



Mercator Research Institute on
Global Commons and Climate Change gGmbH

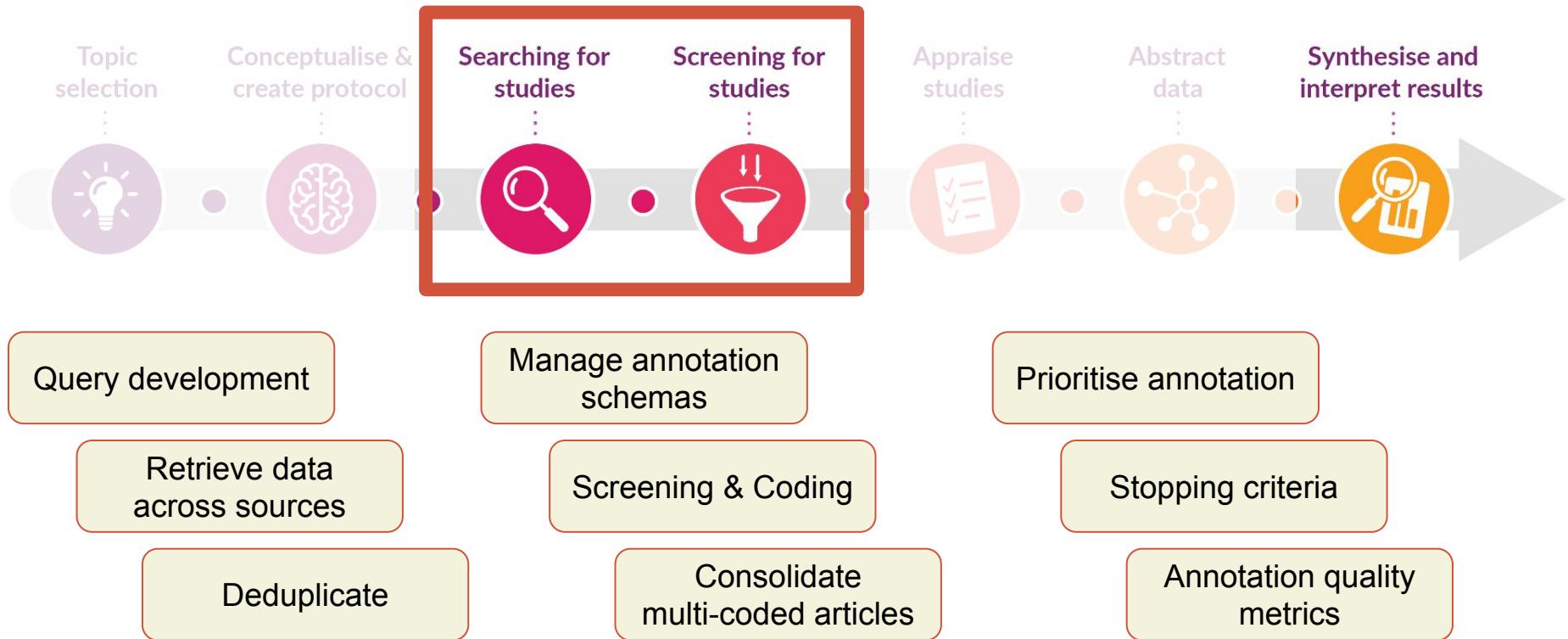
Computer-aided research synthesis

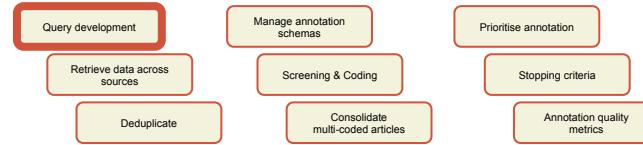


Overview of the NACSOS screening platform
developed in the APSIS group

Max Callaghan
Tim Repke

Steps in a systematic review





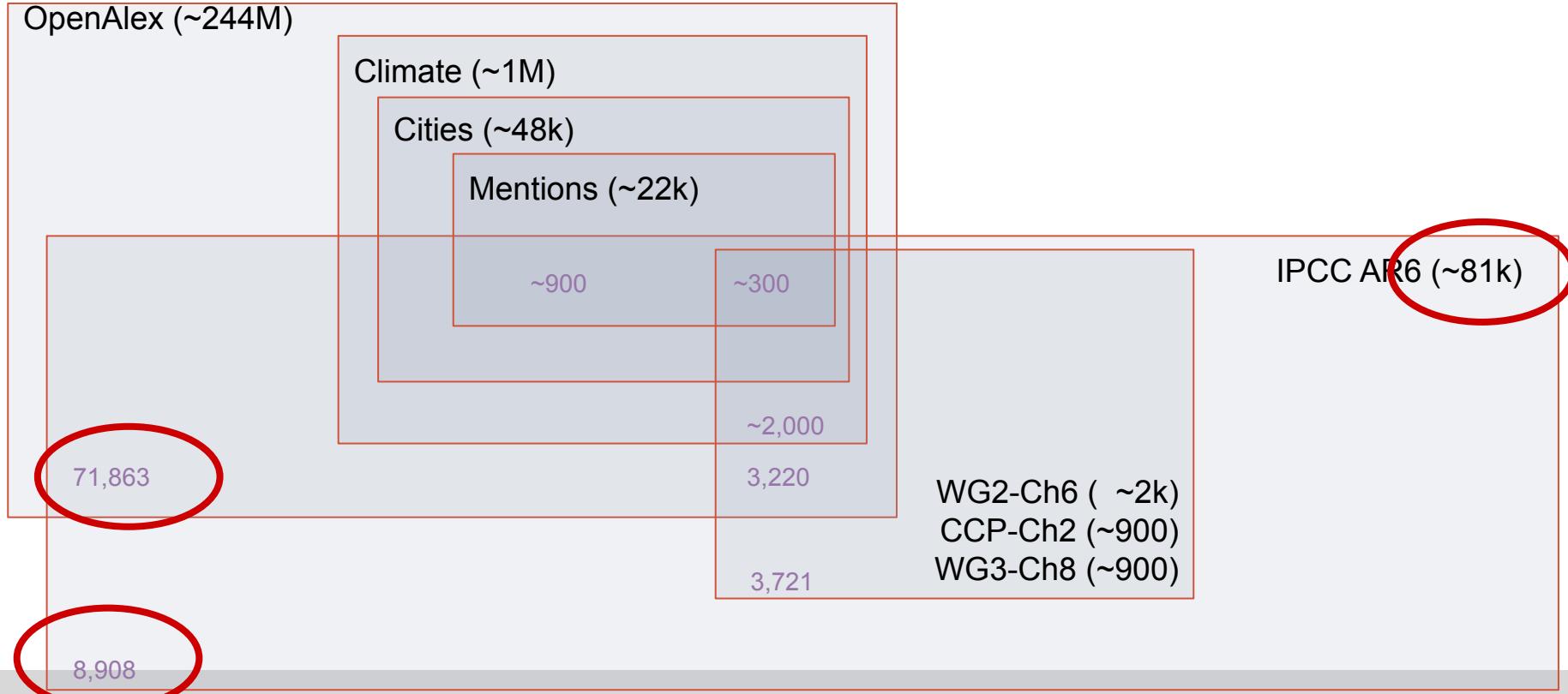
Query development

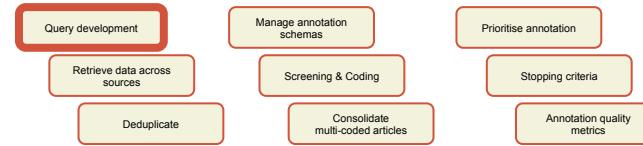
3 OpenAlex

- Open source scientific database
- 244M publications (bibliographic meta-data, citations, ...)
- Self-hosted mirror for full control (solr and postgres)
 - Snowballing via citations or similarities (vector search)
 - Better understanding of wildcard expansion
 - Reference metrics on entire fields or all of science

Coverage of OpenAlex

OpenAlex (~244M)

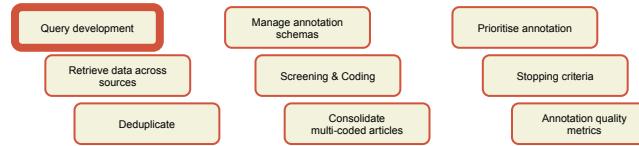




Query development

- Query OpenAlex (our solr mirror) directly from the platform
- Compare to “gold standard” references (*soon, probably*)
- Compare different queries (*soon, possibly*)
- Keyword recommendations (*let's see*)
- Similar documents (*maybe*)
- Follow citations (*likely*)

Query development



- Query OpenAlex (our solr mirror) directly from the platform
- Compare to “gold standard”
- Compare different query engines
- Keyword recommendation
- Similar documents (maybe)
- Follow citations (likely)
- Instant topic maps (??)

apache solr

options

Results

All retrieved results (141)

Apache Solr - Wikipedia

Solr (pronounced “solar”) is an open-source search engine written in Java, from the Apache Lucene project. It includes full-text search, hit highlighting, full-text search, indexing, dynamic clustering, database integration, and more.

https://en.wikipedia.org/wiki/Apache_Solr

BING GOOGLE WIKIPEDIA

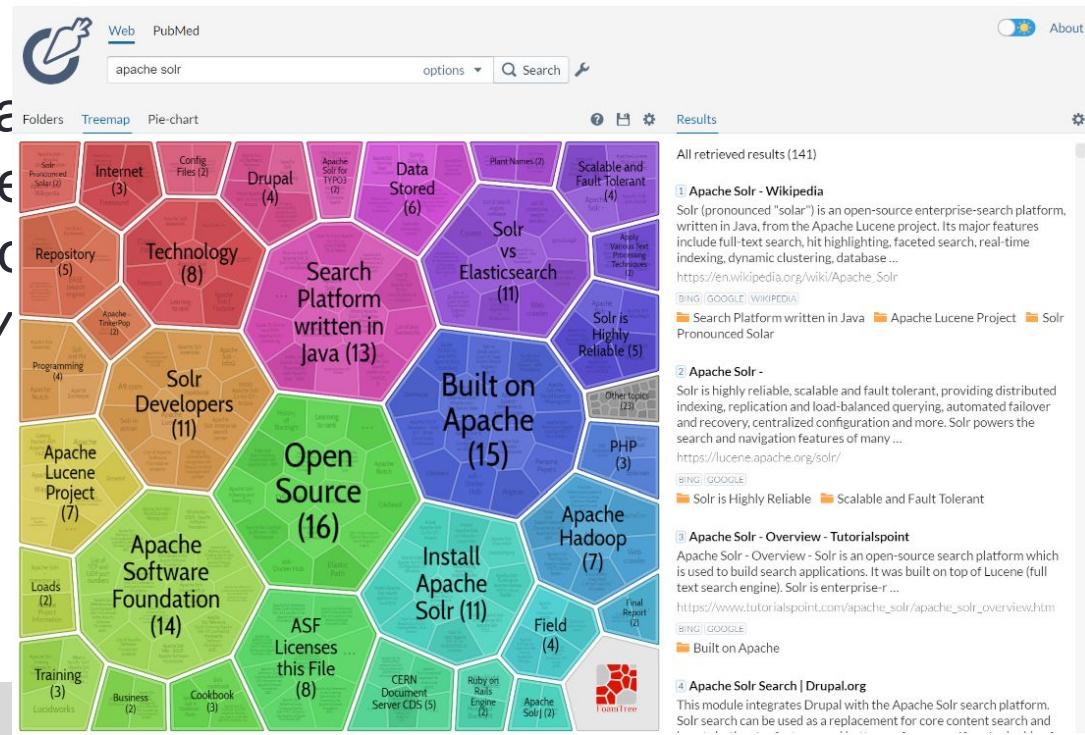
Search Platform written in Java

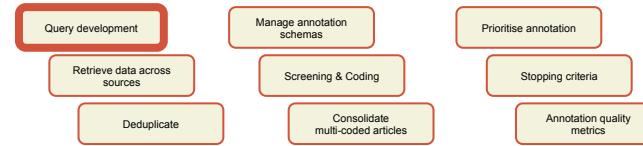
Apache Solr -

Solr is highly reliable, scalable and fault tolerant. It uses indexing, replication and load-balanced query processing and recovery, centralized configuration and administration, and search and navigation features of many ...

https://lucene.apache.org/solr/

BING GOOGLE





Query development

Projects About tim.repke ▾

Search OpenAlex (Solr)

Query setup

"climate change" and "CDR"

Limit: 20, Offset: 0, OP: AND, Field: title_abstract, Parser: edismax

Histogram: 1990 - 2024

Query: Found 321 works in 1123ms

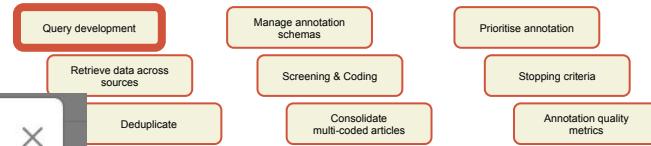
Is carbon dioxide removal 'mitigation of climate change'?

Carbon dioxide removal (CDR) is often characterized as separate from climate change mitigation. Discussion of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on mitigation apply. Similarly, many climate policy processes have ignored CDR. CDR may have been discursively held separate from 'mitigation' due to a partial conceptual overlap with 'geoengineering'. We unpack how the 'mitigation of climate change' – as defined in the United Nations Framework Convention on Climate Change and its Paris Agreement – includes CDR as defined by the Intergovernmental Panel on Climate Change. We point to important implications and opportunities for strengthening governance by enhanced clarity regarding parties' obligations, principled equitable distribution of removal efforts, prioritization of rapid emissions reductions and careful paths to long-term removals, and a need for considering sustainability and human rights issues in the pursuit of CDR.

2021 Matthias Honegger · William C. G. Burns · David R. Morrow

The influence of learning about carbon dioxide removal (CDR) on support for mitigation policies

A wide range of carbon dioxide removal (CDR) strategies has been proposed to address climate change. As most CDR strategies are unfamiliar to the public, it is unknown how increased media and policy attention on CDR might affect public sentiment about climate change. On the one hand, CDR poses a potential moral hazard: if people perceive that CDR solves climate change, they may be less likely to support efforts to reduce carbon emissions. On the other hand, the need for CDR may increase the perceived severity of climate change and, thus, increase support for other types of mitigation. Using an online survey of US adults (N = 984), we tested these competing hypotheses by exposing participants to information about different forms of CDR. We find that learning about certain CDR strategies indirectly reduces support for mitigation policies by reducing the perceived threat of climate change. This was found to be true for participants who read about CDR in general (without mention of specific strategies), bioenergy with carbon capture and storage, or direct air capture. Furthermore, this risk compensation pattern was more pronounced among political conservatives than liberals, although in some cases, was partially offset by positive



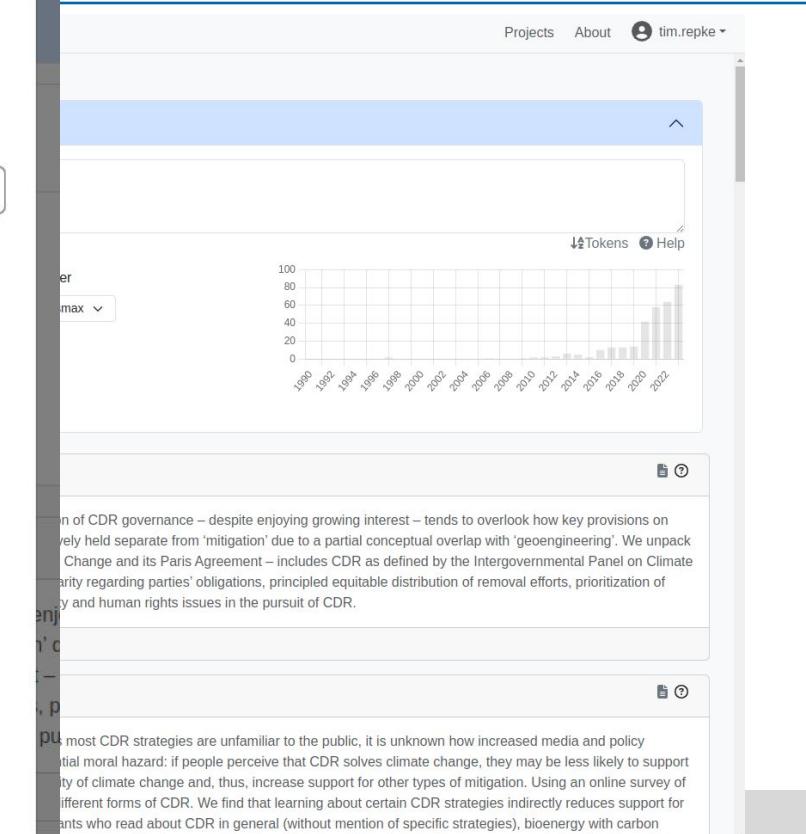
Token Wildcard Expansion

Wildcards (e.g. `clim*`) are very bad for query performance and add a lot of uncertainty. It is much better to explicitly expand those (e.g. `(climate OR climatic)`) and in the process exclude false-positives like `climb`, `climax`, ...

Results include term frequency (ttf) and document frequency (df).

trad 

<code>traditional</code> (4,367,640 3,336,034)	<code>trade</code> (2,283,075 1,258,966)	<code>tradition</code> (741,911 571,371)
<code>traditionally</code> (469,056 452,426)	<code>traditions</code> (344,721 277,624)	<code>trading</code> (473,445 267,987)
<code>tradicional</code> (127,853 106,850)	<code>tradeoff</code> (104,027 84,757)	<code>traded</code> (91,789 68,194)
<code>traders</code> (102,810 61,627)	<code>traduction</code> (83,917 57,091)	<code>tradeoffs</code> (68,101 56,926)
<code>tradicionales</code> (63,504 55,028)	<code>tradicional</code> (108,371 54,205)	<code>trades</code> (63,587 48,194)
<code>traduit</code> (47,691 45,129)	<code>tradicionais</code> (45,346 36,977)	<code>traducción</code> (44,601 34,270)
<code>tradición</code> (34,400 30,833)	<code>trademark</code> (57,896 30,607)	<code>tradicionalmente</code> (31,277 30,366)
<code>tradisi</code> (73,246 28,825)	<code>tradicion</code> (31,723 26,931)	<code>traditionnelle</code> (24,946 22,342)
<code>tradução</code> (30,221 19,116)	<code>traditionnelles</code> (20,291 18,486)	<code>tradiciones</code> (20,334 17,648)
<code>tradição</code> (19,750 16,574)	<code>traducción</code> (28,661 16,187)	<code>trader</code> (21,439 15,380)
<code>traditionnels</code> (16,714 15,022)	<code>traditionnel</code> (15,655 14,048)	<code>trademarks</code> (19,576 12,831)
<code>traduire</code> (14,010 12,736)	<code>tradicao</code> (15,428 12,401)	<code>tradizione</code> (13,908 12,394)
<code>trad</code> (17,082 11,847)	<code>tradable</code> (18,884 11,597)	<code>traditionellen</code> (12,698 11,532)
<code>traditionnellement</code>	<code>traductions</code> (13,149 10,265)	<code>traduzione</code> (12,246 10,032)



Projects About  tim.repke

Tokens Help

1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012 2014 2016 2018 2020 2022

Annotation of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on...
...ely held separate from 'mitigation' due to a partial conceptual overlap with 'geoengineering'. We unpack...
...Change and its Paris Agreement – includes CDR as defined by the Intergovernmental Panel on Climate...
...arity regarding parties' obligations, principled equitable distribution of removal efforts, prioritization of...
...y and human rights issues in the pursuit of CDR.

Annotation of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on...
......ely held separate from 'mitigation' due to a partial conceptual overlap with 'geoengineering'. We unpack...
......Change and its Paris Agreement – includes CDR as defined by the Intergovernmental Panel on Climate...
......arity regarding parties' obligations, principled equitable distribution of removal efforts, prioritization of...
......y and human rights issues in the pursuit of CDR.

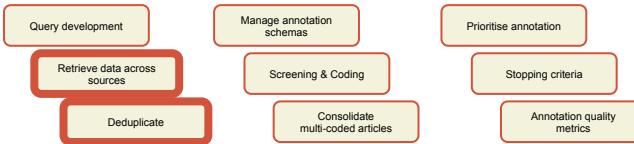
Annotation of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on...
......ely held separate from 'mitigation' due to a partial conceptual overlap with 'geoengineering'. We unpack...
......Change and its Paris Agreement – includes CDR as defined by the Intergovernmental Panel on Climate...
......arity regarding parties' obligations, principled equitable distribution of removal efforts, prioritization of...
......y and human rights issues in the pursuit of CDR.

Annotation of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on...
......ely held separate from 'mitigation' due to a partial conceptual overlap with 'geoengineering'. We unpack...
......Change and its Paris Agreement – includes CDR as defined by the Intergovernmental Panel on Climate...
......arity regarding parties' obligations, principled equitable distribution of removal efforts, prioritization of...
......y and human rights issues in the pursuit of CDR.

Annotation of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on...
......ely held separate from 'mitigation' due to a partial conceptual overlap with 'geoengineering'. We unpack...
......Change and its Paris Agreement – includes CDR as defined by the Intergovernmental Panel on Climate...
......arity regarding parties' obligations, principled equitable distribution of removal efforts, prioritization of...
......y and human rights issues in the pursuit of CDR.

Annotation of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on...
......ely held separate from 'mitigation' due to a partial conceptual overlap with 'geoengineering'. We unpack...
......Change and its Paris Agreement – includes CDR as defined by the Intergovernmental Panel on Climate...
......arity regarding parties' obligations, principled equitable distribution of removal efforts, prioritization of...
......y and human rights issues in the pursuit of CDR.

Data gathering (importing queries)



- Import data from
 - Scopus (via CSV)
 - Web of Science (via RIS)
 - OpenAlex (via solr)
 - Any other source (you just need to translate it to the platform format)
- Automatic deduplication of publications during import

Create new data import

An "import" provides the scope for how data enters a project. You can either configure a query to import from external scholarly databases, Twitter, or upload files directly. Please note, that only one "type" of data can exist in a project. This project is configured for **academic**.

Basic information

Name for this import

New import

Description of this import

Select import type

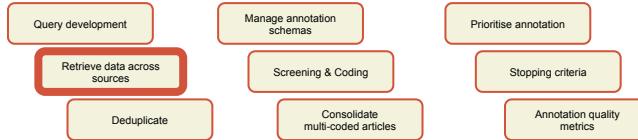
Select import type

- Upload JSON file (AcademicItemModel)
- Upload OpenAlex file**
- Import from OpenAlex (Solr)
- Upload Scopus CSV file(s)
- Upload Web of Science text file(s)

Import stats

load

Dataset insights



*more to come!



Projects About tim.repke

Overview

Import

Dataset

Explore

Statistics

Export

Annotation

Artefacts

Pipelines

Annotations

Project

Project statistics

Core stats

This project contains 51625 items imported via 8 imports. There are 3 annotation schemes and 22 assignment scopes amounting to a total of 129236 labels for 5491 (10.64%) unique documents.

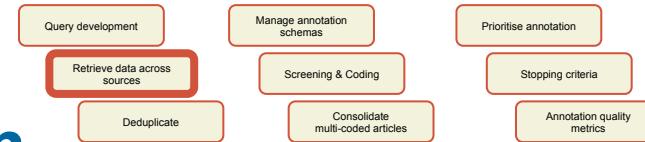
Annotator Leaderboard

1. **Niklas Döbbeling**: 94954 labels for 4385 items
2. **Alessandra Landa**: 4827 labels for 1758 items
3. **Arianna Avallone**: 2217 labels for 1057 items
4. **Klaas Miersch**: 12534 labels for 868 items
5. **Jan Minx**: 10140 labels for 591 items
6. **Finlay Hatch**: 2398 labels for 361 items
7. **Leonhard Schneider**: 2166 labels for 360 items

Histogram

Items

Year	Items
1990	~100
1993	~200
1996	~300
1999	~500
2002	~1,000
2005	~1,500
2008	~2,000
2011	~2,500
2014	~3,000
2017	~3,500
2020	~4,000
2023	~3,500



Searching on the platform

Search the project dataset
with our expressive query
language, e.g.

- Title, abstract, publication year, annotations,

The screenshot shows the NACoSOS platform's navigation menu on the left. The menu items are: Overview, Import, Dataset (highlighted in blue), Explore, Statistics, Export, Annotation, Artefacts, Pipelines, Annotations (with a gear icon), and Project. The 'Dataset' item is currently selected.

Urban Planning and Policy Faces Climate Change

Scientific opinion is overwhelmingly convinced that global temperatures are likely to continue to rise with concomitant extreme weather patterns and events. There is a protean body of scientific literature available on global warming and climate change, on modelling techniques for estimating 'heat island' and related effects in urban areas and on the design of more environmentally sustainable buildings. But the planning knowledge on climate change and its implications for human settlement is just emerging. This paper examines the state of planning knowledge in the area of climate change and urban settlement and seeks to assist planners in responding to climate change.

2006 · Edward J. Blakely · John Dee · Rafael E. Pizarro

Trade-off between wheat yield and drainage under current and climate change conditions in northeast Germany

In the temperate climate of northeast Germany, a trade-off exists between water use in agricultural crop production and groundwater recharge which is important for urban water supply, irrigation, forestry and peat protection. The APSIM-Nwheat model was used to analyze the impact of climate change scenarios on deep drainage (DD), the water loss below the maximum root zone as the main source of groundwater recharge, and wheat production for two main soil types. A linear and a nonlinear climate scenario were used in this study: The linear scenario for 2001–2050 was based on a simple linearly modified historical climate record from 1951 to 2000. The nonlinear scenario used the same 1951–2000 historical climate record but combined it nonlinearly with a Global Circulation Model climate scenario for 2001–2050. Simulation results showed different distributions of deep drainage and grain yield with the linear and nonlinear scenarios, but no difference in the 50-year averages. Hence, a linear manipulation of climate records can be as effective for climate change impact studies on deep drainage and grain yield as nonlinearly manipulated climate data, if long-term average changes are of main interest. The simulation results indicated that a trade-off between deep drainage and grain yields can be potentially controlled through N management. However, such control mechanism was more effective under current climate conditions than under future climate and on a better water-holding silt soil compared to a poorer water-holding loamy sand. It is suggested that areas with poor water-holding soils should be managed extensively for groundwater recharge harvesting while better water-holding soils should be used for high input grain production.

2006 · Gerd Wessolek · Senthil Asseng

The Science and Politics of Global Climate Change: A Guide to the Debate

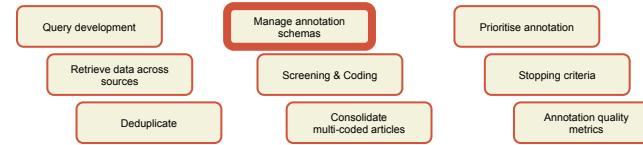
Search options

Number of documents: 479 / 112,995

Help (title: "climate change" OR abstract: CCS AND YR: >=2000 AND <2008)

Query

Reset

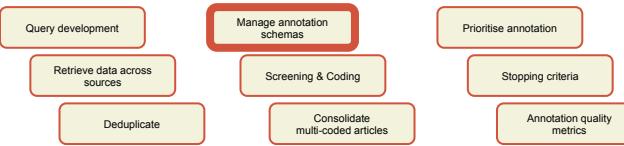


Annotation schema

Arbitrarily complex (doc-level) from screening to coding

- Simple yes/no labels
- Text fields (e.g. for comments)
- Labels with several options
- Multi-labels (for tagging multiple classes)
- Nesting of labels for “conditional labels”

Annotation schema



Related to cities •

Yes

No

Maybe

Explicit mention of one or more cities •



Climate

document is about climate change



Comment

↳ screening to coding

↳
“le classes)
al labels”

Annotation

Related to cities

Yes

No

Maybe

Explicit mention of one or more cities



Climate

document is about climate change



Comment

City related

Annotations for IPCC references related to cities

Note: Please use positive numbers only for choice values (zero included).

Related to cities cities Hint message

Type
Max. Repeat
 Required
 Use dropdown

single
1

Choices:

Yes
 1
 Hint message

Sub-annotations:

Explicit mention of one or more cities
 mention
 Hint message

Type
Max. Repeat
 Required

bool
1

Add Label

No
 0
 Hint message

Maybe
 2
 Hint message

Sub-annotations:

Explicit mention of one or more cities
 mention
 Hint message

Type
Max. Repeat
 Required

bool
1

Add Label

Add choice

Climate
 climate
 document is about climate ch

Type
Max. Repeat
 Required

bool
1

Add Label

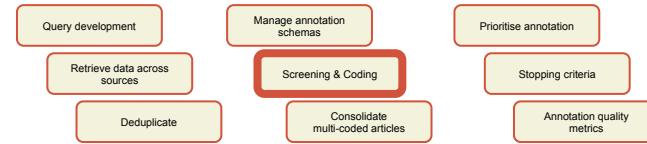
Add choice

Comment
 com
 Hint message

Type
Max. Repeat
 Required

str
1

Add Label



Annotating

Intuitive annotation interface for users

Projects About  tim.repke ▾



NACoS

- Overview
- Import
- Dataset
- Annotation
- Artefacts
- Pipelines
- Annotations
- Project

The Genome of Cardinium cBtQ1 Provides Insights into Genome Reduction, Symbiont Motility, and Its Settlement in *Bemisia tabaci*

Many insects harbor inherited bacterial endosymbionts. Although some of them are not strictly essential and are considered facultative, they can be a key to host survival under specific environmental conditions, such as parasitoid attacks, climate changes, or insecticide pressures. The whitefly *Bemisia tabaci* is at the top of the list of organisms inflicting agricultural damage and outbreaks, and changes in its distribution may be associated to global warming. In this work, we have sequenced and analyzed the genome of Cardinium cBtQ1, a facultative bacterial endosymbiont of *B. tabaci* and propose that it belongs to a new taxonomic family, which also includes *Candidatus Amoebophilus asiaticus* and *Cardinium cEper1*, endosymbionts of amoeba and wasps, respectively. Reconstruction of their last common ancestors' gene contents revealed an initial massive gene loss from the free-living ancestor. This was followed in Cardinium by smaller losses, associated with settlement in arthropods. Some of these losses, affecting cofactor and amino acid biosynthetic encoding genes, took place in Cardinium cBtQ1 after its divergence from the Cardinium cEper1 lineage and were related to its settlement in the whitefly and its endosymbionts. Furthermore, the Cardinium cBtQ1 genome displays a large proportion of transposable elements, which have recently inactivated genes and produced chromosomal rearrangements. The genome also contains a chromosomal duplication and a multicopy plasmid, which harbors several genes putatively associated with gliding motility, as well as two other genes encoding proteins with potential insecticidal activity. As gene amplification is very rare in endosymbionts, an important function of these genes cannot be ruled out.

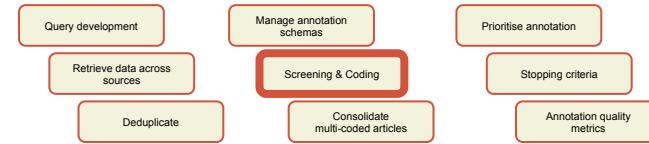
2014

Diego Santos-Garcia · Pierre-Antoine Rollat-Farnier · Francisco J. Beitia · Einat Zchori-Fein · Fabrice Vavre · Laurence Mouton · Andrés Moya · Amparo Latorre · Francisco J. Silva

Annotation Panel

- Show scheme description
- Is about cities
- Mentions one or more cities
- Topic: Adaptation
- Topic: Mitigation
- Topic: Impacts

Save & Previous Save & Next



Annotating

Intuitive annotation interface for users

The screenshot shows the NACoS annotation interface. The left sidebar has a 'Annotation' tab selected. The main area displays a document summary: 'The Genome of Cardinium cBtQ1 Provides Insights into Genome Reduction, Symbiont Motility, and Its Settlement in Bemisia tabaci'. Below the summary is a large yellow box containing the following text:

- Progress tracking
- Keyboard support
- Mobile friendly
- Customisable keyword highlighting

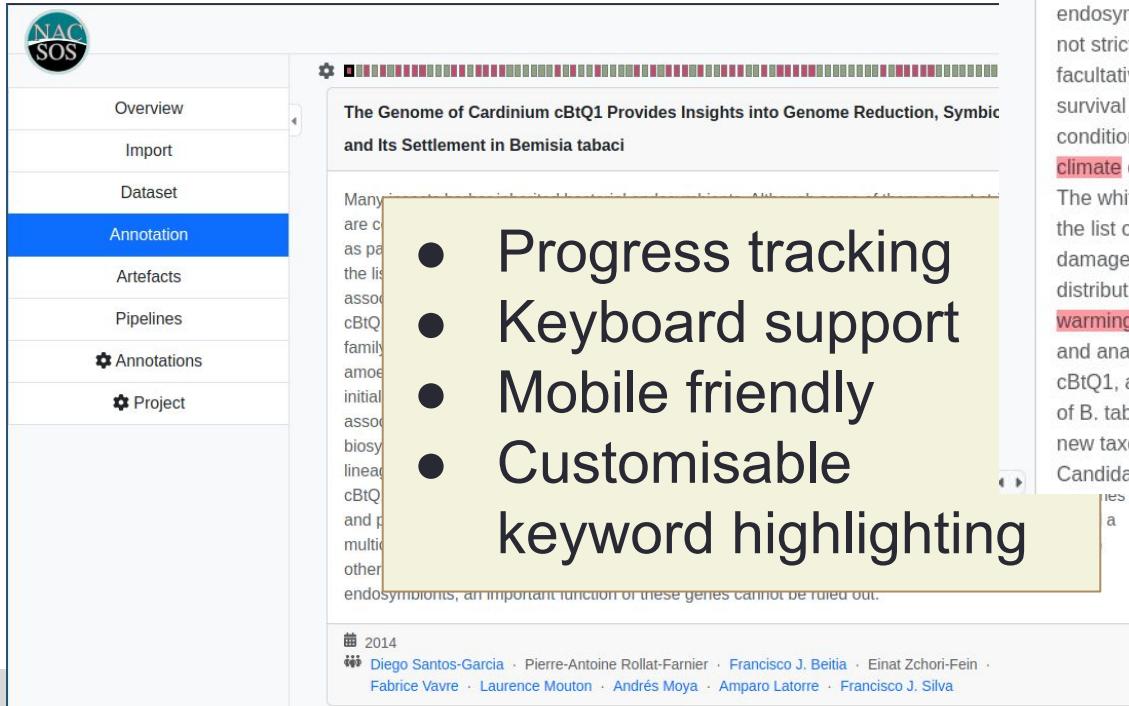
On the right is the 'Annotation Panel' with several filter options and buttons:

- Show scheme description
- Is about cities:
- Mentions one or more cities:
- Topic: Adaptation:
- Topic: Mitigation:
- Topic: Impacts:

At the bottom are 'Save & Previous' and 'Save & Next' buttons.

Annotating

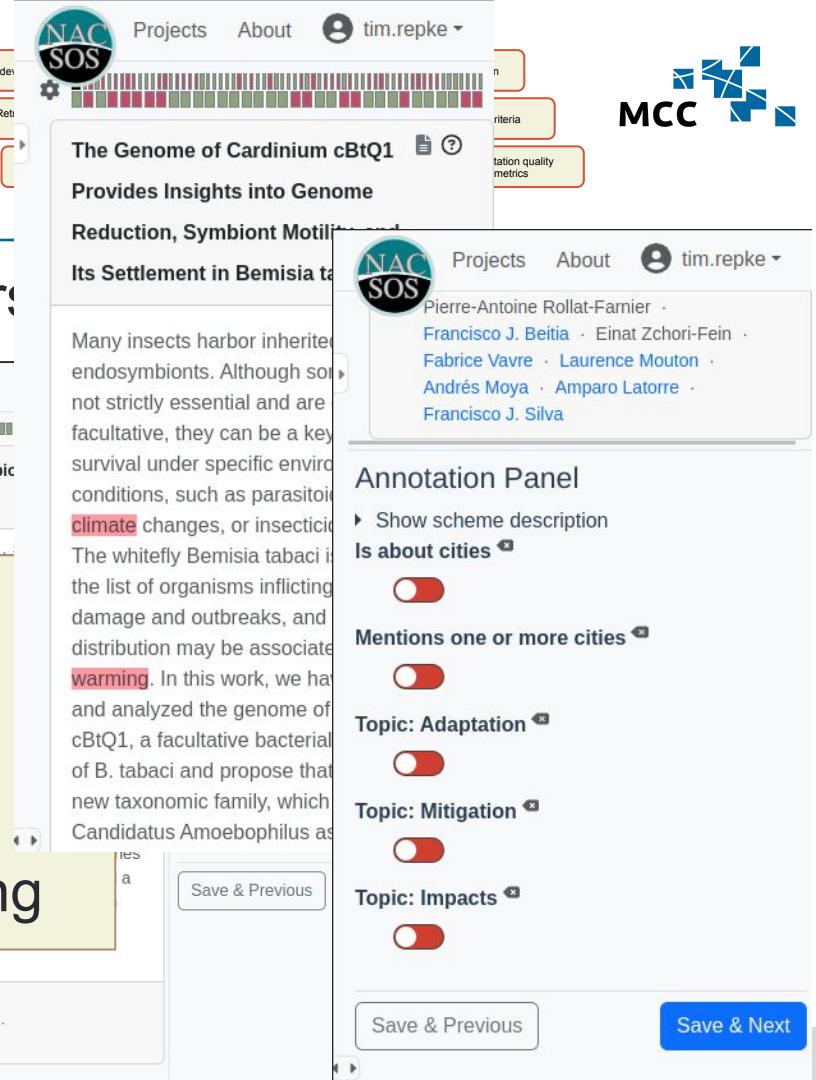
Intuitive annotation interface for users



The screenshot shows the NAC-SOS annotation interface. The left sidebar has a navigation menu with 'Annotation' selected. The main content area displays a list of publications, with one entry expanded to show its details. The expanded entry is for the paper: 'The Genome of Cardinium cBtQ1 Provides Insights into Genome Reduction, Symbiotic Motility, and Its Settlement in Bemisia tabaci'. The text content of the paper is visible, and a large yellow box highlights the following features:

- Progress tracking
- Keyboard support
- Mobile friendly
- Customisable keyword highlighting

At the bottom of the main content area, there is a footer with the year 2014 and a list of authors: Diego Santos-Garcia, Pierre-Antoine Rollat-Farnier, Francisco J. Beitia, Einat Zchori-Fein, Fabrice Vavre, Laurence Mouton, Andrés Moya, Amparo Latorre, and Francisco J. Silva.

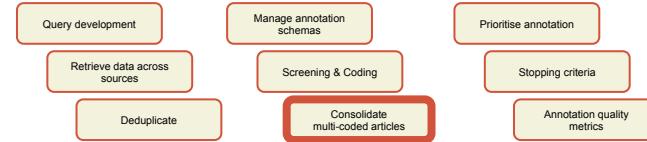


The screenshot shows the NAC-SOS annotation interface with the 'Annotation Panel' open. The panel contains several search filters with toggle switches:

- Show scheme description
- Is about cities (checked)
- Mentions one or more cities (checked)
- Topic: Adaptation (checked)
- Topic: Mitigation (checked)
- Topic: Impacts (checked)

At the bottom of the panel are 'Save & Previous' and 'Save & Next' buttons. The background shows a blurred view of the NAC-SOS interface, including a header with 'Projects', 'About', and a user profile, and a sidebar with 'Annotation' selected.





Label consolidation

- Resolve annotations by multiple users
- Platform proposes a majority vote resolution
- Annotators can resolve disagreements
- Since original annotations are kept, we can analyse how challenging the coding task was and estimate the quality of annotations

Label cons

- Resolve annotations
- Platform projects
- Annotators can disagree
- Since original data is challenging to work with, annotations are often inconsistent

Resolve Annotations

Annotation Export Configuration

Descriptive name for this annotation export
Resolve_20230911_01_prioritised_LS_FH

Annotation Scheme
Carbon pricing map

Source assignment scopes

- 20230901_03_prioritised_LS_FH
- 20230927_prioritised_LS_FH
- 20230911_02_prioritised_LS_FH
- 20230503_relevant_AL_AA_KM_ND
- 20230505_relevant_AL_AA_KM
- 20230505_relevant_AL_AA
- 20230515_prioritised_AL_ND
- 20230522_prioritised_AL_AA_ND
- 20230901_01_prioritised_ND_LS_FH_training
- 20230901_02_prioritised_ND_LS_FH_learning
- 20230911_01_prioritised_LS_FH

Scheme labels to resolve

- meth – Method
- outc – Analysed outcome
- polname – Policy name
- sect – Sector
- otherpol – Interaction with other policies
- exp – ex-post/ex-ante
- imp – Implemented policy
- cp – Carbon pricing

Repeats to resolve ?

- 1
- 2
- 3
- 4

Ignore scheme hierarchy

Ignore annotation order (repeats)

Source assignment scopes

Annotation selector

- leonhard.schneider – Leonhard Schneider
- niklas.doebbeling – Niklas Döbbeling
- alessandra.landa – Alessandra Landa
- klaas.miersch – Klaas Miersch
- finlay.hatch – Finlay Hatch
- arianna.avallone – Arianna Avallone

Resolution algorithm

- majority vote

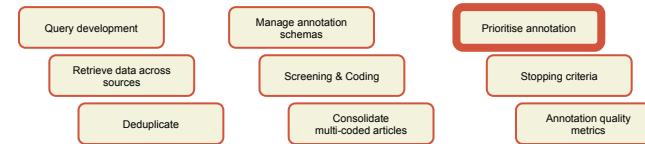
Show text

Load

Item # Filter item_id

#	Item	cp	imp	exp	meth	otherpol	outc	polname	sect
0	02340393-fc53-402d-aaa8-eba37b8f82be9	1 1	1 1	1 1	1 1	2 5	2 5	10 9	5 1
1	063b3764-bc81-4447-86f3-b5c360fd1236d	1 1	1 1	1 1	1 1	0 0	0 0	2 6	4 1
2	08e4bc1e-c327-4c2d-90e-79458b2b631e	1 1	1 1	1 1	1 1	3 3	3 3	1 10	5 1
3	0958b1d1-13d4-414c-86a8-db32e3cb5573	1 1	1 1	2 1	1 1	1 1	1 1	10 5	5 1
4	091cd2e5-766e-4988	1 1	1 1	1 1	1 1	1 1	1 1	10 5	5 1

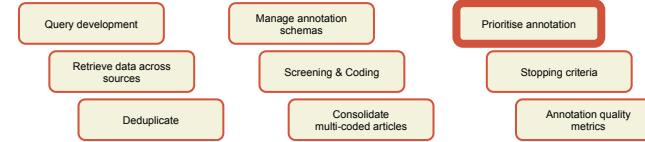
Save



Managing assignments

Fine-grained control over who has access to what in your project

The screenshot shows the NACoS (Niklas' Annotation Configuration System) interface. The left sidebar has a 'Project' tab selected, with other tabs like 'Overview', 'Import', 'Dataset', 'Annotation', 'Artefacts', 'Pipelines', and 'Annotations' visible. The main area shows 'Project information' with a project ID (748e739d-f011-44de-9cb0-c9cb4bb18d08), a 'Project name' (Carbon pricing map (dedup)), a 'Project type' (Academic (abstracts)), and a 'Project description' (empty). A 'Save' button is present. Below this is a section for 'Motivational quotes during annotation' with a toggle switch. The 'User permissions' section shows two users: 'niklas.doebbeling' and 'alessandra.landa'. For 'niklas.doebbeling', permissions are listed: owner, imports_edit, pipelines_edit, annotations_read, annotations_edit, artefacts_read, dataset_read, dataset_edit, and imports_read. For 'alessandra.landa', only 'dataset_read' is listed. Buttons for 'Apply preset', 'Remove', and 'Save' are at the bottom of the permissions table. A '+' button is at the bottom right of the user list.



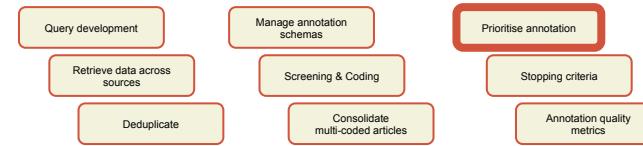
Managing assignments

Assign a user to annotate a specific document using an annotation scheme (in batches/scopes).

- Random / rule-based sampling
 - Determine how many documents to code by how many users
 - Exclude/include specific documents per batch of assignments

Next week (or two):

- Prioritised screening with machine learning
- First platform to implement stopping criteria!!



Managing assignments

Assign a user to annotate a specific *document* using an annotation *scheme* (in *batches/scopes*).

Assignment strategy settings

Random assignment

Configure random assignments

Number of items # multi-coded items Min. # coders per item Max. # coders per item Random seed

150 150 2 2 10829

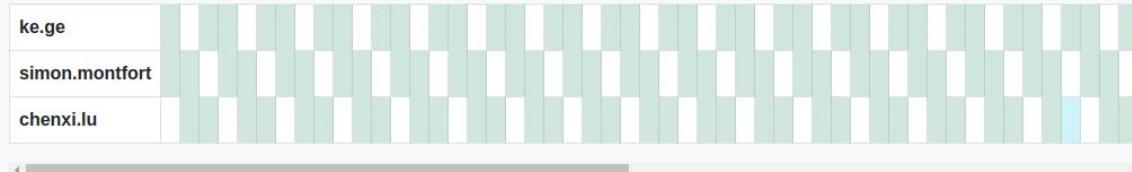
Make assignments

Results

(Re)load stats

Assignments: 400 (open: 4 | partial: 0 | done: 396)

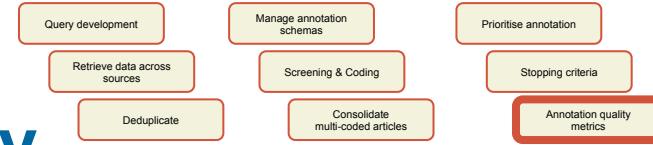
ke.ge
simon.montfort
chenxi.lu



Next week

- Prioritise annotation
- First platform to implement stopping criteria!!

any users assignments



Monitor annotation quality

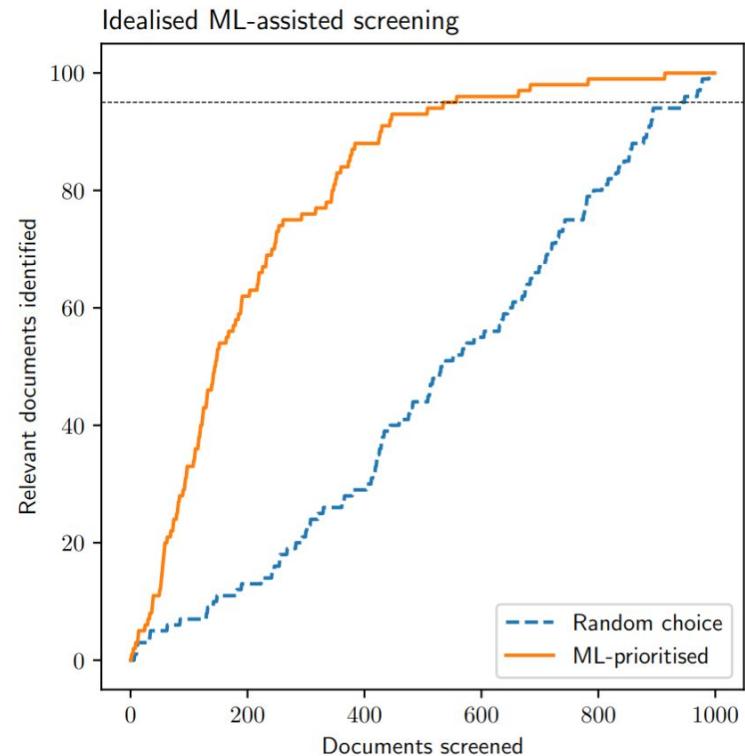
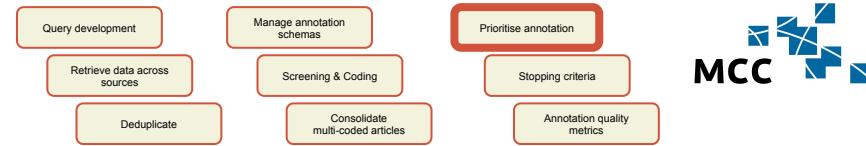
- Annotator leaderboard
- Track assignment progress

Coming soon...

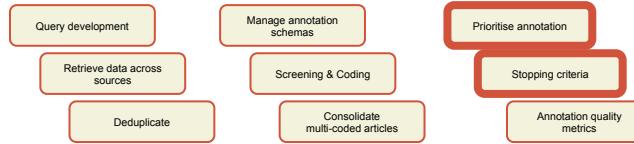
- Screening/coding coverage of dataset
- Statistics on inter-rater agreement
- Stopping criterion plots
- Recall targets

ML-assisted document screening

- Growing number of “researcher-in-the-loop” machine learning applications for screening documents for systematic reviews (O’Mara-Eves et al., 2015; van de Schoot et al., 2021).
- Using machine learning to prioritise documents likely to be relevant, we can achieve high levels of recall without screening all documents.

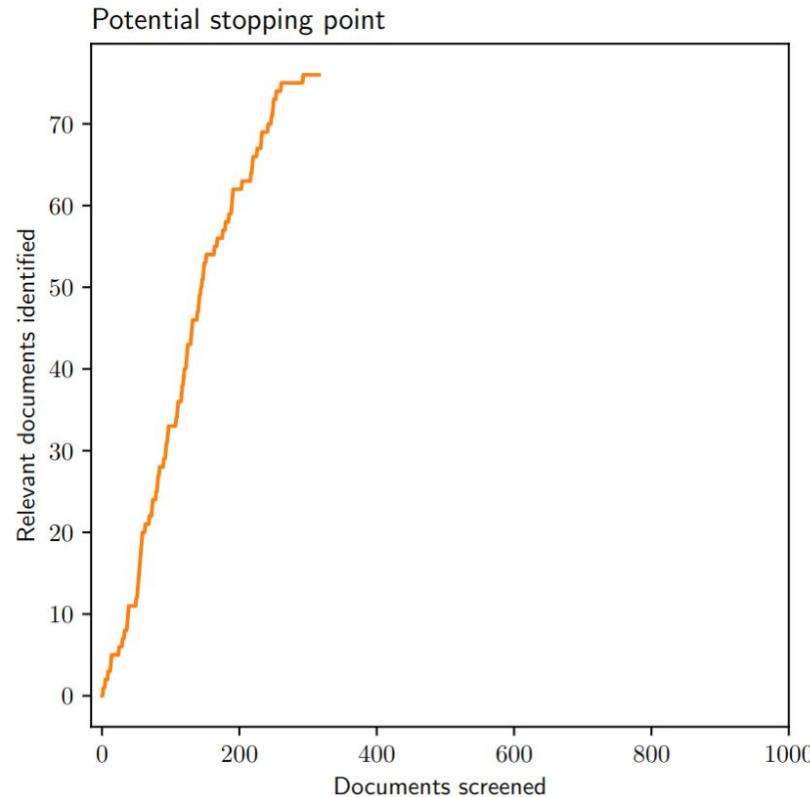


Unleashing AI; But when do we stop?



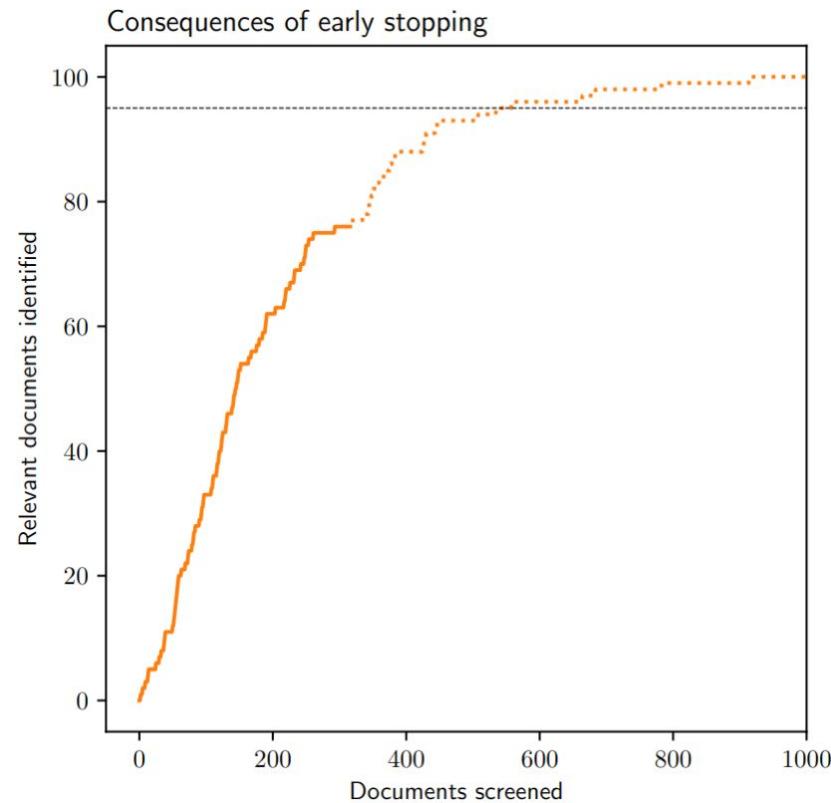
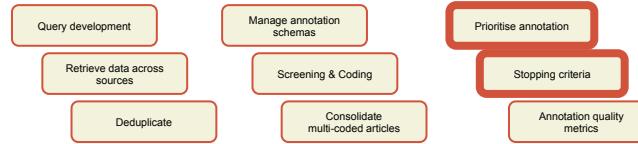
- We do not know *a priori* the true number of relevant documents

→ We need criteria when to stop screening!

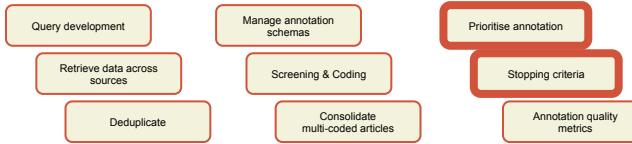


Unleashing AI; But when do we stop?

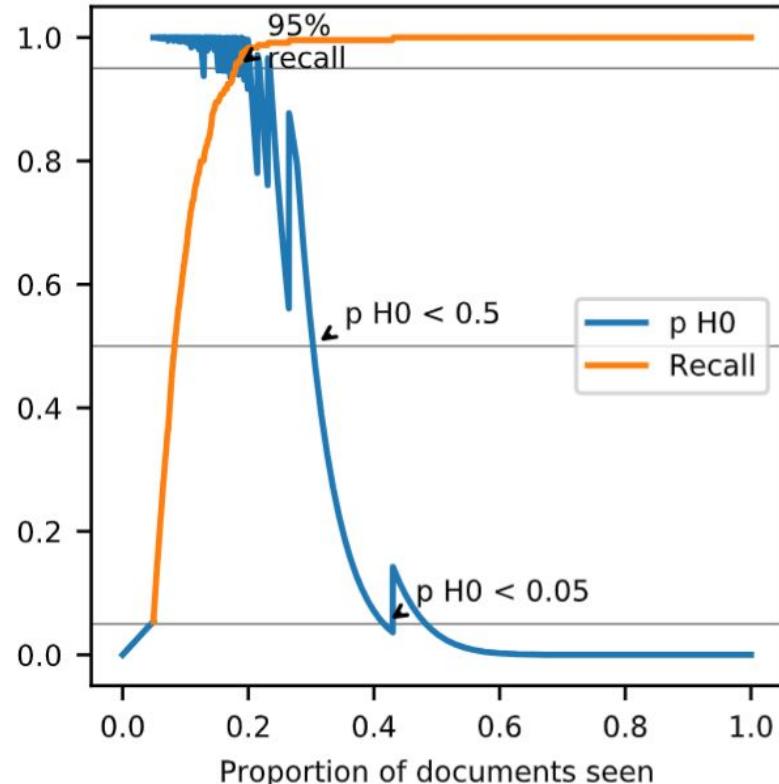
Stopping too early can
lead to huge biases in
reviews!



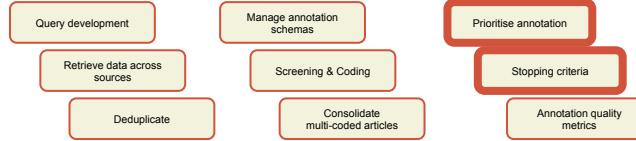
Unleashing AI; But when do we stop?



- Our stopping criterion works by treating documents as if they were white (not relevant) and red (relevant) marbles drawn from an urn without replacement.
- The hypergeometric distribution describes the probability of observing k red marbles in a sample of n marbles, given an urn with N marbles, of which K are red.
- We formulate a null hypothesis H_0 that a given recall target (e.g. 95% of relevant documents) has been missed.
- We calculate a p-score for H_0 and stop screening if this falls below a selected threshold.



Unleashing AI; But when do we stop?



Also available as an R package
<https://mcallaghan.github.io/buscarR/>



Is the first (and only) platform
with a stopping criterion!



Beyond the platform

NAC SOS

- Overview
- Import
- Dataset**
- Explore
- Statistics
- Export**
- Annotation
- Artefacts
- Pipelines
- Annotations
- Project

Download/Export Data

Options

Ignore annotation hierarchy
 Ignore annotation order

Users

chenxi.lu
 felix.creutzig
 jan.minx
 ke.ge
 max.callaghan
 simon.montfort
 tim.repke

Document fields

text
 title
 doi
 wos_id
 scopus_id
 openalex_id
 publication_year
 source

Resolved annotations / BotAnnotations

Settlements annotations

Assignment Scopes

WG2, Chapter 6
 City related

CCP, Chapter 2
 City related

WG3, Chapter 8
 City related

Batch 1
 Cities and topics

settlements_01_CL_KG_SM
 Settlements

settlements_02_CL_KG_SM

Labels

adaptation all none
 false true

cities all none
 0 1 2

city all none
 false true

climate all none
 false true

com Include strings

human_settlements all none
 false true

Projects About  tim.repke ▾

Beyond the platform (soon)

Living evidence maps

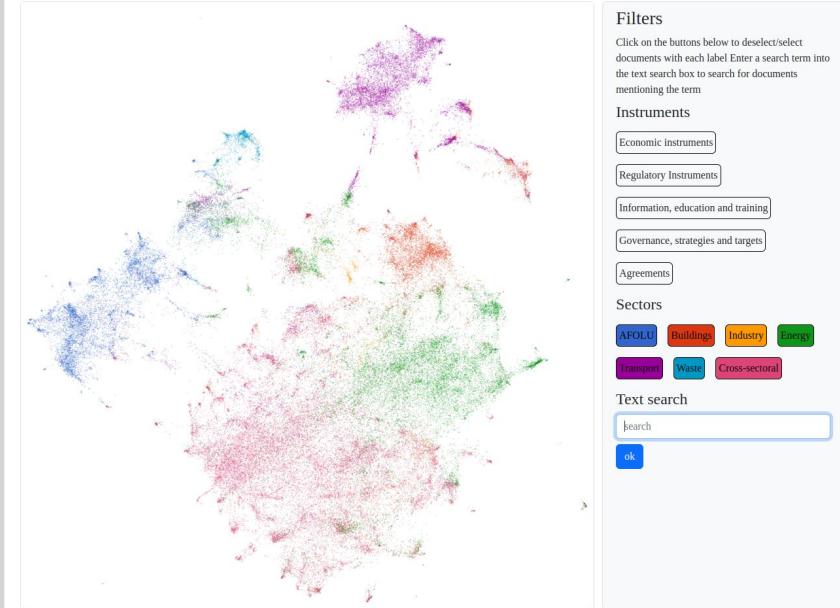
- Regularly update queries
- Classify new data
- Feed filtered and enriched data into map explorer
- Provide open, up-to-date datasets for relevant research areas

Climate Policy Instruments

Max Callaghan

This interactive website accompanies the paper [x], which uses machine learning to identify and classify the literature on climate policy instruments. You can explore this literature in the map below, where each paper is represented by a dot, and papers which are linguistically similar are placed close together on the plot. Hovering over the map will show the titles of the papers.

You can select papers by clicking and dragging on the map to zoom in on an area. Or you can choose a different type of selection method using the icons in the top left. Once you have selected documents, a sample of these will be shown in the box below. You will also have the opportunity to download the complete selection of documents, including the machine-learning generated labels.



<https://apsis.mcc-berlin.net/climate-policy-instruments-map/>

Disclaimer / Terms of Service

- This is **not a** (commercial) **product** and we are not service providers.
 - We cannot offer 24/7 tech support.
 - We cannot offer to implement custom features.

but...

- We are open to invite others to use the platform.
 - Our “**fee**” is a **co-authorship** in respective publications.
 - If it is in line with our needs, we can implement features.
 - If resources permit, we can run additional analyses.
 - We can support the onboarding and introduction to the platform.
 - You could set up your own instance, it’s open source and “well documented”.
- **Data ownership**
 - We do **not guarantee** the **data is safe** forever. We do make backups and do our best though.
 - We would also like to use the data, but of course, you will be included in the process and we will never publish anything novel with your data.

Links

APSiS Website

<https://apsis.mcc-berlin.net/>

NACSOS

<https://apsis.mcc-berlin.net/nacsos/>

NACSOS Documentation

<https://apsis.mcc-berlin.net/nacsos-docs/>

Source code

<https://gitlab.pik-potsdam.de/mcc-apsis/nacsos>





Tim Repke
repke@mcc-berlin.net

Interested?
Get in touch!

Max Callaghan
callaghan@mcc-berlin.net

Other tools

- AS-Review <https://asreview.nl/>
- EPPI Reviewer <https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=2914>
- Covidence <https://www.covidence.org/>
- 3ie Evidence Gap Explorer <https://developmentevidence.3ieimpact.org/>
- ...