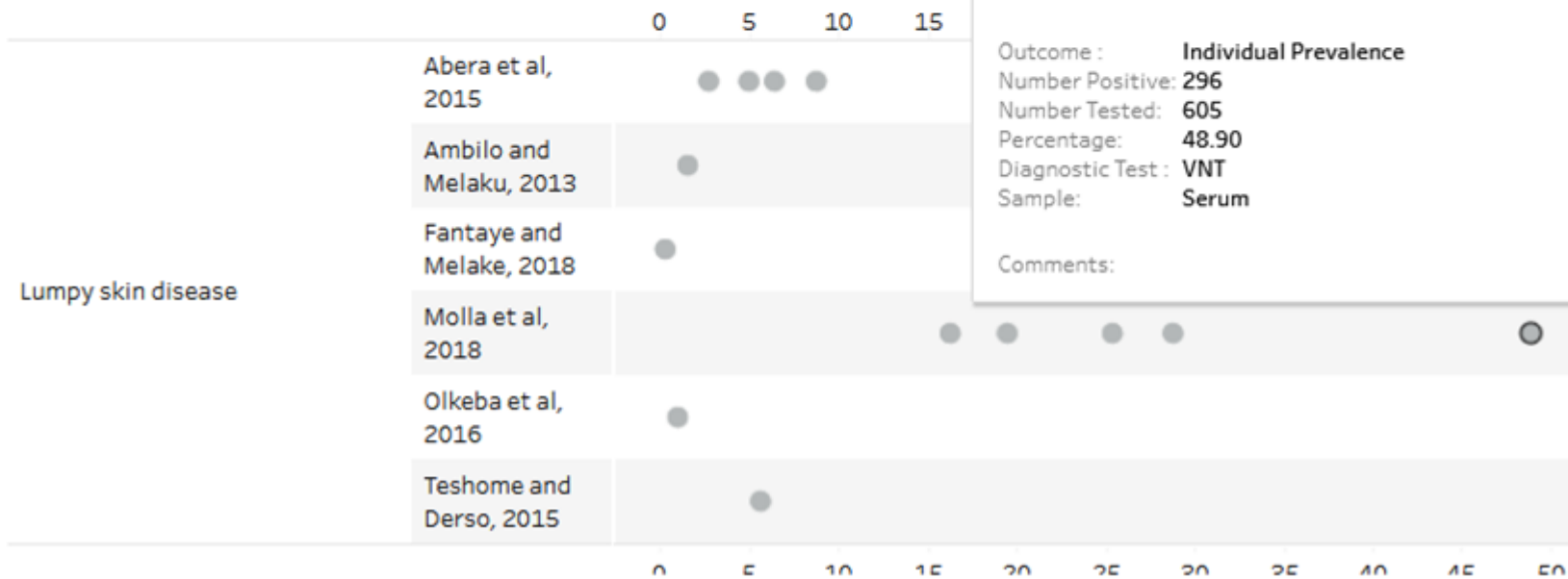


Study Results

Each line represents a study and each dot a data point (see colour legend below), use filters below

■ Individual Prevalence

Outcome: Individual Prevalence
Disease: Lumpy skin disease
Species: Cattle
State: (All)



Molla et al, 2018

Disease : Lumpy skin disease

State : Oromia
Ecosystem : Highland, Lowland, Midland
Production System : Mixed
Species : Cattle
Age Category :
Age Detail :
Breed : Local, Holstein Friesian Zebu cross

Outcome : Individual Prevalence
Number Positive: 296
Number Tested: 605
Percentage: 48.90
Diagnostic Test : VNT
Sample: Serum

Comments:

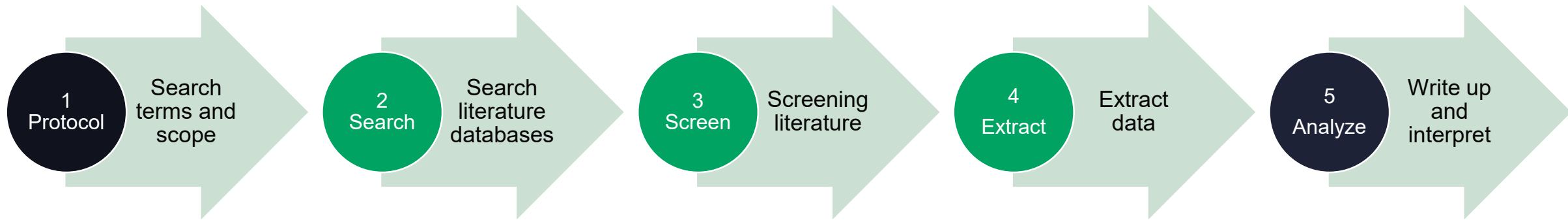
Hypothesis Generator



litXpress a tool for searching for a signal in noise



Evidence Synthesis pathway LitXpress



Define research question, search strategy including search strings and inclusion criteria

Training data

Input queries and run the task

Review screened results

Review dataset and clean results

Analyse results, gluts and gaps and overall findings **living map**

Produces a living map – timely evidence



Future litXpress functionality

1. Years   Weeks
2. Variety of projects
3. Project Management and Interface





Future litXpress project management interface

SYSTEMAP TOOL Account Log In

Project Portal About Collaborators FAQ Contact Us

VTA EXTRACTION

Project

Bacteremia seroprevalence in cattle near a wildlife reserve in Kenya. Enstirobo et al. 2017

Screen

LOCATION

STUDY PERIOD

AGRO-ECOLOGY

SPECIES

DISEASE

AGE

SAMPLE TYPE

DIAGNOSTIC TEST

SAMPLE SIZE

HERD PREVALENCE

INDIVIDUAL PREVALENCE

BREED

PUBLICATION TYPE

COMMENTS

[SAVE & EXIT](#)

[SAVE & FINISH](#)

SYSTEMAP TOOL Account Log In

Project Portal About Collaborators FAQ Contact Us

SCREEN DOCUMENTS

Queries

Label

Classifier Model

Search Engine

Maximum number of Results

Start Year

End Year

Email Results

Just Me

Confience Threshold

[Submit Task](#)

PubMed_Search_2.SCI

GoogleScholar_Search_2.SCI

Web of Science_Search_2.SCI

[Combine Results](#)

SYSTEMAP TOOL Account Log In

Project Portal About Collaborators FAQ Contact Us

MACHINE REVIEW - OF

EXCLUDED DOCUMENTS

EXCLUDED	SCREENED	ISSUES	ABSTRACT	SCREEN SOURCE
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

Low-Cost Extract

TITLE	AUTHORS	DOI	ABSTRACT	SCREEN SOURCE
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

SYSTEMAP TOOL Account Log In

Project Portal About Collaborators FAQ Contact Us

CONSISTENCY CHECK

Project

Screen

Observe All

Review

Kappa

EXTRACTOR

Cohen's kappa coefficient (k) =

A consistency check of the results is carried out and a conclusion should then draw based on the results and discussions between reviewers. Boxes are colour coded based on the conclusion that reviewers have drawn from the title and abstract of a specific paper

Sorted in order of best agreement

	R1	R2	R3	C
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

Kappa (1) = 0.64

LitXpress
expansion to other subject areas
Making an impact

Engineering

- As a test case the tool was adapted to a small part of a literature review for the Microfluidics department in Engineering
- Microfluidics confirmed the value of the tool

Engineering : Microfluidics

- Key issue is the recognition of Equations within a body of text
- Texts are typically PDFs, and text conversion from PDF breaks on encountering an equation
- Recognising equations is a Vision AI problem, rather than an NLP AI problem
- Engineering have 200 PhD and 100 Postdoc researchers who are expected to run ~2 reviews every three years, each review taking ~150 hours. Automation could reduce this to ~15 hours, a total time saving of 40 person-years in researcher time per year for Engineering alone.

Microfluidics

$n_i(\mathbf{r}, t_+)$ and $n_{i'}(\mathbf{r} + \mathbf{c}_i, t_+)$, corresponding to velocities \mathbf{c}_i and $\mathbf{c}_{i'} (\mathbf{c}_{i'} = -\mathbf{c}_i)$ parallel to the link connecting \mathbf{r} and $\mathbf{r} + \mathbf{c}_i$, the notation $n_i(\mathbf{r}, t_+) = n_i(\mathbf{r}, t) + \Delta_i(\mathbf{r}, t)$ is used to indicate the post-collision distribution, (2.7). The velocity of the boundary node \mathbf{u}_b is determined by the solid-particle velocity \mathbf{U} , angular velocity $\boldsymbol{\Omega}$, and centre of mass \mathbf{R} ,

$$\mathbf{u}_b = \mathbf{U} + \boldsymbol{\Omega} \times (\mathbf{r} + \frac{1}{2}\mathbf{c}_i - \mathbf{R}). \quad (2.8)$$

By exchanging population density between n_i and $n_{i'}$, the local momentum density can be modified to match the velocity of the solid-particle surface at the boundary node, without affecting either the mass density or the stress, which depend only on the sum $n_i + n_{i'}$. The precise form for the boundary-node collision operator is

$$\left. \begin{aligned} n_i(\mathbf{r} + \mathbf{c}_i, t+1) &= n_{i'}(\mathbf{r} + \mathbf{c}_i, t_+) + 2a_1^{c_i} \rho \mathbf{u}_b \cdot \mathbf{c}_i, \\ n_{i'}(\mathbf{r}, t+1) &= n_i(\mathbf{r}, t_+) - 2a_1^{c_i} \rho \mathbf{u}_b \cdot \mathbf{c}_i \end{aligned} \right\} \quad (2.9)$$

As a result of the boundary-node interactions (2.9), forces are exerted on the solid particles,

$$\mathbf{f}(\mathbf{r} + \frac{1}{2}\mathbf{c}_i, t + \frac{1}{2}) = 2[n_i(\mathbf{r}, t_+) - n_{i'}(\mathbf{r} + \mathbf{c}_i, t_+) - 2a_1^{c_i} \rho \mathbf{u}_b \cdot \mathbf{c}_i] \mathbf{c}_i; \quad (2.10)$$

thus momentum is exchanged locally between the fluid and the solid particle, but the combined momentum of solid and fluid is conserved. The forces and torques on the solid particle, obtained by summing $\mathbf{f}(\mathbf{r} + \frac{1}{2}\mathbf{c}_i)$ and $(\mathbf{r} + \frac{1}{2}\mathbf{c}_i) \times \mathbf{f}(\mathbf{r} + \frac{1}{2}\mathbf{c}_i)$ over all the boundary nodes associated with a particular particle, are then used to update the particle velocity and angular velocity, according to the laws of Newtonian mechanics. The mass and moment of inertia of the particle are preassigned, depending on the

Refugee Law

- Canada is one of the few countries to publish its legal decisions on admitting refugees in full
- This is the subject of Claire Barale's PhD : **Enabling ethical human-AI reasoning in international law**
- Rather than develop a new NLP system Claire is adapting our tool to process the Canadian database
- The aim will be to identify automatically anomalous decisions that should be appealed



RPD File No. / N° de dossier de la SPR : MB2-06059

Private Proceeding / Huis clos

Reasons and decision – Motifs et décision

Claimant(s) Demandeur(e)(s) d'asile

XXXX XXXX XXXX

Date(s) of hearing

Date(s) de l'audience

October 28, 2013

Place of hearing

Lieu de l'audience

Montréal, Quebec

Date of decision
and reasons

Date de la décision
et des motifs

December 16, 2013

Panel

Tribunal

Stéphane Morin

Counsel for the claimant(s)

Conseil(s) du (de la/des)
demandeur(e)(s) d'asile

M^e Marie-José Blain

Designated representative

Représentant(e) désigné(e)

N/A

Counsel for the Minister

Conseil du (de la) ministre

N/A

2013 CanLII 92087 (CA IRB)

Dataset

First page

Keywords

REFUGEE PROTECTION DIVISION / EMPLOYMENT / EXCLUSION / ART. 1F(A) / CRIMES
AGAINST HUMANITY / COMPLICITY / PROFILE / FAILURE TO DISASSOCIATE / MALE /
NEGATIVE / PARTICULAR SOCIAL GROUP / FAMILY / REFUGEE SUR PLACE / WELL-FOUNDED
FEAR OF PERSECUTION / FEMALE / CHILDREN / POSITIVE / ETHIOPIA

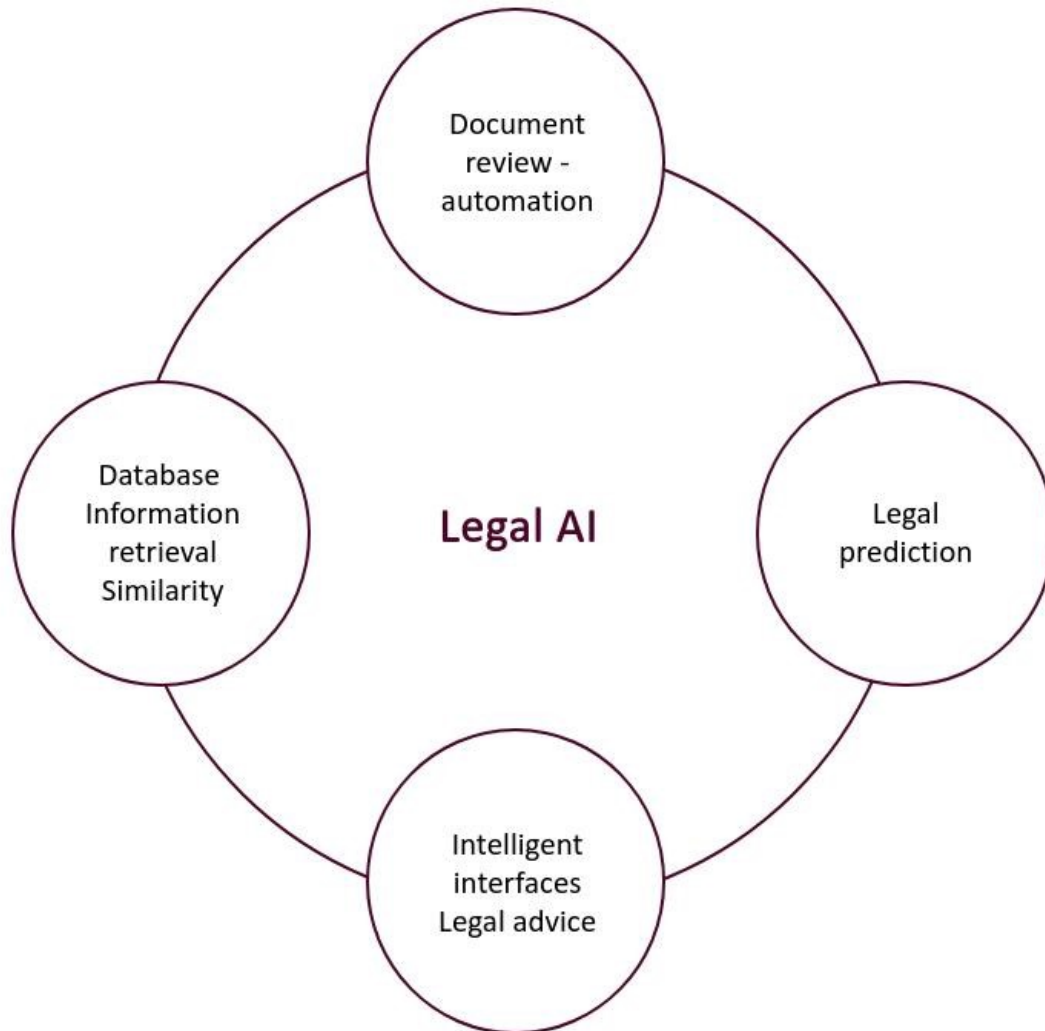
Features

>>> Other text data:

- + country reports
- + international conventions
- + local regulation, guidelines and laws

Background

Legal AI – LegalTech



UNHCR, Refworld, <https://www.refworld.org/>

Chen, D. L., & Egel, J. (2017, June). **Can machine learning help predict the outcome of asylum adjudications?**. In *Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law* (pp. 237-240).

Aletras, N., Tsarapatsanis, D., Preoțiu-Pietro, D., & Lampos, V. (2016). **Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective**. *PeerJ Computer Science*, 2, e93.

Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). **A general approach for predicting the behavior of the Supreme Court of the United States**. *PloS one*, 12(4), e0174698.

Undavia, S., Meyers, A., & Ortega, J. E. (2018, September). **A comparative study of classifying legal documents with neural networks**. In *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 515-522). IEEE.

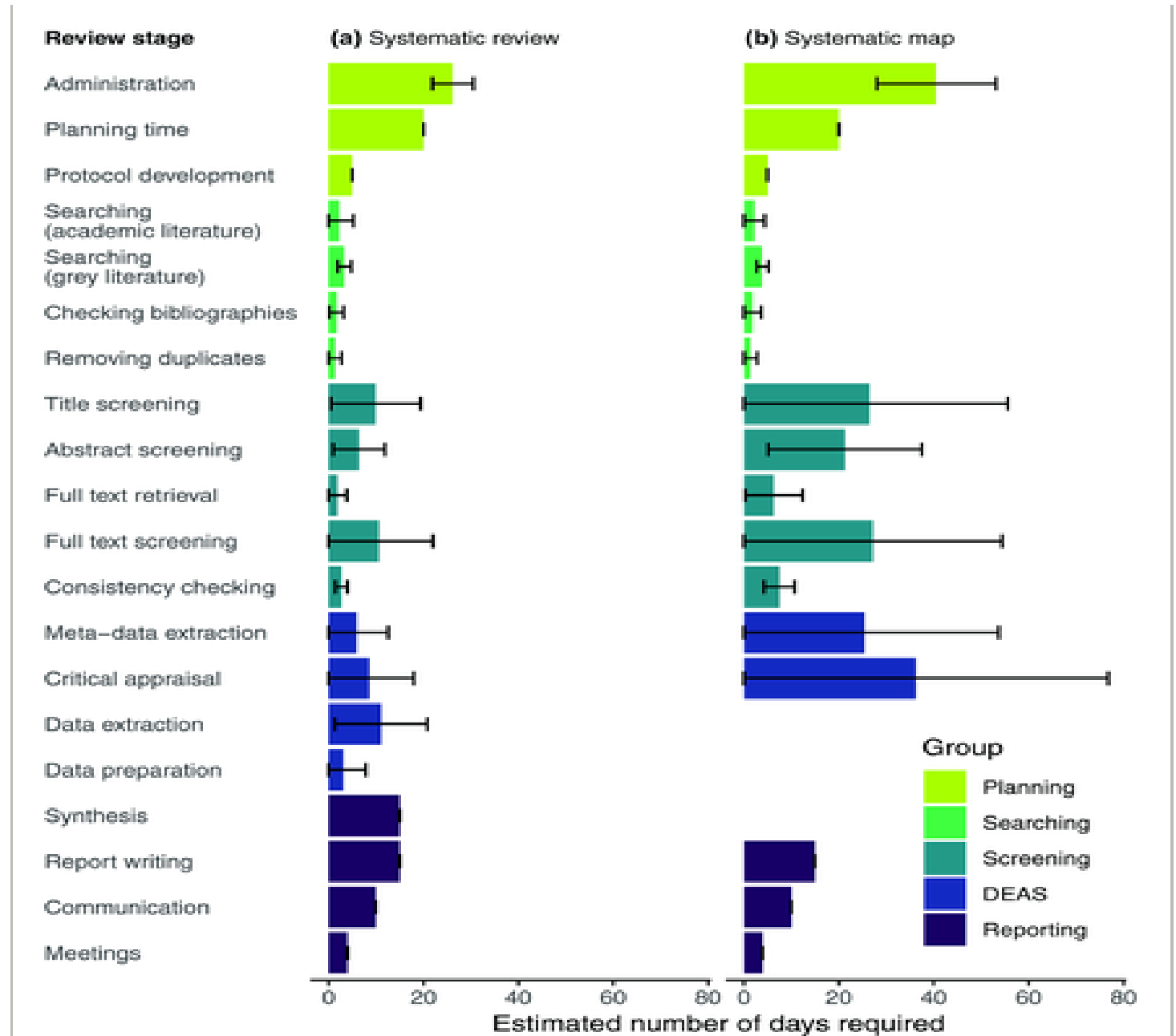
Medvedeva, M., Vols, M., & Wieling, M. (2020). **Using machine learning to predict decisions of the European Court of Human Rights**. *Artificial Intelligence and Law*, 28(2), 237-266.

Next Steps

- Continue to work with Microfluidics before involving Materials, moving eventually to a general solution for Engineering
- Develop a tool for Refugee Law document processing, then move to automating the search for relevant case law, eventually developing a general legal research tool
- Look into offering the potential value of equation processing to schools of Physics and Mathematics
- Investigate other interested subjects, and look at larger scale collaborations with Cochrane and JISC

Impact

- Example of a paper published in Conservation Biology
- Up to 50% of time in these cases would be saved by automation
- In the order of ~100 days per review



Impact

- Depending on the review, 50%-90% of the work involved can be automated by a machine learning approach
- There is an initial investment of time, often requiring a manual review to be shadowed and learnt from
- Impact for any given department performing reviews is likely to be in the order of hundreds of days
- Machine accuracy for a well trained classifier is in the order of 5% better than human, so quality of results, and therefore research, is improved

Impact

- There are many disparate efforts within the Schools to apply ML to Systematic Reviews, however they are poorly funded and require individuals familiar with ML to run them. There is significant duplication of effort
- Review time that is automated away leads directly to researchers being able to spend more time on actual research – substantially increasing research output for the University of Edinburgh
- Systematic Review is a necessary step before conducting research, and is viewed as a barrier to initiating research. Reduction in the size of the barrier means more, better research will be initiated



AMU and AMR, **litXpress a tool** for
searching for a signal in noise



AMR and litXpress

Run query on PubMed 'Poultry AND antibiotic* and *Intensive': 3 errors, 74 low confidence, 26 excluded

Example paper: Antimicrobial resistance among *Campylobacter* spp. strains isolated from different poultry production systems at slaughterhouse level

Labels: Species, Disease, Study date, Sample

Abstract

The aim of the current work was to evaluate the prevalence and antimicrobial susceptibility of **Camp [Entity: DISEASE, Score: 1.000]** **##ylobacter [Entity: DISEASE, Score: 0.915]** **spp [Entity: DISEASE, Score: 0.999]** **##.** isolated from different **chicken [Entity: SPECIES, Score: 0.999]** production systems at the slaughterhouse level. **Chicken [Entity: SAMPLE TYPE, Score: 0.981]** sampling at slaughterhouse was performed for cecum, carcass, and breast meat from flocks of organic (n = 6), extensive indoor (n = 14), and intensive production (n = 14), totaling 34 ceca pools, 64 neck skin pools, and 132 breasts, representing 96 **##,3 [Entity: SAMPLE SIZE, Score: 0.549]** **##86 [Entity: DISEASE, Score: 0.994]** **chickens [Entity: SPECIES, Score: 0.994]** **##.** A collection of **167 [Entity: SAMPLE SIZE, Score: 0.990]** strains were identified as **Campylobacter [Entity: DISEASE, Score: 0.910]** **coli [Entity: DISEASE, Score: 0.996]** (n = 85) and **Camp [Entity: DISEASE, Score: 1.000]** **##ylobacter [Entity: DISEASE, Score: 0.776]** **jejuni [Entity: DISEASE, Score: 0.999]** (n = 82) and were tested for susceptibility to 11 antimicrobial agents by the **disk diffusion [Entity: DIAGNOSTIC TEST, Score: 0.996]** **method [Entity: DIAGNOSTIC TEST, Score: 0.998]** **##.** The frequency of **Campyl [Entity: DISEASE, Score: 0.988]** **##obacter [Entity: DISEASE, Score: 0.899]** in **chicken [Entity: SPECIES, Score: 1.000]** samples from different production systems was between 79 and 100%. **Campylobacter [Entity: DISEASE, Score: 0.858]** isolated from all origins were resistant to the fluoroquinolones studied (80-98%). However, for ciprofloxacin and ofloxacin, the **Campyl [Entity: DISEASE, Score: 0.987]** **##obacter [Entity: DISEASE, Score: 0.890]** isolates from **extensive [Entity: PRODUCTION SYSTEM, Score: 0.998]** indoor **chicken [Entity: SPECIES, Score: 0.998]** were significantly (P < 0.05) less resistant (77 and 58%) than that from organic (97 and 91%) and intensive production (96 and 95%). A high probability of tetracycline resistance occurrence was also found for the **Camp [Entity: DISEASE, Score: 1.000]** **##ylobacter [Entity: DISEASE, Score: 0.852]** **spp [Entity: DISEASE, Score: 0.999]** **##.** tested **(5 [Entity: SAMPLE SIZE, Score: 0.949]** **##8%** for **C. [Entity: DISEASE, Score: 0.987]** **jejuni [Entity: DISEASE, Score: 0.996]** and **76 [Entity: SAMPLE SIZE, Score: 0.989]** **##%** for **C. [Entity: DISEASE, Score: 0.996]** **coli [Entity: DISEASE, Score: 1.000]** **##).** A more frequent profile of multidrug resistance was noticed for isolates from intensive and organic production than for extensive indoor production. These results reinforce the need of efficient strategy implementation to control and reduce **Campyl [Entity: DISEASE, Score: 0.892]** **##obacter [Entity: DISEASE, Score: 0.986]** in **chickens [Entity: SPECIES, Score: 1.000]** at production and slaughter levels, and the necessity to reduce the use of antimicrobials in **poultry [Entity: SPECIES, Score: 0.996]** sector.

Methods

METHODS Sampling and Isolation of **Campyl [Entity: DISEASE, Score: 0.986]** **##obacter** Different **chicken [Entity: SPECIES, Score: 0.999]** flocks from organic (n = 6), extensive indoor (n = 14), and intensive production (n = 14), representing a total of **96,386 [Entity: SAMPLE SIZE, Score: 0.997]** **birds [Entity: SPECIES, Score: 0.956]** **##.** were sampled during the slaughter and deboning process from **December 2007 to October [Entity: STUDY DATE, Score: 0.877]** **2008, [Entity: STUDY DATE, Score: 0.999]** according to ISO 17604:2003 (International Organization for Standardization, 2003) and the guidelines of The National Veterinary Institute (2010). Sampling was performed for cecum, carcass (neck skin), and deboned poultry breast pieces, maintaining each flock's traceability. For each **chicken [Entity: SPECIES, Score: 0.999]** flock, a pool of 10 ceca was **randomly [Entity: STUDY DESIGN, Score: 1.000]** sampled, and the neck skins of 5 carcasses, after the inside-outside shower and before rapid cooling, were collected twice constituting 2 pool samples. A total of 34 ceca pools and 64 neck skin pool samples were collected. Four carcasses per flock were marked and, after deboning (performed according 1579 to industrial practices), their **breast [Entity: SAMPLE TYPE, Score: 0.642]** **##s** were sampled (n = 132) and delivered to the laboratory in a commercial refrigerated vehicle. **Campyl [Entity: DISEASE, Score: 0.573]** **##obacter** isolation was performed according to ISO 10272-1:2006 (International Organization for Standardization, 2006). Bacterial strains were grown and purified on Columbia Blood agar (BioMerieux, Marcy l'Etoile, France), supplemented with 5%



Go to www.menti.com and use the code 1600 1042

 Mentimeter

What would you want from literature data mining? What would be your concerns?

www.menti.com

Code 1600 1042



sebiadmin@ed.ac.uk