Finding the needles in the evidence haystack: smart sorting for conservation decision making

SNAPP

CONSERVATION INTERNATIONAL

DataKind

One of the SNAPP Evidence-Based Conservation working group projects focused on …

How do we find evidence?

How do we communicate evidence?

How do we use evidence?

Global scientific output doubles every nine years…

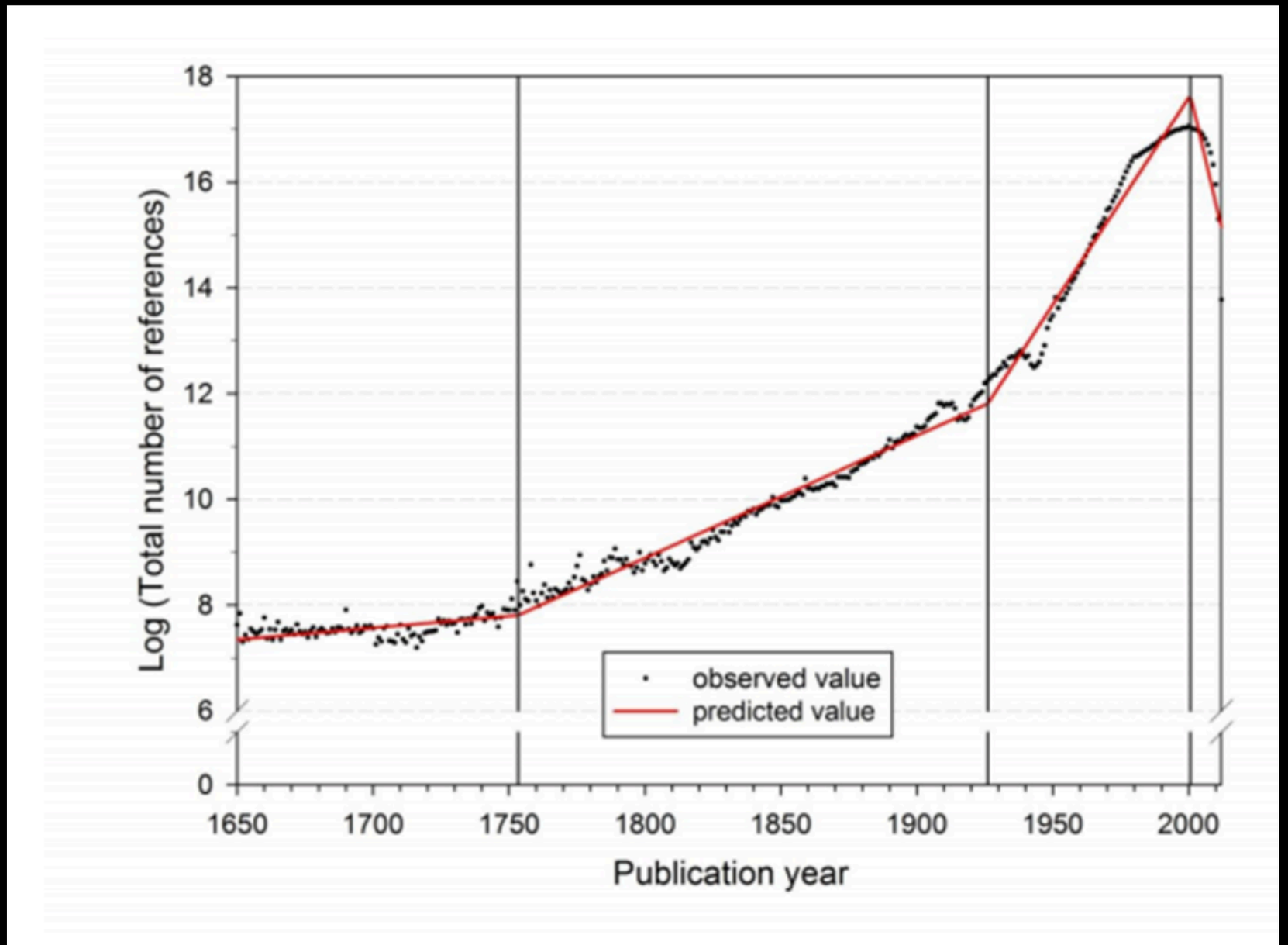


Figure 2. Segmented growth of the annual number of cited references from 1650 to 2012 (citing publications from 1980 to 2012)

Van Noorden, 2014

Wonkblog

The solutions to all our problems may be buried in PDFs that nobody reads

By Christopher Ingraham May 8, 2014

What if someone had already figured out the answers to the world's most pressing policy problems, but those solutions were buried deep in a PDF, somewhere nobody will ever read them?

According to a recent report by the World Bank, that scenario is not so far-fetched. The bank is one of those high-minded organizations -- Washington is full of them -- that release hundreds, maybe thousands, of reports a year on policy issues big and small. Many of these reports are long and highly technical, and just about all of them get released to the world as a PDF report posted to the organization's Web site.

The World Bank recently decided to ask an important question: *Is anyone actually reading these things?* They dug into their Web site traffic data and came to the following conclusions: Nearly one-third of their PDF reports had never been downloaded, not even once. Another 40

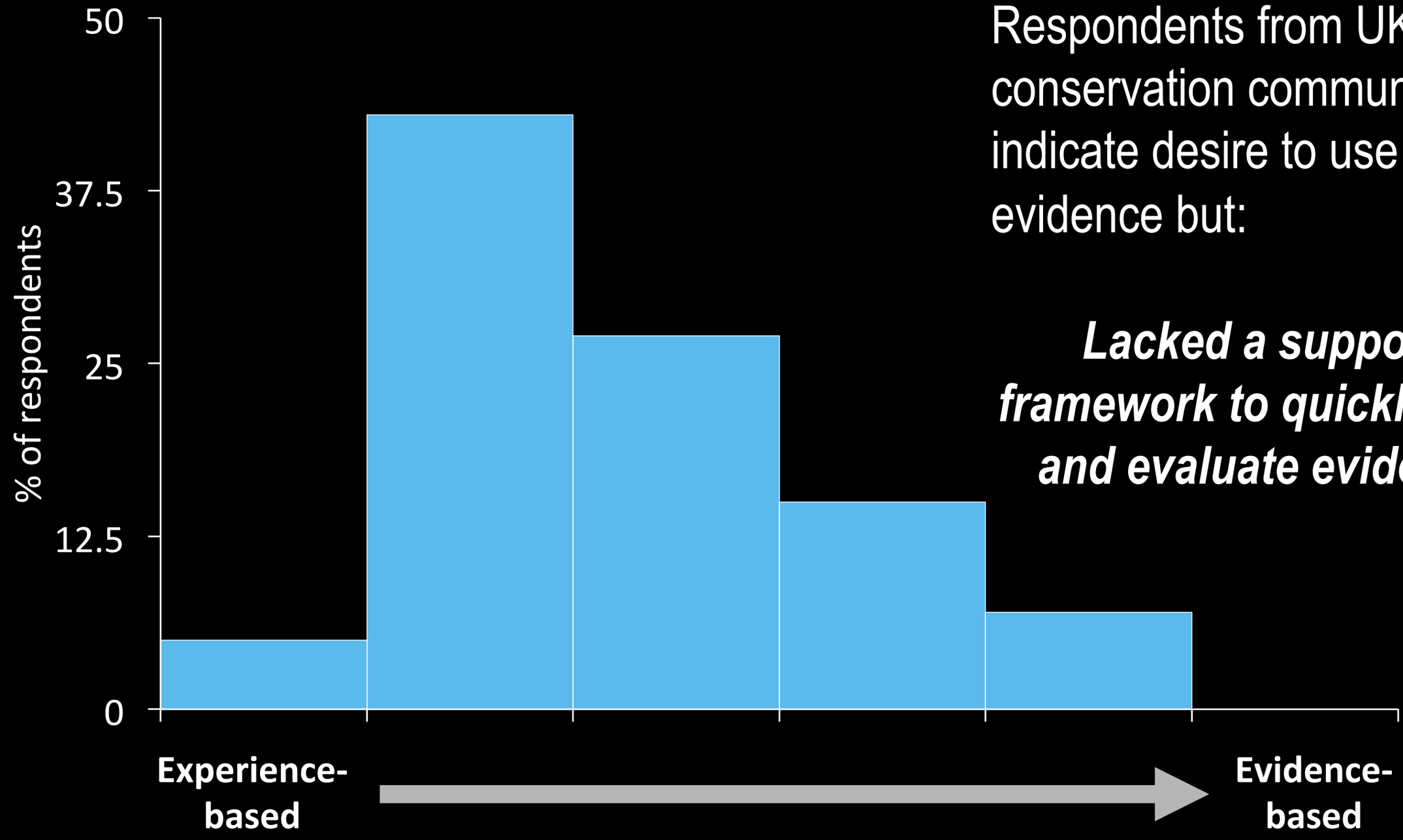
The Washington Post

Academics can Change the World - If They Stop Talking Only to Their Peers

By The Conversation Africa

The Huffington Post

Evidence gap



Respondents from UK conservation community indicate desire to use evidence but:

Lacked a support framework to quickly sort and evaluate evidence

modified from Pullin et al. 2004

The need

Practitioners need **access to research insights** from academic and grey literature for evidence-based decision making

Researchers need a framework to follow to create these resources

Oxford

A solution

Systematic review/map

PICO

Test Library, Boolean Search String

(Peer review, select grey literature)

Database Query: SCOPUS & IEEE

(past two years, language restricted)

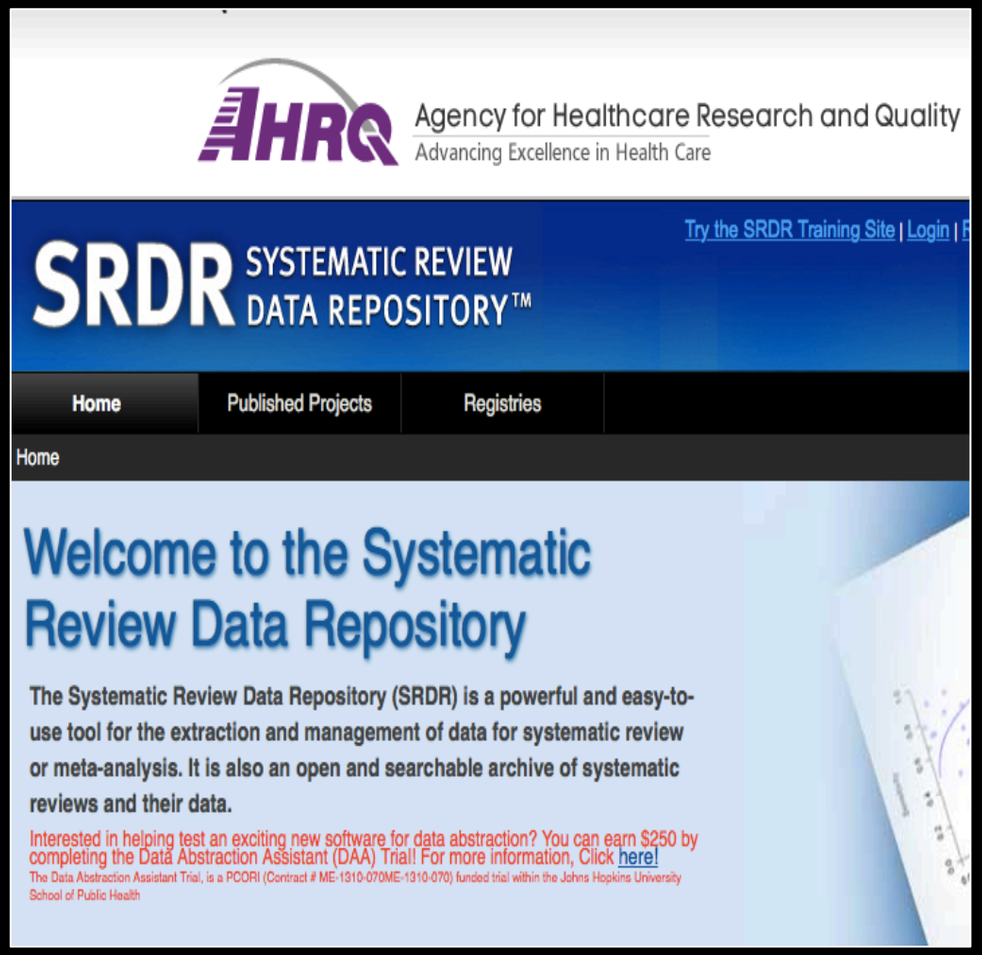
Citation databasing

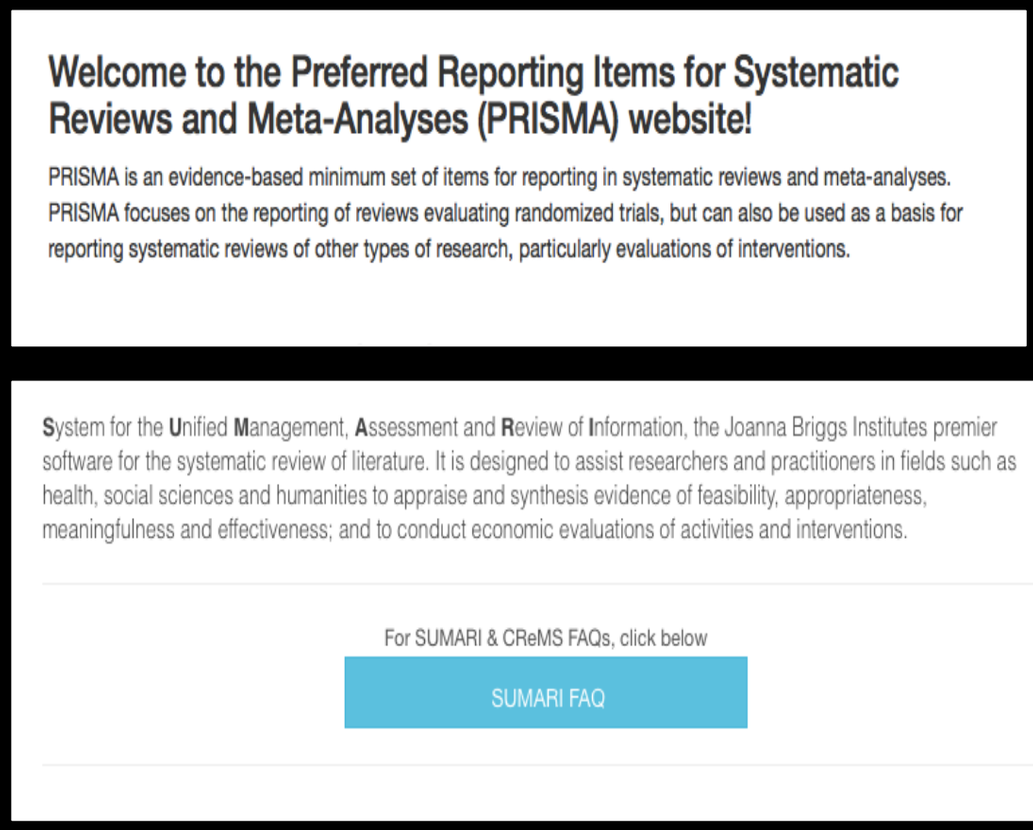
Feature tagging

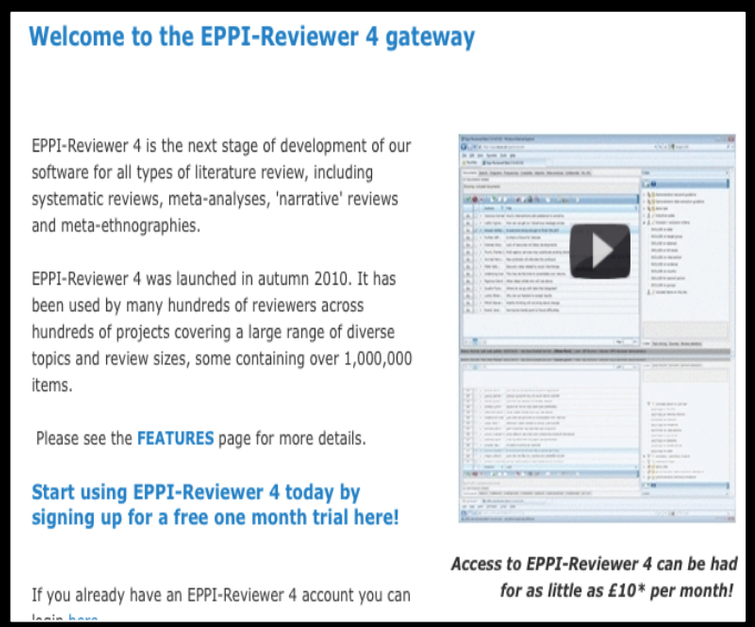
(direct keywording, abstract/paper review)

"[process] aims to provide a complete, exhaustive summary of current literature relevant to a research question."

Problem #1: there are tools out there





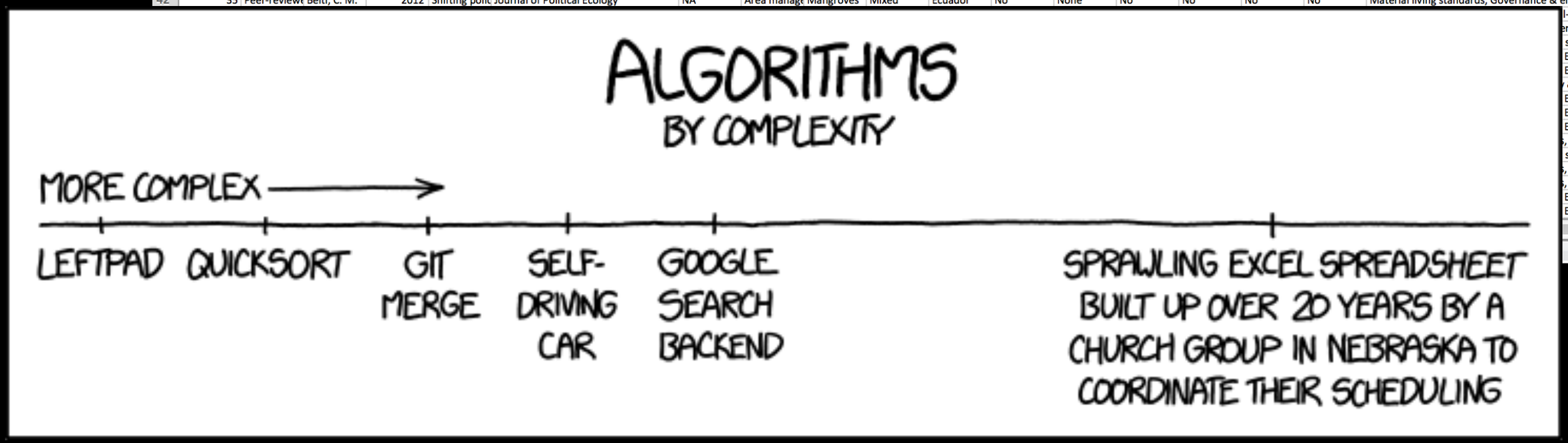


Software name	Setting up the review	Scoping/pilot study	Literature searching	Duplicate checking	Article screening	Data coding	Critical appraisal	Synthesis	Documentation
E.g.	Facilitation of question formulation and/or stakeholder engagement	Protocol development, PICO* elements specified	Software integrated with publication databases	Automated marking of duplicates	For study selection	Tagging and extraction to support meta-analyses	Risk of bias assessments	Facilitates quantitative/ qualitative syntheses of results	Output of text, figures or tables to assist with report writing
CADIMA									
* Colandr									
Covidence									
DistillerSR									
EROS									
* EPPI-Reviewer 4									
HAWC									
METAGEAR package for R									
PARSIFAL									
* Rayyan									
REviewER									
RevMan 5									
RevMan Web	Data unavailable								
SESRA									
* SLR-Tool									
SLuRp									
SRDB.PRO									
SRDR									
StArt									
SUMARI									
* SWIFT-Review									
SyRF									
TOTAL	5	10	13	11	20	19	12	15	13

Kohl et al. 2017

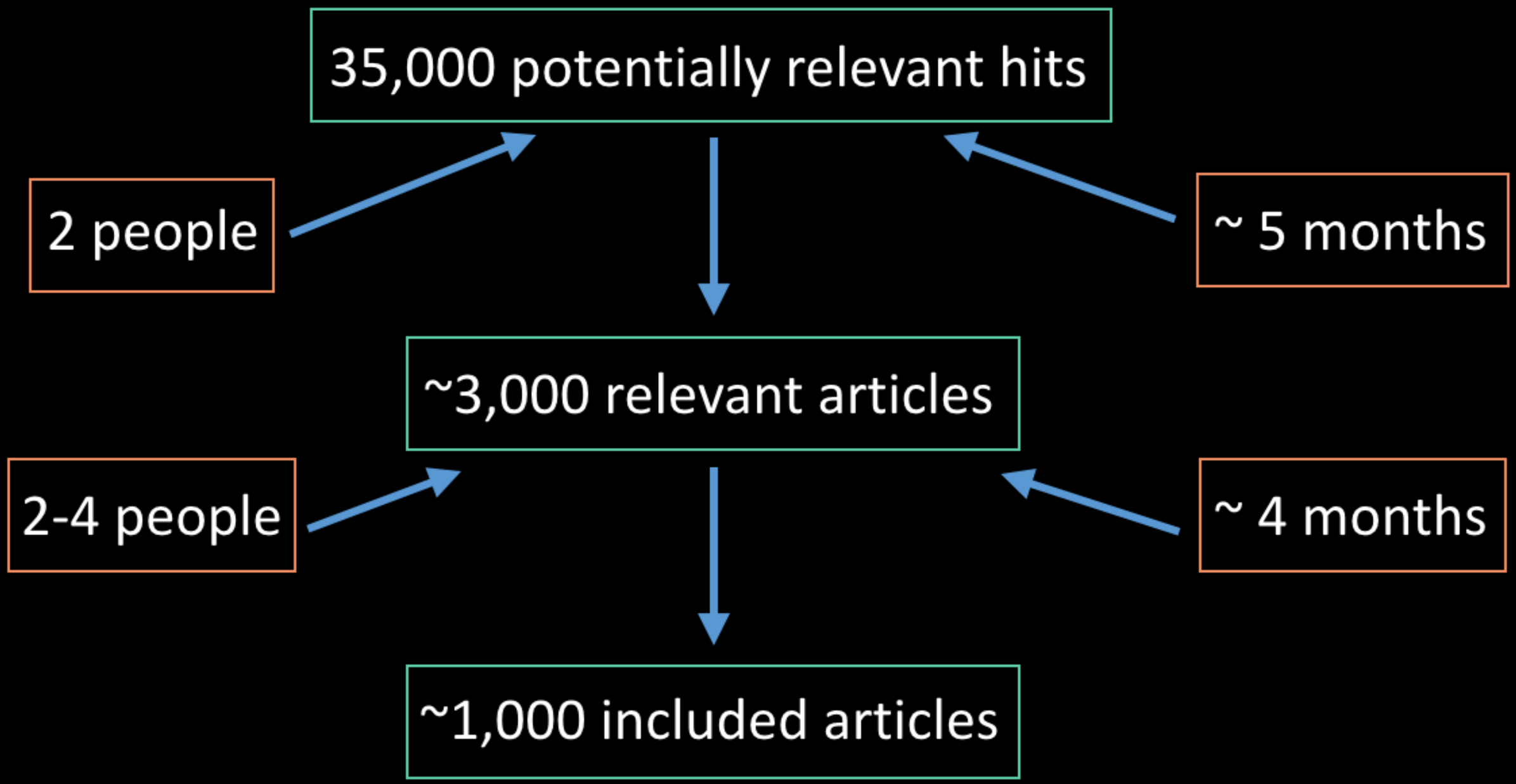
Problem # 2: the common practice is messy

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Article ID	Type of publ	Author(s)	Year of publ	Title	Journal Nam	Volume	Page #	Publisher	Type of inte	Type of bion	Type of data	Country of s	Does the stu	If so, what ty	Does the stu	Does the stu	Does the stu	Does the stu	Type of outcomes measured by study	Theory of cha	If so, wha
2	1	Peer-reviewe	Abbott, J.; Ca	2007	Rivers as res	The Canadia	51(3)	280-302	Canadian As	Species man	Flooded Gra	Qualitative	Namibia	Yes	sites	No	Yes	No	No	Social relations, Governance & empowerme	No	
3	1	Peer-reviewe	Abbott, J.; Ca	2007	Rivers as res	The Canadia	51(3)	280-302	Canadian As	Species man	Flooded Gra	Qualitative	Zambia	Yes	sites	No	Yes	No	No	Social relations, Governance & empowerme	No	
4	2	Conference p	Adeyemo, A.	2012	Lake Chad re	21st Century	May 27-June	Bari, Italy	ASABE	Resource ma	Large Lakes	?	Chad	Yes	sites	No	?	No	No	Governance & empowerment, Subjective we	No	
5	2	Conference p	Adeyemo, A.	2012	Lake Chad re	21st Century	May 27-June	Bari, Italy	ASABE	Resource ma	Large Lakes	?	Niger	Yes	sites	No	?	No	No	Governance & empowerment, Subjective we	No	
6	2	Conference p	Adeyemo, A.	2012	Lake Chad re	21st Century	May 27-June	Bari, Italy	ASABE	Resource ma	Large Lakes	?	Nigeria	Yes	sites	No	?	No	No	Governance & empowerment, Subjective we	No	
7	3	Peer-reviewe	Adam, Y. O.;	2013	Contribution	Agricultural S	117	90-97	Elsevier	Enterprises &	Tropical/Sub	?	Sudan	No	None	No	?	No	No	Economic living standards	No	
8	4	Peer-reviewe	Agea, J. G.; C	2009	Efficacy of fo	The Social Sc	4(3)	295-303	Medwell Pub	Policies & re	Tropical/Sub	?	Uganda	Yes	sites	No	?	No	No	Education, Social relations, Governance & er	No	
9	5	Peer-reviewe	Akyeampong	2011	Pro-poor tou	Journal of Su	19	197-213	NA	Area protect	Tropical/Sub	?	Ghana	No	None	No	?	No	No	Economic living standards, Education, Social	No	
10	6	Peer-reviewe	Aladuwaka, S	2010	Sustainable (	Gender and	18(1)	43-58	Taylor & Fra	Resource ma	Tropical/Sub	Qualitative	Sri Lanka	Yes	None	Continuous	Yes	No	No	Economic living standards, Health, Educatio	No	
11	7	Peer-reviewe	Aldon, M. E.	2011	Socio-cultura	Fisheries Res	107	112-121	Elsevier	Area protect	Tropical Cora	Mixed	Philippines	No	None	No	No	No	No	Material living standards, Economic living sta	No	
12	8	Peer-reviewe	Ali, T.; Ahma	2007	Impact of pa	Internationa	14	211-223	Taylor & Fra	Resource ma	Temperate B	Mixed	Pakistan	Yes	Presence/Ab	No	No	No	Yes	Material living standards, Economic living sta	Yes	Sustaina
13	9	Peer-reviewe	Allendorf, T.	2013	Gender and /	Society and I	26	962-976	NA	Area protect	Tropical/Sub	Qualitative	Myanmar	Yes	user groups	No	Yes	No	No	Material living standards, Economic living sta	No	
14	10	Peer-reviewe	Amati, C.	2013	We all voted	Journal of Ea	7(4)	650-670	Routledge	Area protect	Tropical/Sub	?	Kenya	No	None	No	?	No	No	Economic living standards, Education, Gover	No	
15	11	Peer-reviewe	Ambrose-Oji	2003	The contribu	Internationa	5	106-117	NA	Resource ma	Tropical/Sub	Quantitative	Cameroon	Yes	socio-econor	No	No	No	No	Environmental, Material living standards	No	
16	12	Peer-reviewe	Ameha, A.; N	2014	Impacts of a	Ecological Ec	97	162-171	Elsevier	Resource ma	Tropical/Sub	Qualitative	Ethiopia	Yes	presence/ab	Punctuated	No	No	Yes	Economic living standards	No	
17	13	Peer-reviewe	Anand, A.; Cl	2012	Homestays a	Mountain Re	32(2)	126-136	Internationa	Enterprises &	Montane Gra	Qualitative	Nepal	Yes	presence/ab	Continuous	Yes	No	No	Environmental, Economic living standards, S	No	
18	14	Peer-reviewe	Ansong, M.;	2011	Determinant	Internationa	7(2)	98-107	Taylor & Fra	Area protect	Tropical/Sub	Mixed	Ghana	Yes	sites	No	No	No	No		No	
19	15	Peer-reviewe	Antunes Zap	2012	The conflict I	Marine Policy			Elsevier	Area protect	Marine-Tem	Mixed	Brazil	No	None	No	No	No	No	Economic living standards, Subjective well-b	No	
20	16	Peer-reviewe	Aswani, S.; F	2007	Do Marine P	Coastal Man	35123:05:00	545-565	Taylor & Fra	Area protect	Tropical Cora	Mixed	Solomon Isla	Yes	presence/ab	Punctuated	No	Yes	Yes	Health, Subjective well-being	No	
21	17	Peer-reviewe	Aswani, S.; W	2004	Scientific Eva	Human Orga	63(3)	301-319	The Society f	Area protect	Tropical Cora	?	Solomon Isla	Yes	presence/ab	Punctuated	?	No	No	Environmental, Economic living standards, S	No	
22	18	Conference p	Azman, N.; H	2010	Public educa	Procedia Soc	7	504-511	NA	Area protect	Tropical/Sub	Qualitative	Malaysia	No	None	No	Yes	No	No	Economic living standards, Education, Social	Yes	Framew
23	19	Peer-reviewe	Baez, S.	2011	Notes: The "I	Fordham Law	80(2)	821-875	Fordham Un	Legislation, (	Tropical/Sub	?	Brazil, Indon	Yes	sites	No	?	No	No		No	
24	20	Peer-reviewe	Balciauskas,	2013	Forty years a	European Jo	60	155-158	NA	Species man	Temperate B	Qualitative	Lithuania	No	None	No	Yes	No	No	Material living standards, Health, Subjective	No	
25	21	Peer-reviewe	Baral, N.; He	2007	Resources us	Environment	34	64-72	Foundation f	Area manage	Tropical/Sub	Mixed	Nepal	Yes	projects	No	No	No	No	Environmental, Material living standards, Go	No	
26	22	Peer-reviewe	Baral, N.; Ste	2011	A comparati	Biodiversity (	20	2407-2426	Springer	Area manage	Tropical/Sub	Mixed	Nepal	Yes	projects	No	No	No	No	Environmental, Economic living standards, S	Yes	Commu
27	23	Peer-reviewe	Baral, N.; Ste	2011	Capital Stock	Society & Na	'24:10	1011-1026	Routledge	Area manage	Tropical/Sub	Mixed	Nepal	No	None	No	No	No	No	Environmental, Economic living standards, S	No	
28	24	Peer-reviewe	Barham, B. L	1999	Rain forest li	Unasylva	50	34-41	NA	Area manage	Tropical/Sub	Mixed	Peru	Yes	sites	No	No	No	No	Economic living standards, Social relations, S	No	
29	25	Peer-reviewe	Barker, A.	2005	Improving Lc	Coastal Educ	42	387-393	Journal of Cc	Area manage	Tropical Cora	Qualitative	Belize	Yes	projects	No	Yes	No	No	Environmental, Education, Social relations, S	No	
30	25	Peer-reviewe	Barker, A.	2005	Improving Lc	Coastal Educ	42	387-393	Journal of Cc	Area manage	Tropical Cora	Qualitative	Fiji	Yes	projects	No	Yes	No	No	Environmental, Education, Social relations, S	No	
31	25	Peer-reviewe	Barker, A.	2005	Improving Lc	Coastal Educ	42	387-393	Journal of Cc	Area manage	Tropical Cora	Qualitative	Indonesia	Yes	projects	No	Yes	No	No	Environmental, Education, Social relations, S	No	
32	25	Peer-reviewe	Barker, A.	2005	Improving Lc	Coastal Educ	42	387-393	Journal of Cc	Area manage	Tropical Cora	Qualitative	Philippines	Yes	projects	No	Yes	No	No	Environmental, Education, Social relations, S	No	
33	25	Peer-reviewe	Barker, A.	2005	Improving Lc	Coastal Educ	42	387-393	Journal of Cc	Area manage	Tropical Cora	Qualitative	Samoa	Yes	projects	No	Yes	No	No	Environmental, Education, Social relations, S	No	
34	26	Peer-reviewe	Barley Kincai	2014	Fisher's perc	Marine Polic	43	226-235	Elsevier	Area protect	Tropical Cora	Mixed	Tanzania, Un	Yes	user groups	No	No	No	No	Subjective well-being	No	
35	27	Peer-reviewe	Bartlett, C. Y	2009	Marine reser	Marine polic	33	673-678	elsevier	Area manage	Tropical Cora	Mixed	Vanuatu	Yes	projects	No	No	No	No	Environmental, Economic living standards, S	No	
36	28	Peer-reviewe	Baticados, D.	2000	Co-managen	Internationa	7(4)	343-355	Taylor & Fra	Species man	Tropical Cora	?	Philippines	Yes	None	Continuous	?	No	No	Material living standards, Social relations, Ge	No	
37	29	Peer-reviewe	Bauer, H.	2003	Local percep	Environment	30	175-181	Foundation f	Area manage	Tropical/Sub	Mixed	Cameroon	No	None	No	No	No	No	Environmental, Economic living standards, S	No	
38	30	Peer-reviewe	Bawa, K. S.; J	2007	Poverty, bioc	Agriculture, I	121	287-295	Elsevier	Area manage	Tropical/Sub	Mixed	India	Yes	None	Continuous	No	No	No	Environmental, Economic living standards, S	No	
39	31	Peer-reviewe	Bayala, J.; Sa	2013	Parklands fo	ScienceDirec	6	28-34	Elsevier	Area manage	Tropical/Sub	Mixed	Region - Afri	Yes	sites	No	No	No	No	Environmental, Economic living standards, H	Yes	Bespoke
40	33	Peer-reviewe	Becker, C. D.	2003	Synergy betv	Conservation	8(1)	1	The Resilienc	Resource ma	Tropical/Sub	Qualitative	Ecuador	Yes	None	Continuous	Yes	No	No	Environmental, Governance & empowermer	No	
41	34	Peer-reviewe	Behera, B.	2009	Explaining th	Ecological Ec	69	177-185	Elsevier	Area manage	Tropical/Sub	Mixed	India	Yes	socio-econor	No	No	No	No	Environmental, Economic living standards, S	No	
42	35	Peer-reviewe	Beitl, C. M.	2012	Shifting polic	Journal of Political Ecology			NA	Area manage	Mangroves	Mixed	Ecuador	No	None	No	No	No	No	Material living standards, Governance & em	No	



xkcd

Problem #3: the process is exceptionally labor intensive



A better solution

Can technology help?

SEP 23, 2016 @ 01:48 AM 41,342 VIEWS The Little Black B

How Machine Learning, Big Data And AI Are Changing Healthcare Forever

Bernard Marr, CONTRIBUTOR

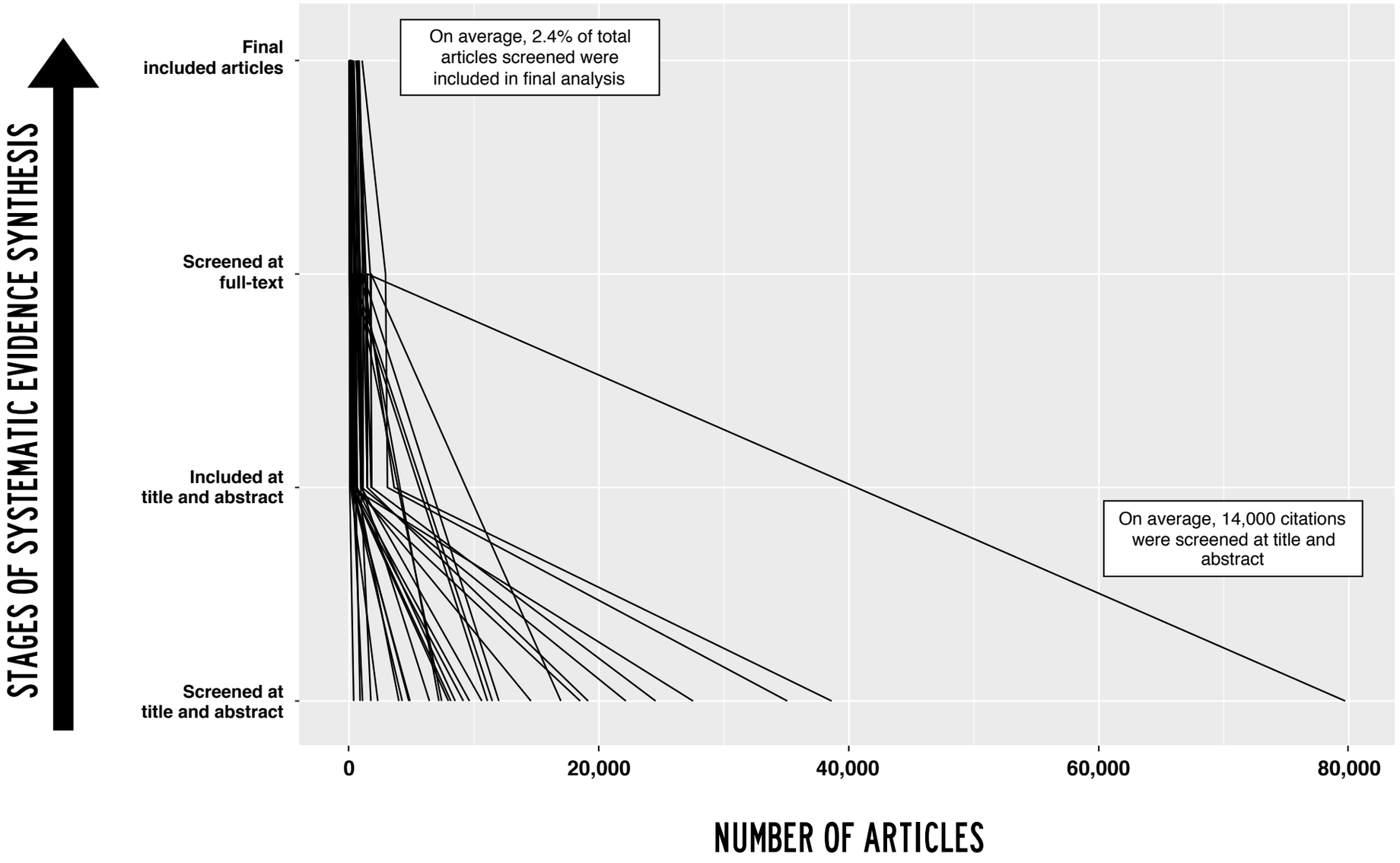
I write about big data, analytics and enterprise performance

FULL BIO

Opinions expressed by Forbes Contributors are their own.

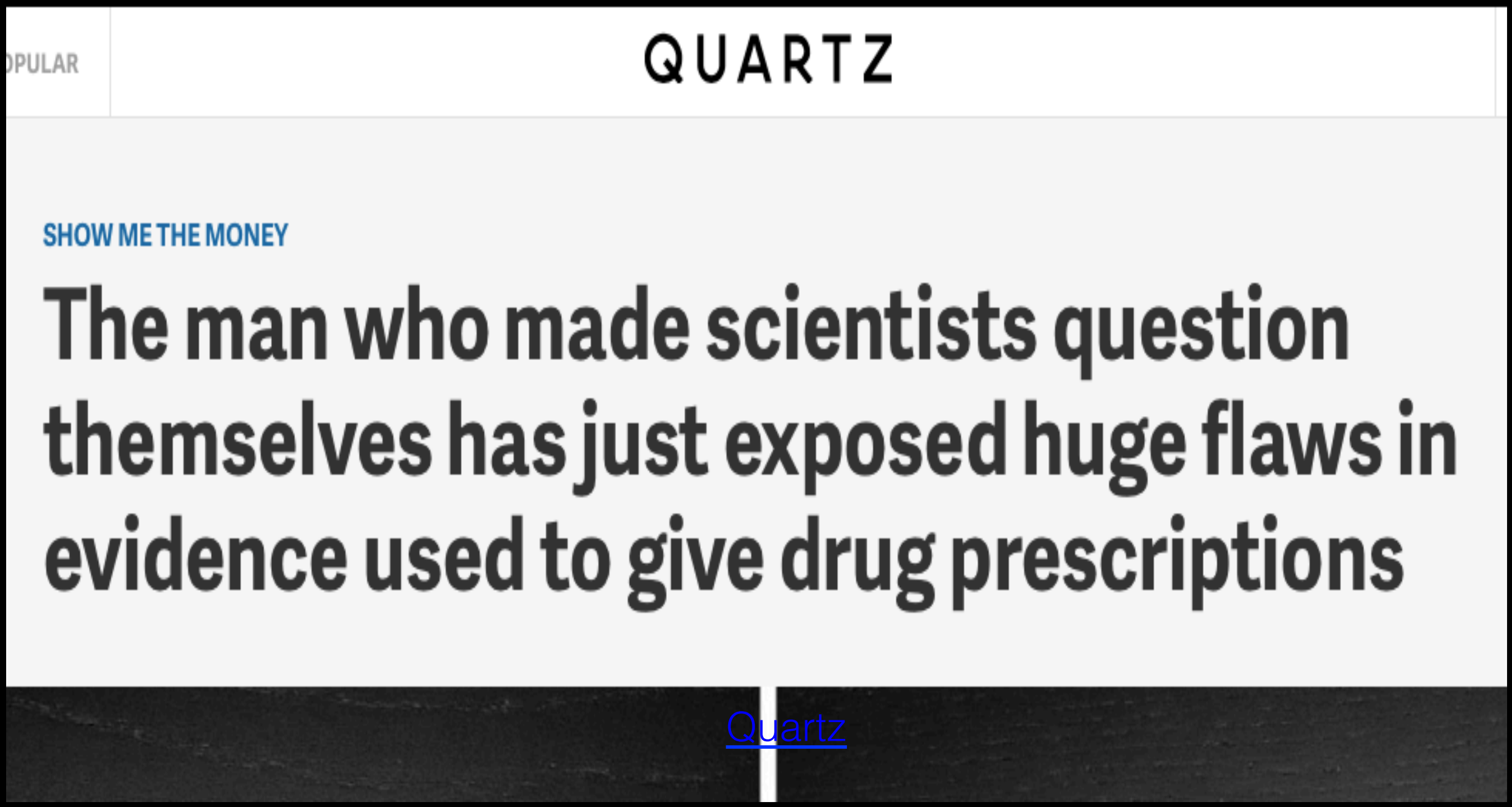
While robots and computers will probably never completely replace doctors and nurses, machine learning/deep learning and AI are transforming the healthcare industry, improving outcomes, and changing the way doctors think about providing care.

Machine learning is improving diagnostics, predicting outcomes, and just beginning to scratch the surface of personalized care.



Cheng et al. (supplemental)

Black box processes aren’t the silver bullet



Quartz

Less

↓


TERMINATOR
RISE OF THE MACHINES
THETERMINATOR@ALICEPOSTA.IT


Wolt
Wolt

↑

More

colandr

Signin Help About

CONSERVATION INTERNATIONAL + DataKind + SNAPP

SIGN IN SIGN UP

Email

caugustin@rsmas.miami.edu

Password

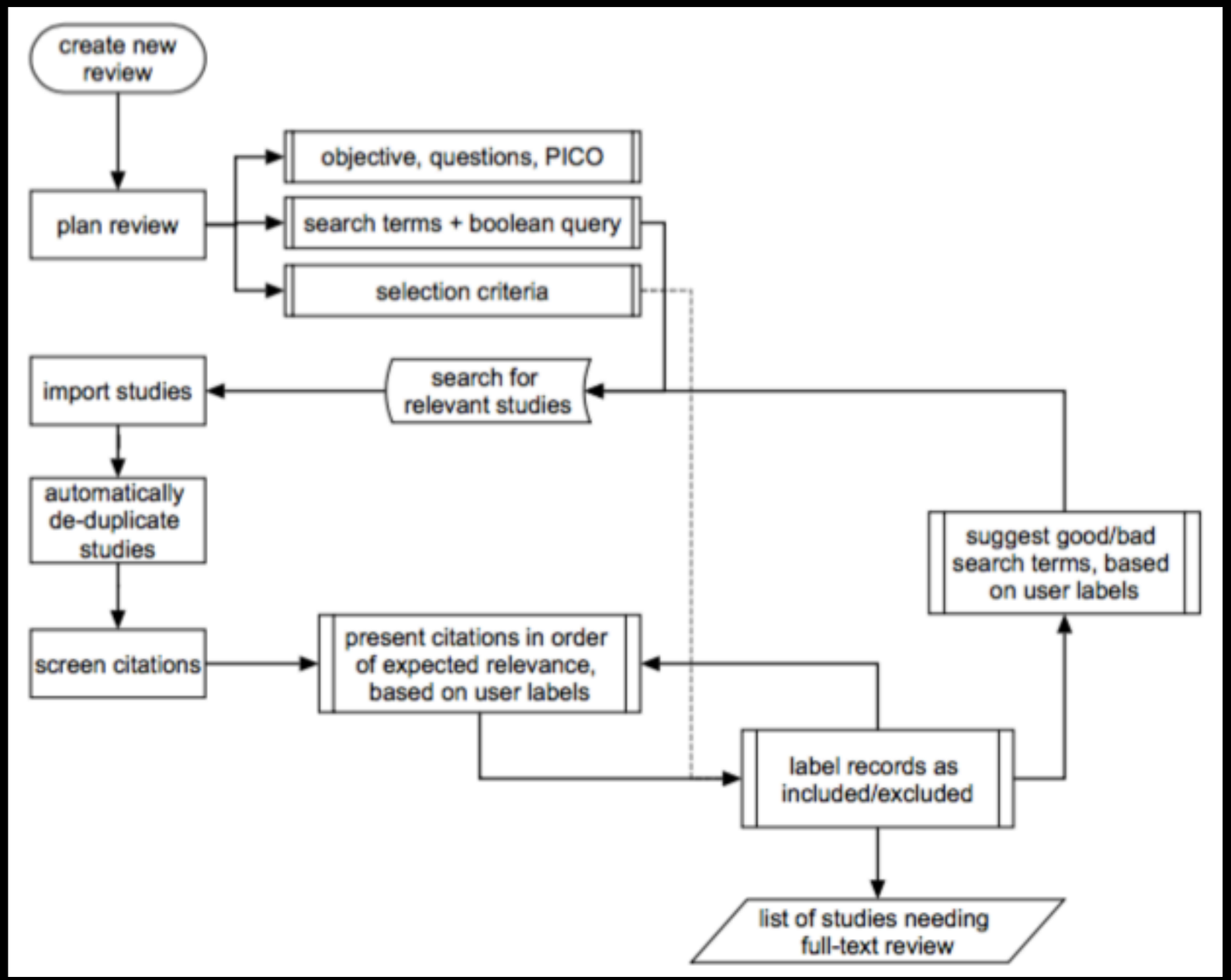
••••••••••

SIGNIN

Forgot your Password?

colandrapp.com

Tool framework



Flask
web development, one drop at a time

PostgreSQL

python 3

Scala

Image Magick

PDFBox

tesseract

Import Citations

FROM FILE | WITH A FORM | IMPORT HISTORY

Import from UPLOAD

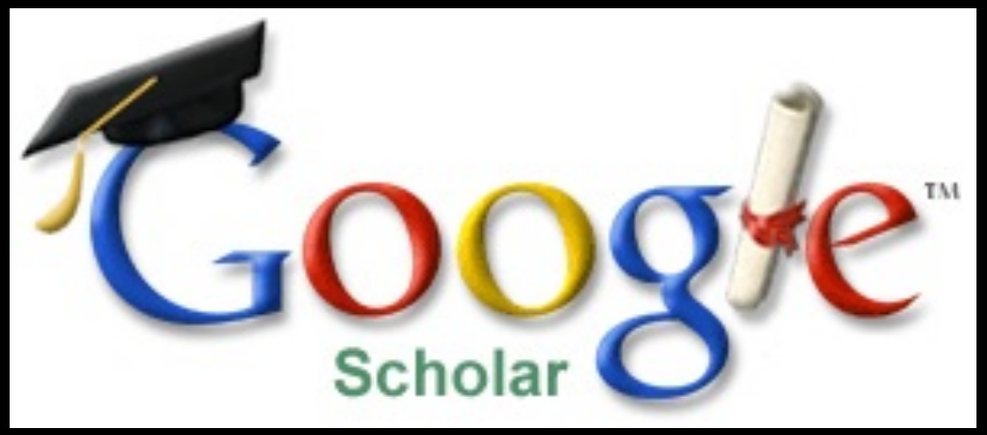
Data Source -- Select One --
database
gray literature

Source Name

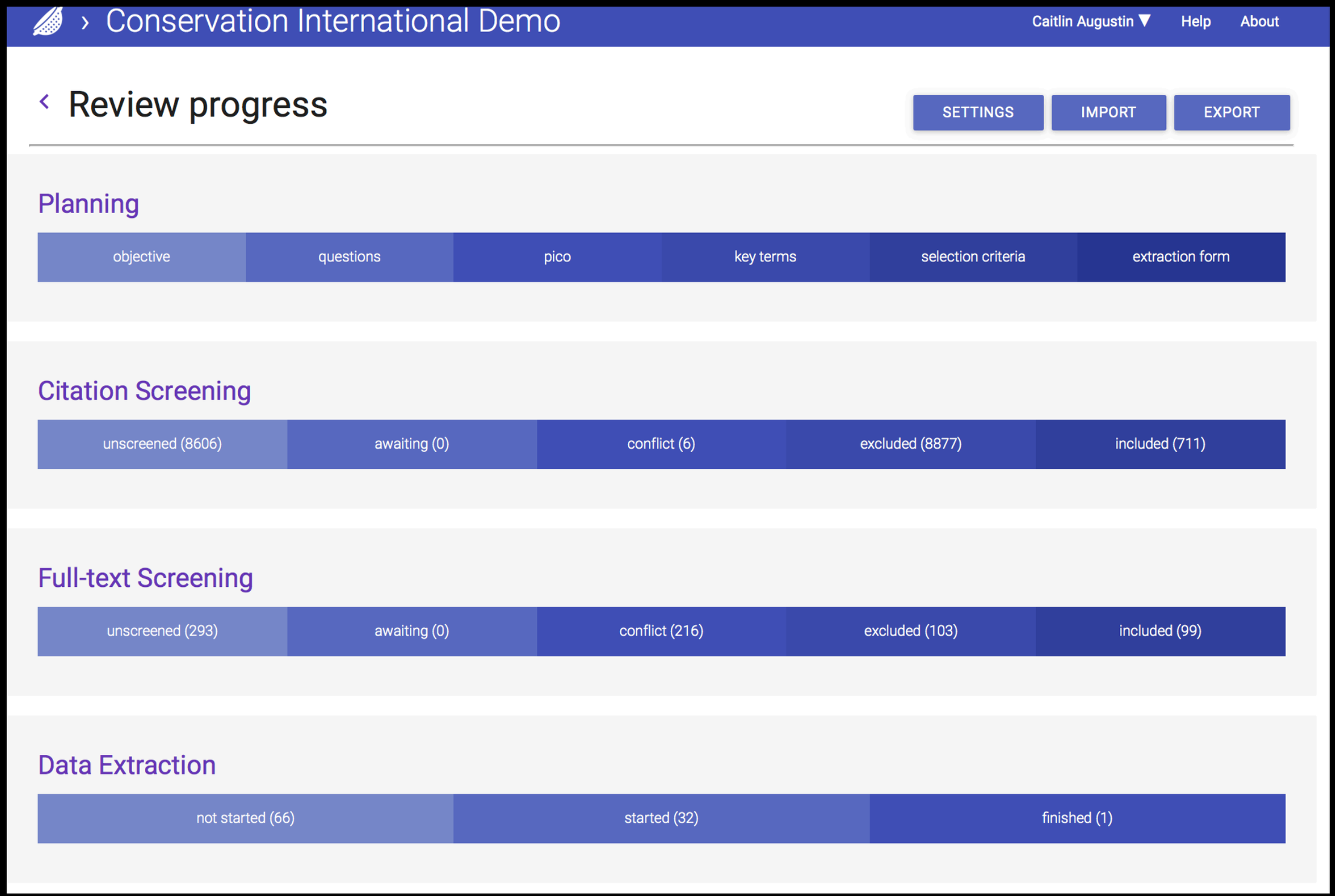
Source URL

Status not screened

IMPORT



Review set-up



System 1: deduplication & relevance ranking for first-pass screening

Citation Screening

SCREEN [8606] AWAITING [0] IN CONFLICT [6] EXCL NCLUDED [711]

HIGHLIGHTS_ON or Filter by tag: Relevance

Konijnendijk, C C,Lövei, G L,Mertz, O,Nielsen, I,Ravnborg, H M

Ecosystem services and **biodiversity** in developing countries

Biodiversity and Conservation (2007)

INCLUDE SKIP EXCLUDE

- [ ] location
- [ ] undefined pop.
- [ ] intervention type
- [ ] in-situ
- [ ] outcome

cancel ok

The concept of **ecosystem services** has become important for our understanding of the role of nature for maintaining **huma** essential to maintain **ecosystem services**? Many studies suggest that higher **biodiversity** allows a higher level of **ecosystem** little hard evidence, especially from tropical environments, to document the necessity of high **biodiversity** for provision of m valuation of **biodiversity** for **ecosystem services** and long-term studies and monitoring are needed to fully understand the c interface. This introduction briefly reviews some of the main arguments in this debate and provides an overview of the other **biodiversity** and **ecosystem** interactions in the context of the provision of **ecosystem services**, these papers address popula importance of dung beetles in agricultural **landscape**s, the **knowledge** and use of palms by local **communities**, bioprospecti **conservation** may have added benefits in terms of improved watershed functions and **health**. © 2007 Springer Science+Business Media B.V.

Keywords: Biodiversity,Bioprospecting,**Conservation**,**Ecological** economics,**Ecosystem services**,Environment,Local **knowledge**,Monitoring,Population,**Sustainable** financing,Valuation

+Tag

Citations are ranked by expected relevance depending on the availability and number of user-labeled examples

- 1st uses search terms from review planning: computes the amount of overlap between those terms and citations' title + abstract + keywords
- 2nd after enough examples have been labeled, uses distributional word vectors (word2vec) as features for a support vector classifier that predicts inclusion or exclusion; use confidence of that classification as expected relevance

System 2: second screening and metadata extraction

Full-paper upload

Behavioural thermoregulation in two freshwater fish species

INCLUDE

EXCLUDE

Marine Policy 42 (2013) 236–244

Contents lists available at SciVerse ScienceDirect

Marine Policy

journal homepage: www.elsevier.com/locate/marpol

Consumptive versus non-consumptive use of sea turtles? Stakeholder perceptions about sustainable use in three communities near Cahuita National Park, Costa Rica

Katharine A. Hart [a,*], Tim Gray [b], Selina M. Stead [a]

[a] School of Marine Science and Technology, Ridley Building 2, Claremont Road, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom
[b] School of Geography, Politics and Sociology, 5th Floor Claremont Tower, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom

Data Extraction - Label Review

Adaptive comanagement of a marine protected area network in Fiji.

LABEL SUMMARY

Sentences related to Intervention: law & policy — **Under Review**

1. Confidence: High

Included in the first iteration of the Kubulau EBM plan was a provision for review and amendment as necessary every 5 years (Wildlife Conservation Society 2009) . This first major revision was expedited to take advantage of new data and improved technical capacity within WCS and in response to donor deadlines . The willingness of Kubulau communities to make early in terms of increasing the area under management was likely due to positive attitudes toward existing protected areas , perceptions that the MPA network had

ACCEPT

SKIP

REJECT

2. Confidence: High

Yet conservation planning and network design have rarely been as dynamic or iterative as intended (Game et al . 2010) , and despite an increasing focus in the literature on the need for adaptive conservation strategies (e.g. , Grantham et al . 2010 ; McCook et al . 2010 ; Ban et al . 2011) , there remain few examples of adaptive management in practice (Holness & Biggs 2011 ; Roux & Foxcroft 2011) . Adaptive management is an iterative process of decision making whereby management strategies are changed or adjusted as new

3. Confidence: Medium

However , no changes were made to these rules . It is possible that communities see the role of the district areas as fulfilling food security objectives Conservation Biology Volume 27 , No. 6 , 2013 Weeks & Jupiter 1241 and the village tabu areas as providing for cultural objectives (ensuring abundant resources at times of social importance) . In common with other planning processes (e.g. , Green et al . 2009 ; Game et al . 2010) , we found habitat and prioritization maps a powerful tool for engaging communities and focusing group

Training data to find sentences that might indicate a label (provide provenance)

- The system over-predicts (predict sentences from a large number of the labels), so that the system can focus on recall, while human annotators can focus on precision
- For locations we use a "Named Entity Recognition" system to find mentioned locations in the document, and suggest these as labels
- For other metadata, we use global vectors for word representation (GloVe) and logistic regression to train a model of ranker-tags
- We show the sentences that best predict labels to the user, who can then use that information to pick the correct labels

Early performance

	Format: Ease of using specific GUI vs. non-specific formats	Error: Catching missed references, mis- assigned tags, duplicates	Efficiency: How many citations screened to find 100 included?
Case 1: Conservation & human well-being (McKinnon et al. 2016)	Version control issues when screening in Microsoft Excel. Oftentimes would crash the program. Multiple columns for exclusion criteria made for lots of unnecessary scrolling back and forth	Many duplicates still cropped up even after data was extracted. The deduplication function in Colandr allowed for us to find duplicates faster than by eye. Colandr also suggested tags for articles that upon closer read, were in fact an appropriate tag for that paper that we had mis-assigned by hand.	Colandr: 250 Manual: 1436
Case 2: Forests & poverty (Cheng et al. 2017)	Screening in EPPI Reviewer is comparative in format, allowing for multiple users and structured format to standardize criteria. However, costs for EPPI quickly rose as we added members to the review team.	Colandr allowed for quicker identification of key sentences that could lead to insight into document tags. Rather than reading through often dense text, it was very useful and efficient to view suggested sentences. While some of the these sentences were not always helpful, having them collated in one place streamlined the process.	Colandr: 167 Manual: 407
Case 3: Synergies, tradeoffs, equity in marine conservation	The GUI facilitated faster title and abstract screening with: clear text layout, highlighted keywords, radio buttons to select reasons for exclusion, and smooth transitions from one entry to another. Also facilitated screening on mobile devices.	Colandr's deduplication function eliminated the need for the reviewer to do this tedious process manually. In total, the app identified 70 duplicates and only missed 7 (90% success rate).	Colandr: <568 Manual: NA

Table 1. Assessing ability of Colandr to optimize process of evidence synthesis using three case Studies (Cheng, et al. supplemental)

Encountered challenges and limitations

- Designing a commercial-competitive product at a non-profit acceptable cost
- Data scarcity/NLP limitations
- Tempering expectations
- Languages

Where are we now, and where do we go from here?

Users

- Over 200 unique registered users, 76 of which are academic users, 30 of which are organizational users
    - Representing World Bank, Conservation International, Mayo Clinic, SEI-International, Stanford, Columbia University, Yale, Duke University, Princeton among others
- 274 reviews created spanning topics of conservation, medicine, education, climate change, marine stewardship and community engagement
- Multi-continent users: users from countries in North America, Europe, and Asia

Established community of practice

User studies + tool inter-comparison studies

Topic expansion

Framework and data set citations / evidence linkages

Trainings and workshops

colandrcommunity.com

Research

- We're current testing colandr a number of different ways and we'd love to add you to our studies!
- Protocols are testing colandr against "standard" review processes
  - University of Illinois work
- Protocols are looking at pain points of using colandr
  - User studies under development
- Many more ideas!

Conservation Biology

Letter | Open Access

Using machine learning to advance synthesis and use of conservation and environmental evidence

S.H. Cheng ✉, C. Augustin, A. Bethel, D. Gill, S. Anzaroot, J. Brun, B. DeWilde, R.C. Minnich, R. Garside, Y.J. Masuda, D.C. Miller, D. Wilkie, S. Wongbusarakum, M.C. McKinnon

First published: 12 April 2018 | https://doi.org/10.1111/cobi.13117

Article impact statement: Machine learning optimizes process of systematic evidence synthes ... More

This article has been accepted for publication and undergone full peer review but has not been ... More

Want to use colandr?

Grab your lunch and join us at 13:00

SIFTING THROUGH EVIDENCE USING COLANDR: APPLIED MACHINE LEARNING FOR SYNTHESIS

in AMPHITHEATER CAQUOT (ground floor)

Questions?

Contact:
Caitlin M. Augustin
caitlin@datakind.org
colandrcommunity.com

@augustincaitlin

Contact:
Samantha H. Cheng
Samantha.Cheng@asu.edu
natureandpeopleevidence.org

@pilesofsquid