MCC

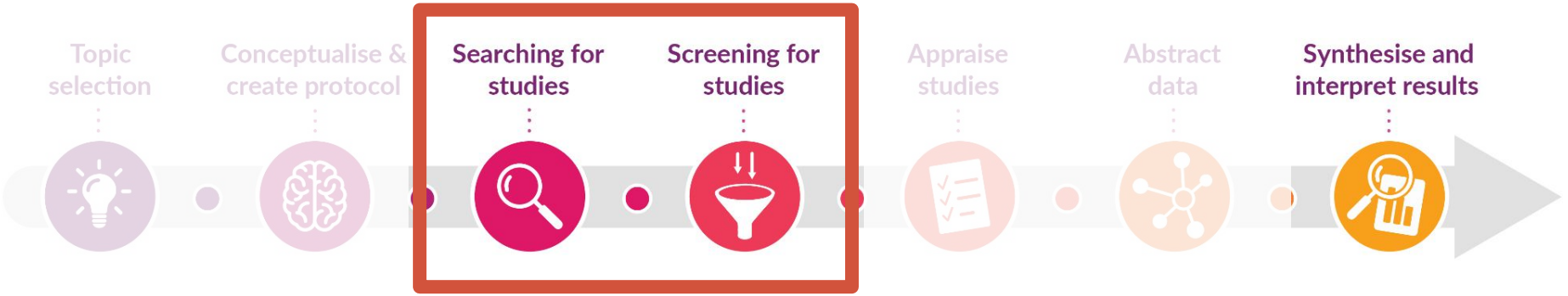
Mercator Research Institute on
Global Commons and Climate Change gGmbH

Computer-aided research synthesis

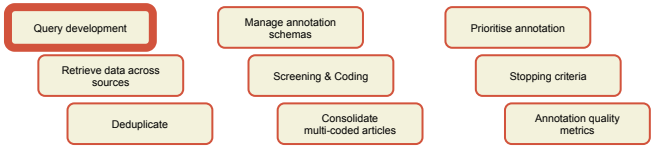
Overview of the *NACSOS* screening platform developed in the APSIS group

Max Callaghan
Tim Repke

Steps in a systematic review



- Query development
- Retrieve data across sources
- Deduplicate
- Manage annotation schemas
- Screening & Coding
- Consolidate multi-coded articles
- Prioritise annotation
- Stopping criteria
- Annotation quality metrics

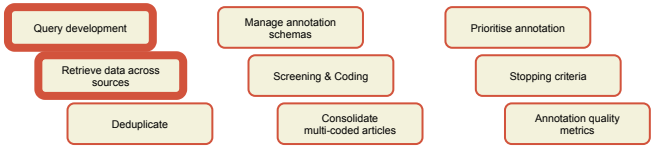


MCC

Query development

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



MCC

Coverage of OpenAlex

OpenAlex (~244M)

Climate (~1M)

Cities (~48k)

Mentions (~22k)

~900

~300

~2,000

3,220

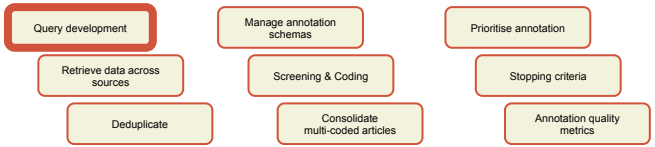
3,721

71,863

8,908

IPCC AR6 (~81k)

WG2-Ch6 (~2k)
CCP-Ch2 (~900)
WG3-Ch8 (~900)



Query development

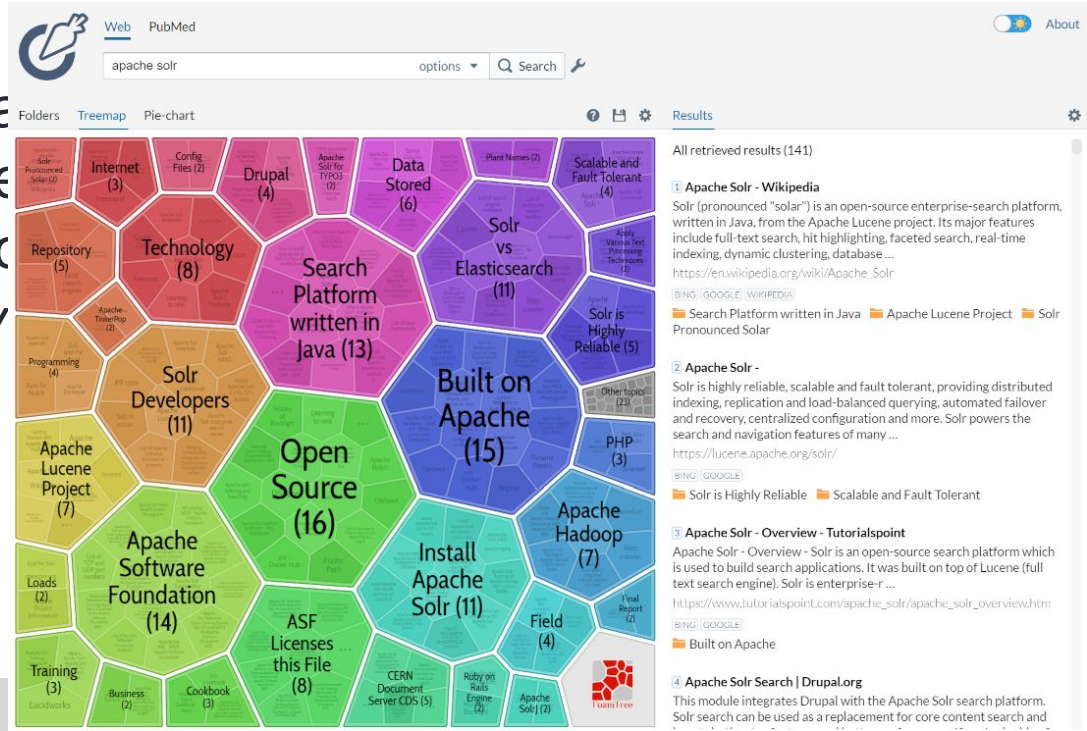
- Query OpenAlex (our solr mirror) directly from the platform

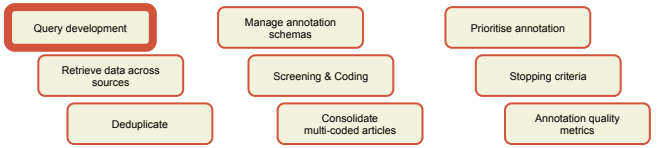
- Compare to “gold standard” references *(soon, probably)*
- Compare different queries *(soon, possibly)*
- Keyword recommendations *(let’s see)*
- Similar documents *(maybe)*
- Follow citations *(likely)*

Query development

- Query OpenAlex (our solr mirror) directly from the platform

- Compare to “gold standa
- Compare different querie
- Keyword recommendatic
- Similar documents *(may*
- Follow citations *(likely)*
- Instant topic maps *(??)*





Query development

Projects About tim.repke

Overview
Import
Imports
OpenAlex Solr
Dataset
Annotation
Artefacts
Pipelines
Annotations
Project

Search OpenAlex (Solr)

Query setup

"climate change" and "CDR"

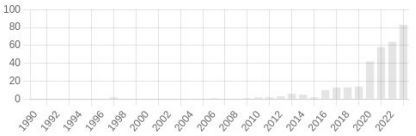
Tokens Help

Limit: 20
Offset: 0 (+20 -20)
OP: AND
Field: title_abstract
Parser: edismax

Histogram
1990 – 2024

Query

Found 321 works in 1123ms



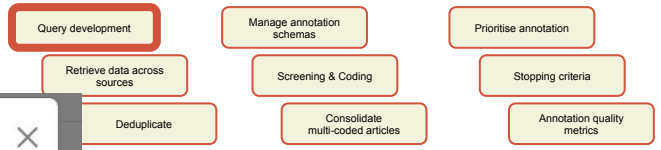
Is carbon dioxide removal 'mitigation of climate change'?

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



MCC

Projects About tim.repke

Tokens Help

100 80 60 40 20 0

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

...n of CDR governance – despite enjoying growing interest – tends to overlook how key provisions on ...vely 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.

... most CDR strategies are unfamiliar to the public, it is unknown how increased media and policy ...tial moral hazard: if people perceive that CDR solves climate change, they may be less likely to support ...ty of climate change and, thus, increase support for other types of mitigation. Using an online survey of ...ifferent forms of CDR. We find that learning about certain CDR strategies indirectly reduces support for ...ants who read about CDR in general (without mention of specific strategies), bioenergy with carbon

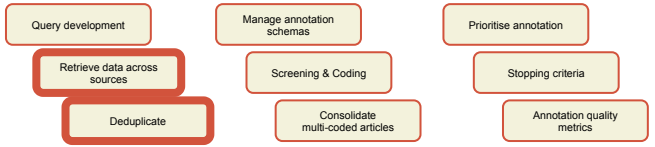
Token Wildcard Expansion

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).

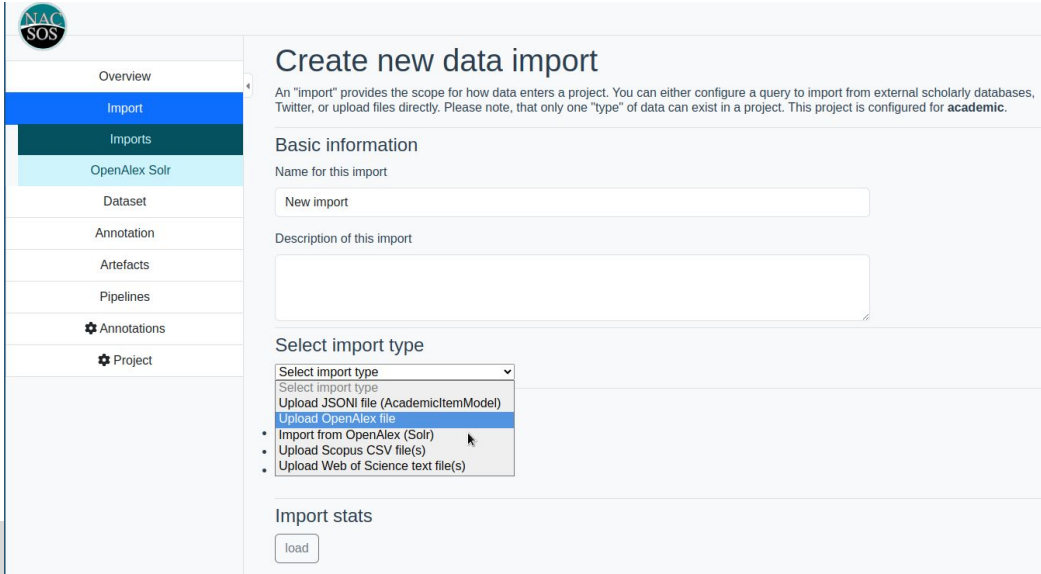
trad

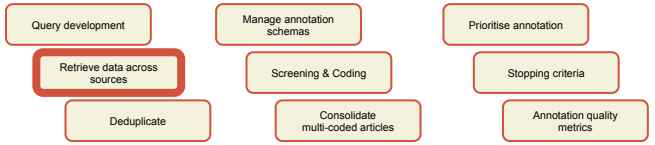
traditional (4,367,640 | 3,336,034)	trade (2,283,075 | 1,258,966)	tradition (741,911 | 571,371)
traditionally (469,056 | 452,426)	traditions (344,721 | 277,624)	trading (473,445 | 267,987)
tradicional (127,853 | 106,850)	tradeoff (104,027 | 84,757)	traded (91,789 | 68,194)
traders (102,810 | 61,627)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

Data gathering (importing queries)



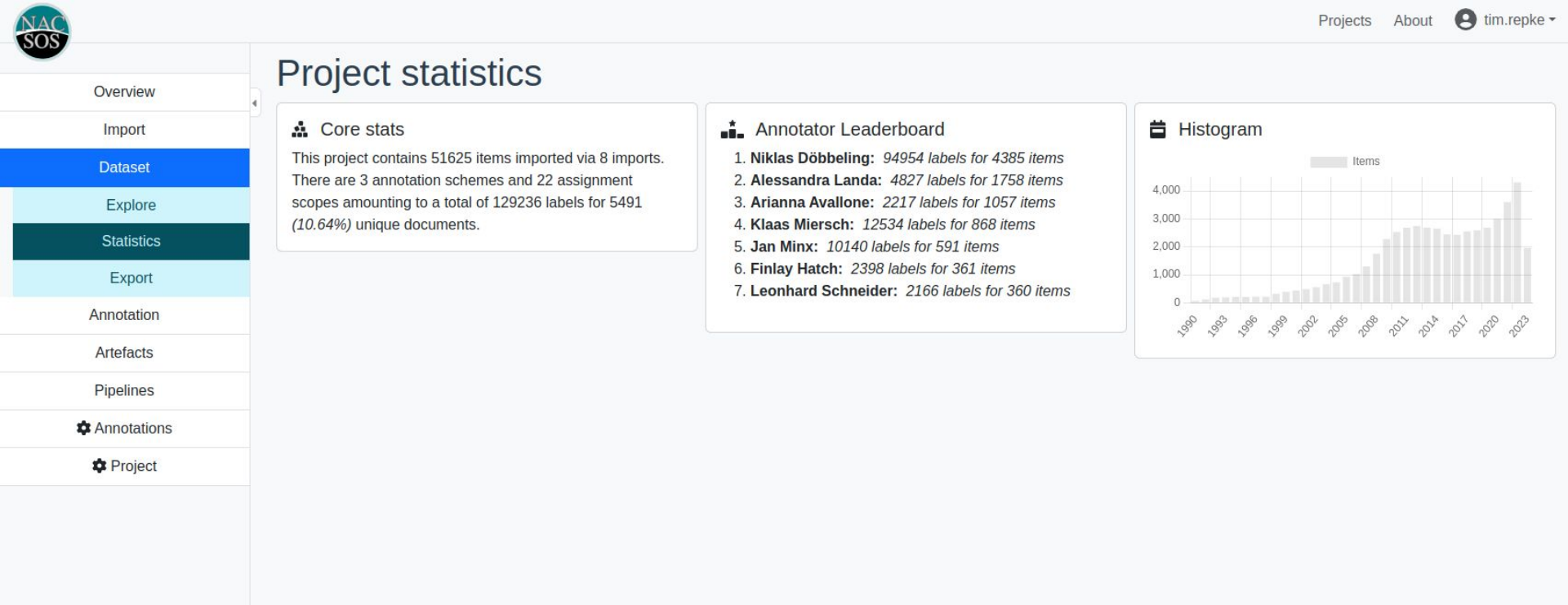
- Import data from
  - Scopus (via CSV)
  - Web of Science (via RIS)
  - OpenAlex (via solr)
  - Any other source (you just need to translate it to the platform format)
- Automatic deduplication of publications during import

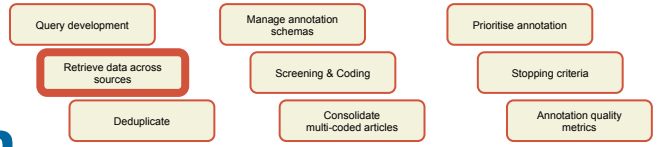




Dataset insights

*more to come!



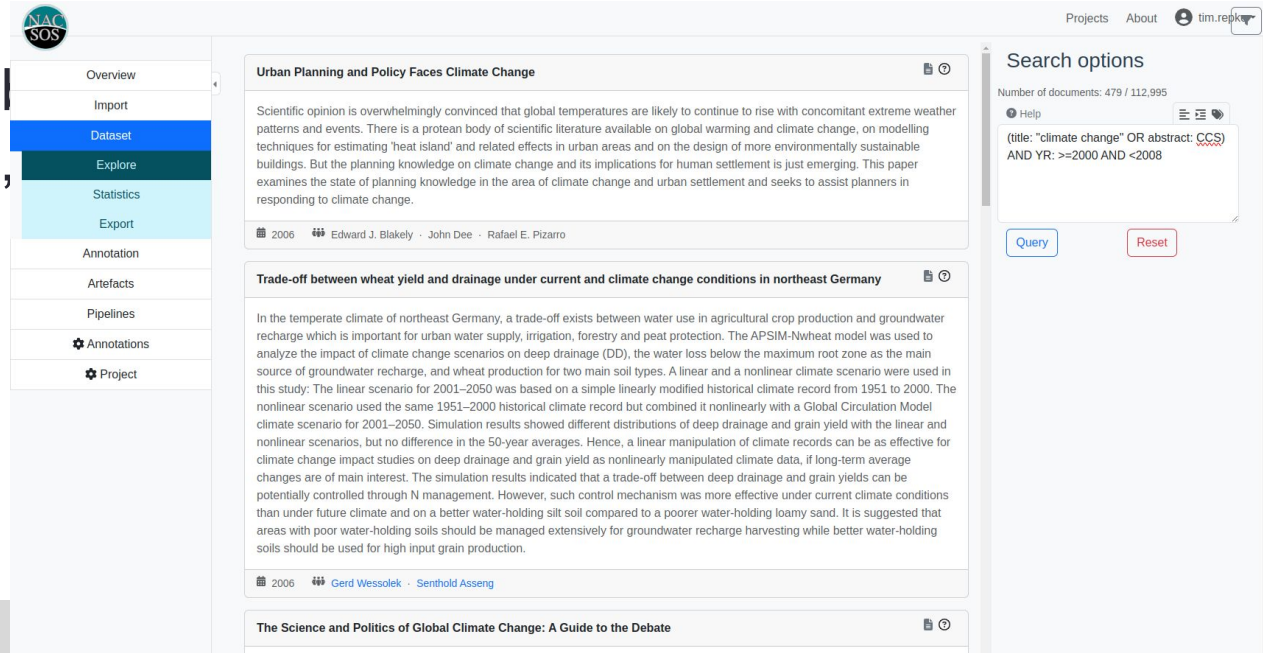


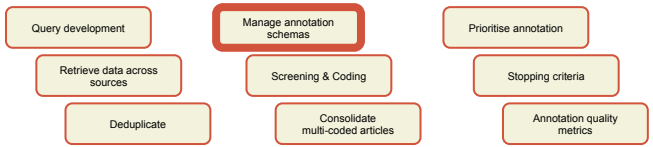
MCC

Searching on the platform

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

- Title, abstract, pul
  year, annotations,

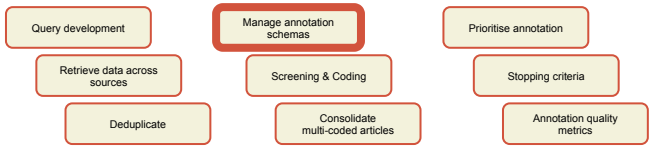




Annotation schema

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

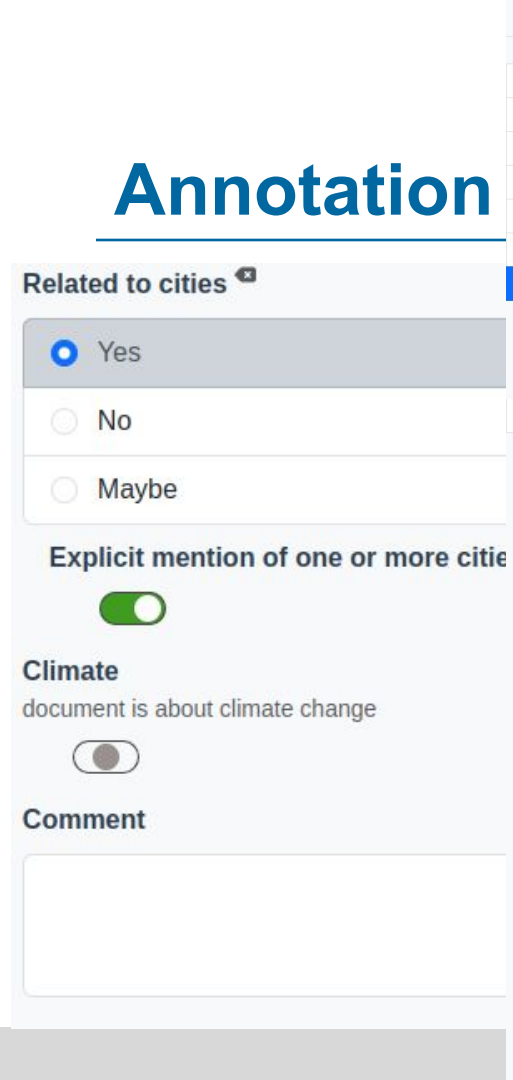
Comment

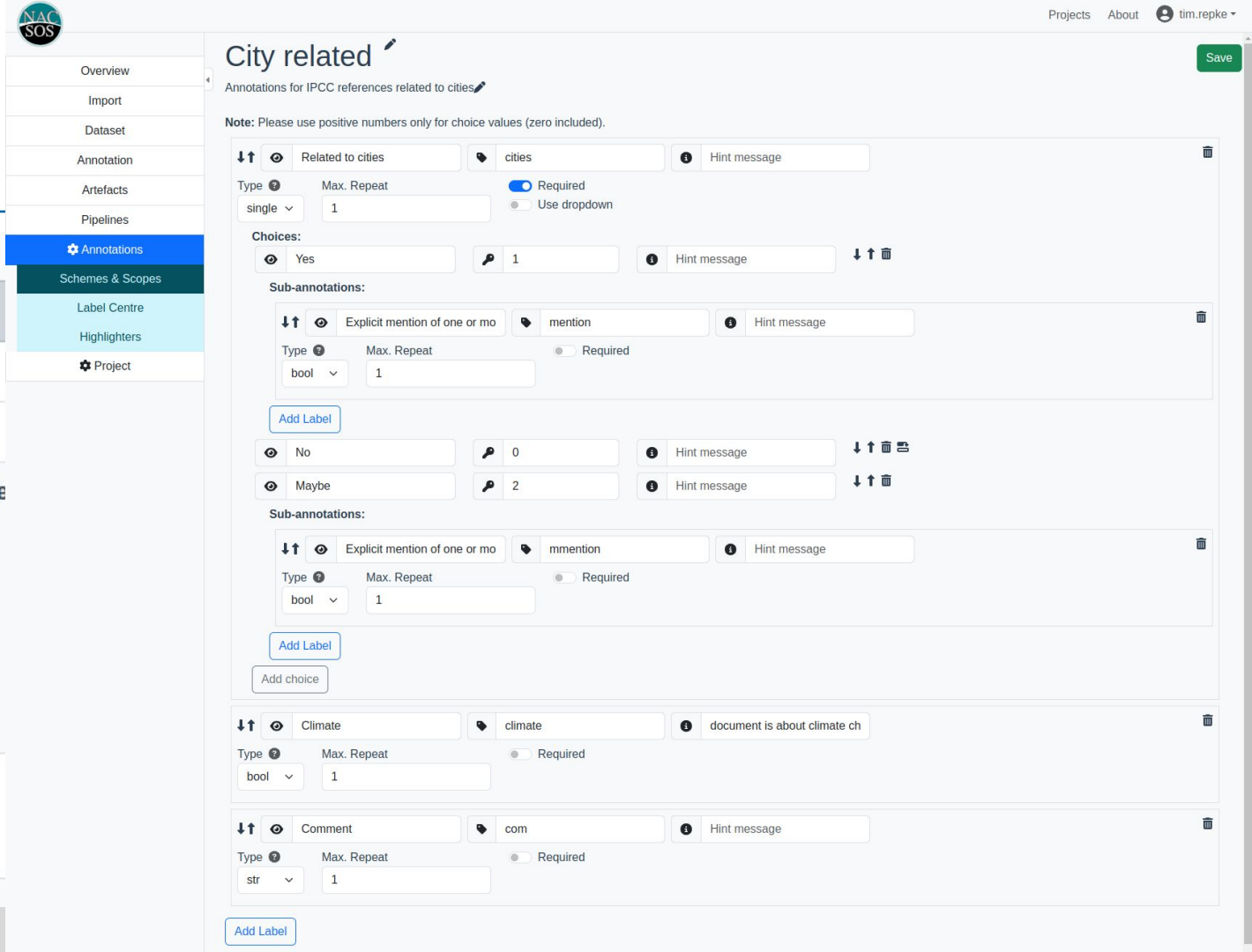
n screening to coding

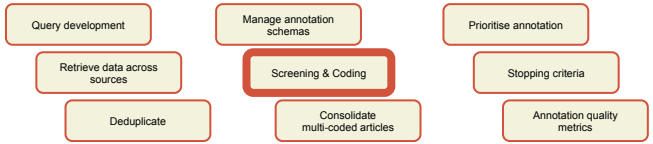
le classes)

al labels”

Annotation

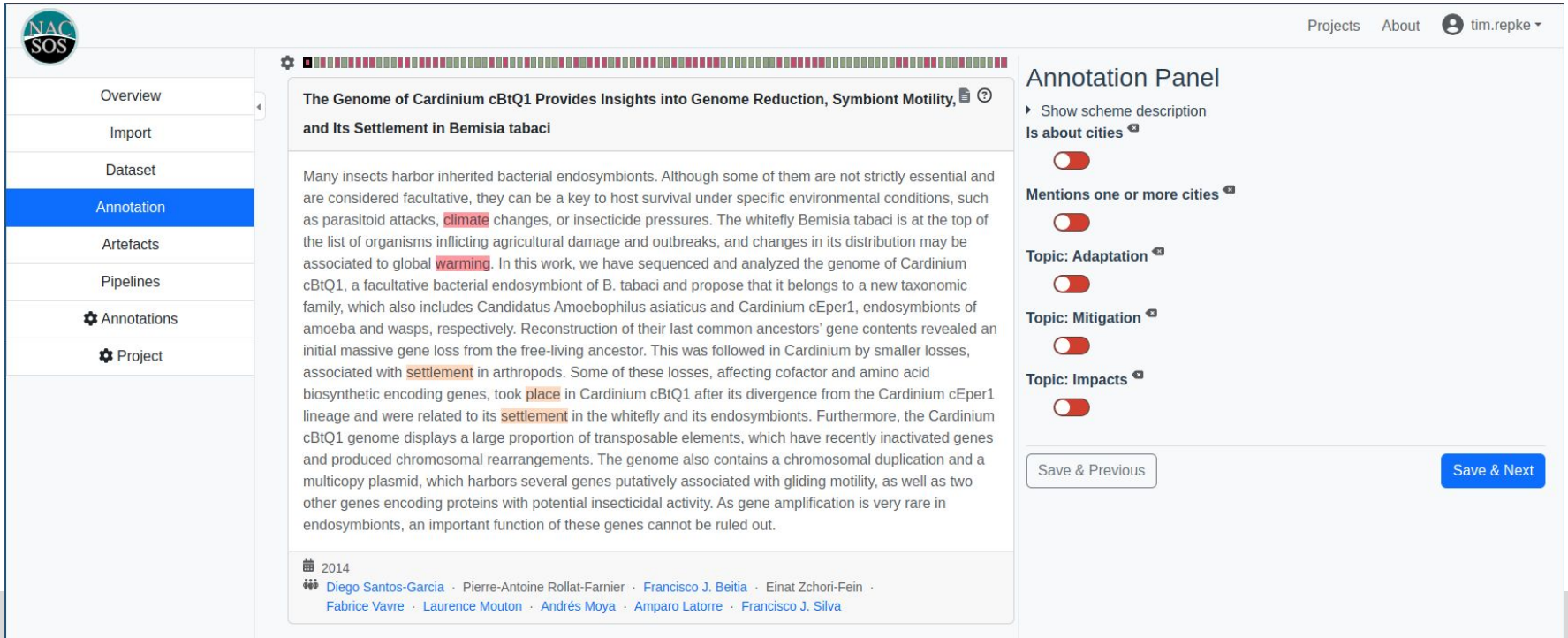

Related to cities
Yes
No
Maybe
Explicit mention of one or more citie
Climate
document is about climate change
Comment

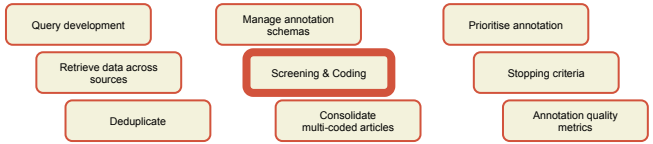

Projects
About
tim.repke
Overview
Import
Dataset
Annotation
Artefacts
Pipelines
Annotations
Schemes & Scopes
Label Centre
Highlighters
Project
City related
Save
Annotations for IPCC references related to cities
Note: Please use positive numbers only for choice values (zero included).
Related to cities
cities
Hint message
Type
Max. Repeat
Required
Use dropdown
single
1
Choices:
Yes
1
Hint message
Sub-annotations:
Explicit mention of one or mo
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 mo
mmention
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
Comment
com
Hint message
Type
Max. Repeat
Required
str
1
Add Label



Annotating

Intuitive annotation interface for users





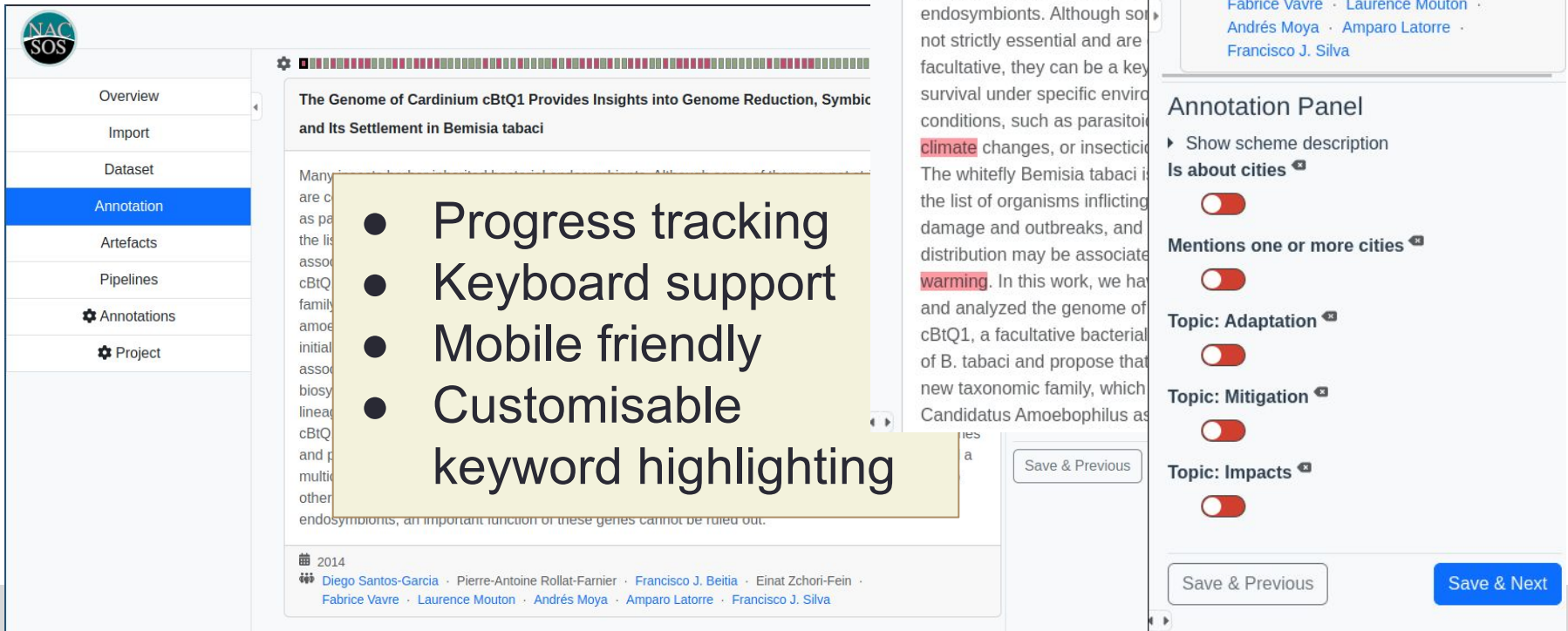
Annotating

Intuitive annotation interface for users

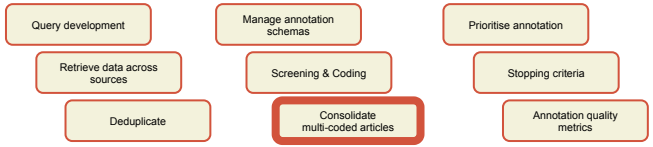
- Progress tracking
- Keyboard support
- Mobile friendly
- Customisable keyword highlighting

Annotating

Intuitive annotation interface for users



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

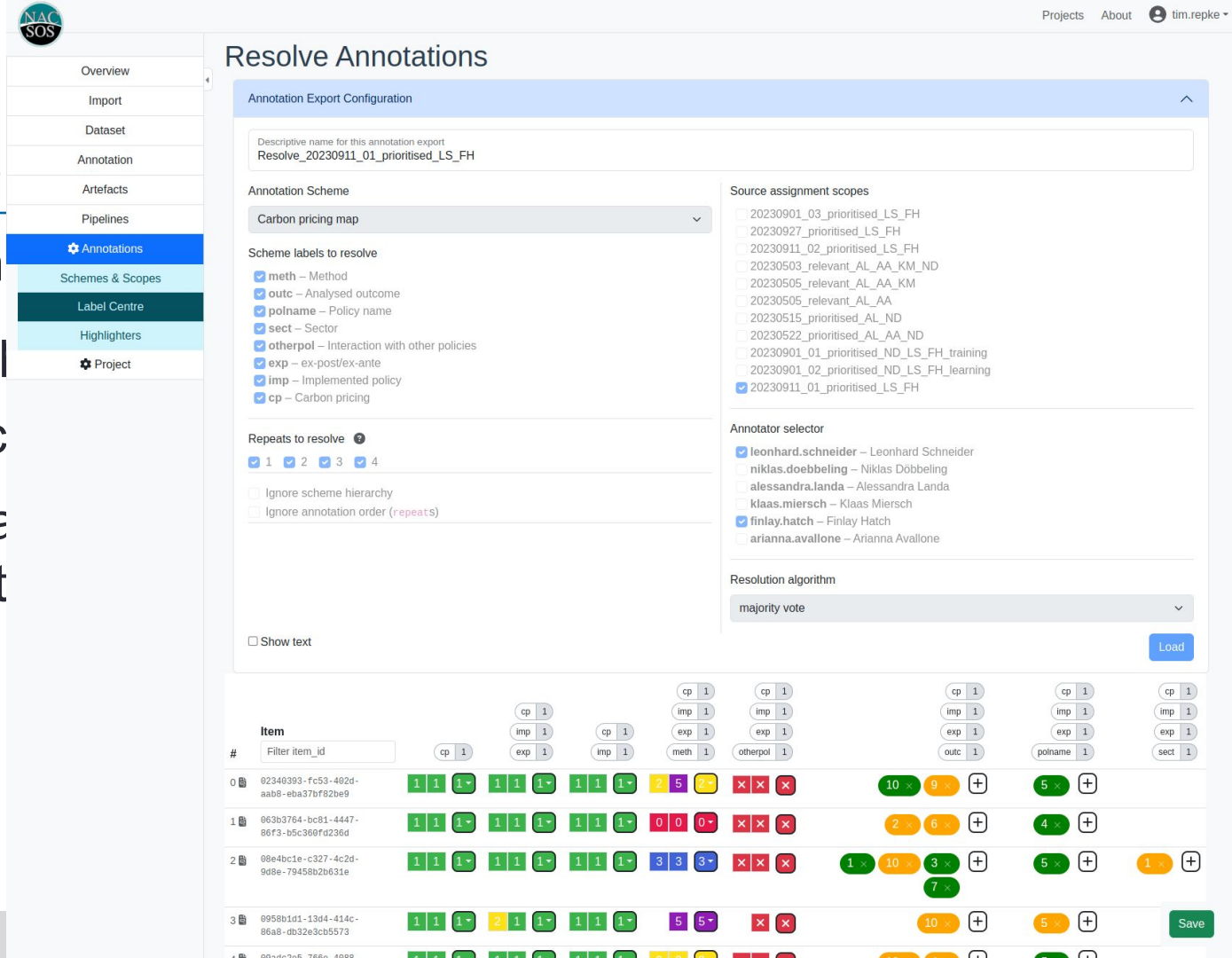


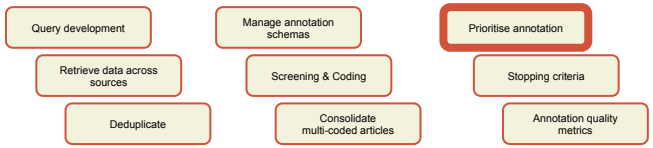
Label consolidation

- 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 conso

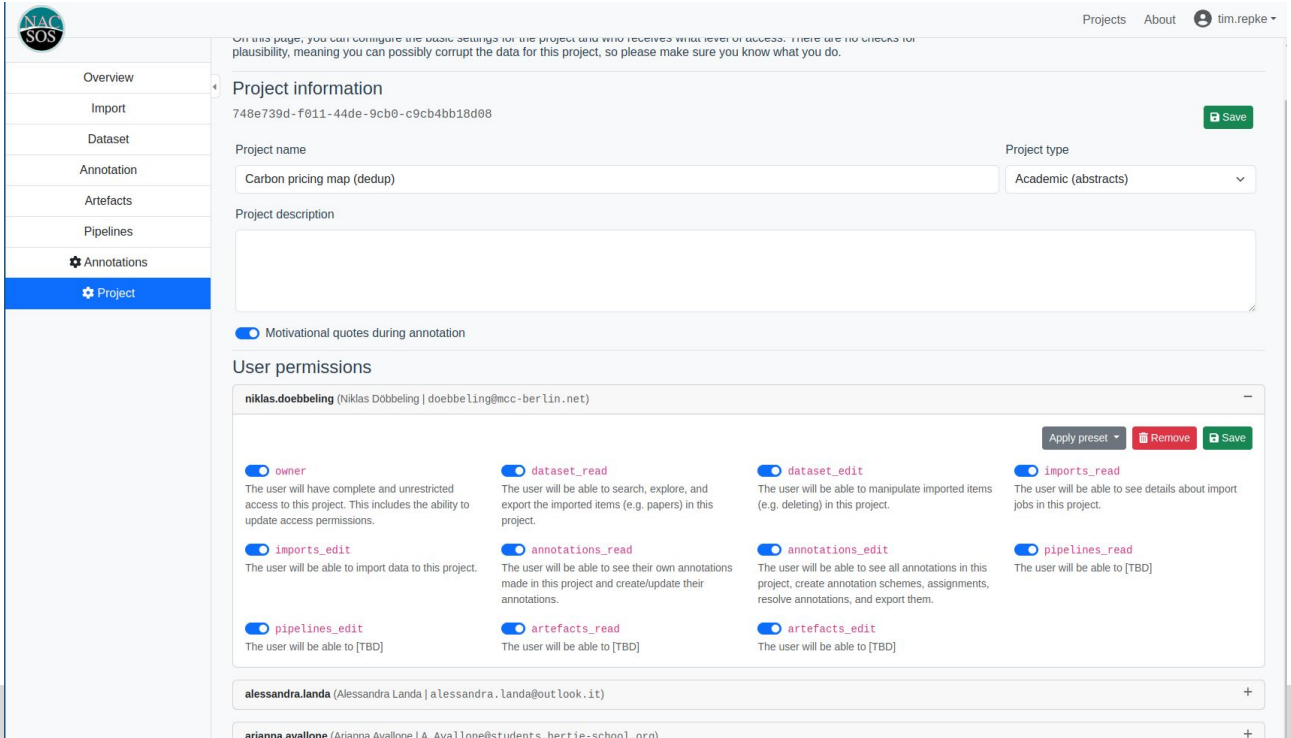
- Resolve ann
- Platform prop
- Annotators c
- Since origina
  challenging t
  annotations

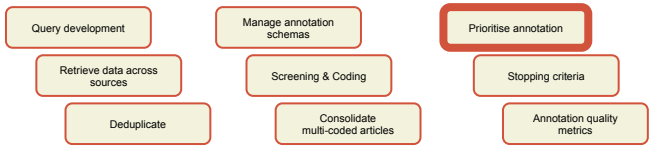




Managing assignments

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





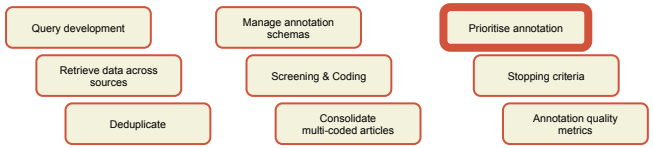
Managing assignments

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).

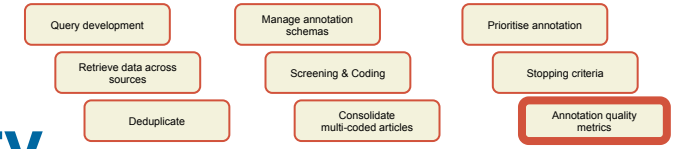
- Ra
  -
  -

any users

signments

Next w

- Pri
- First platform to implement stopping criteria!!



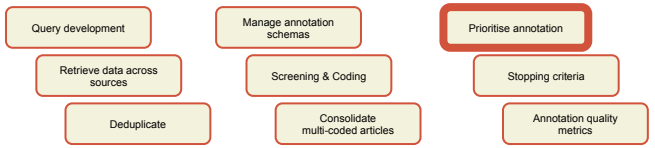
Monitor annotation quality

- Annotator leaderboard
- Track assignment progress

Coming soon…

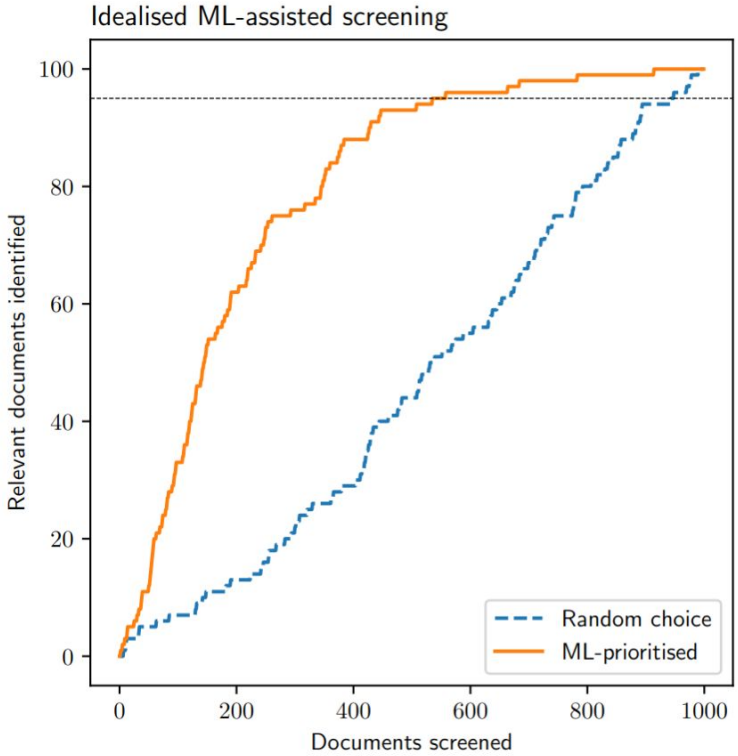
- Screening/coding coverage of dataset
- Statistics on inter-rater agreement
- Stopping criterion plots
- Recall targets

ML-assisted document screening

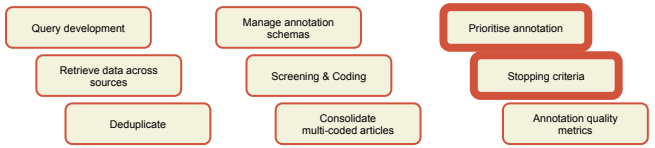


MCC

- 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.



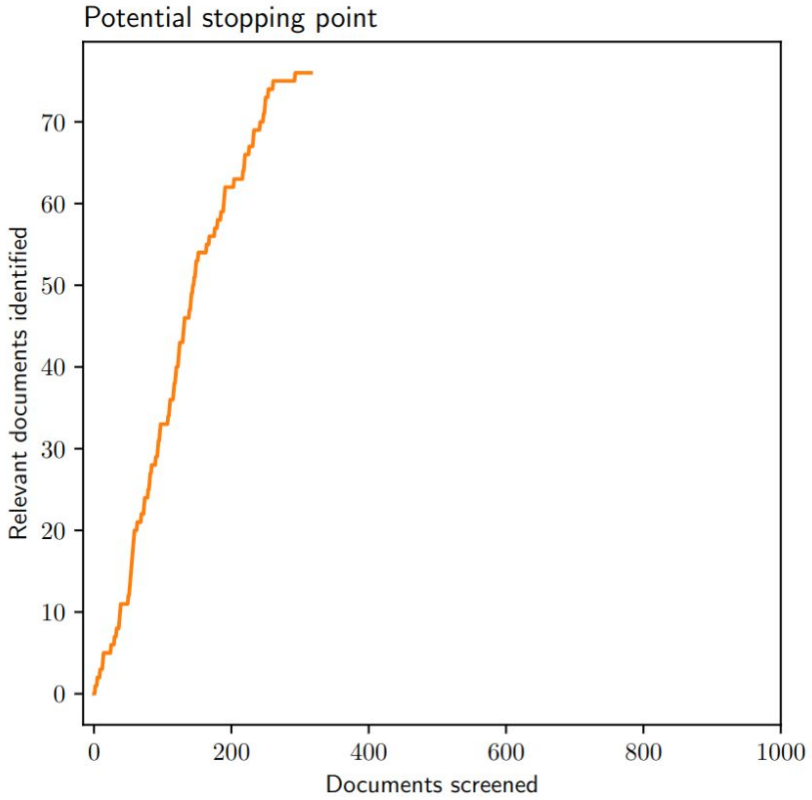
Unleashing AI;
But when do we stop?



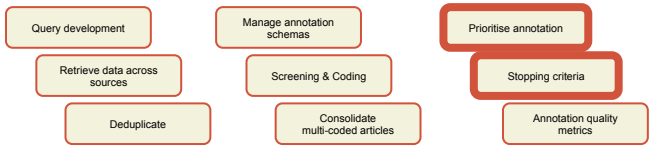
MCC

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

→We need criteria when to stop screening!

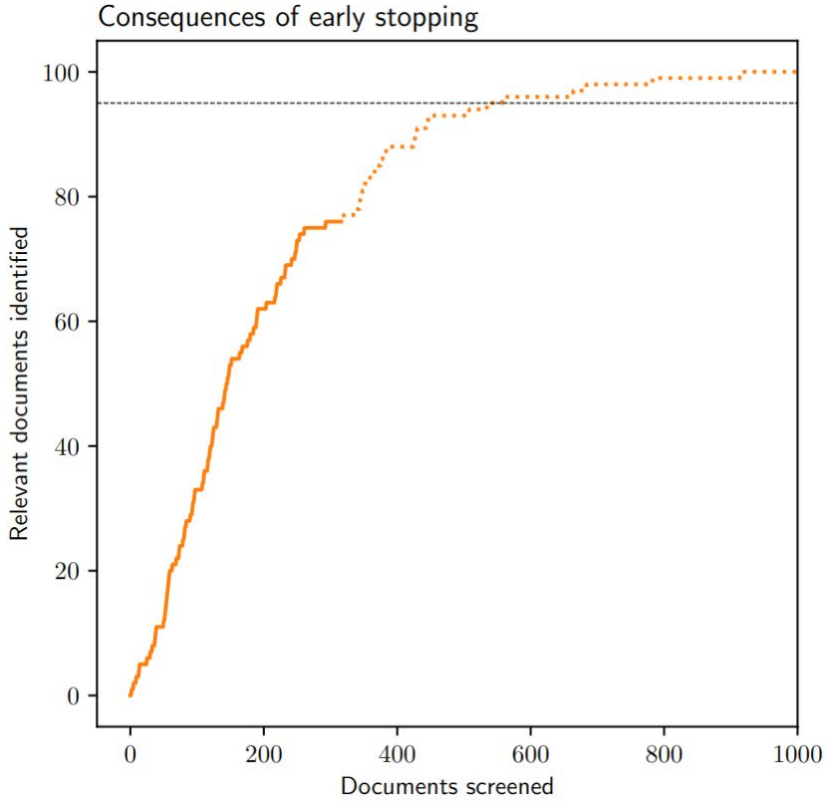


Unleashing AI;
But when do we stop?

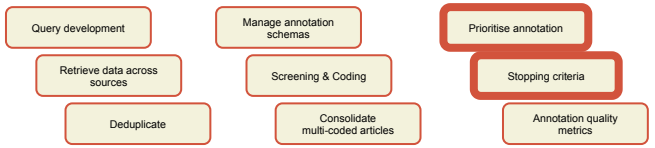


MCC

Stopping too early can lead to huge biases in reviews!

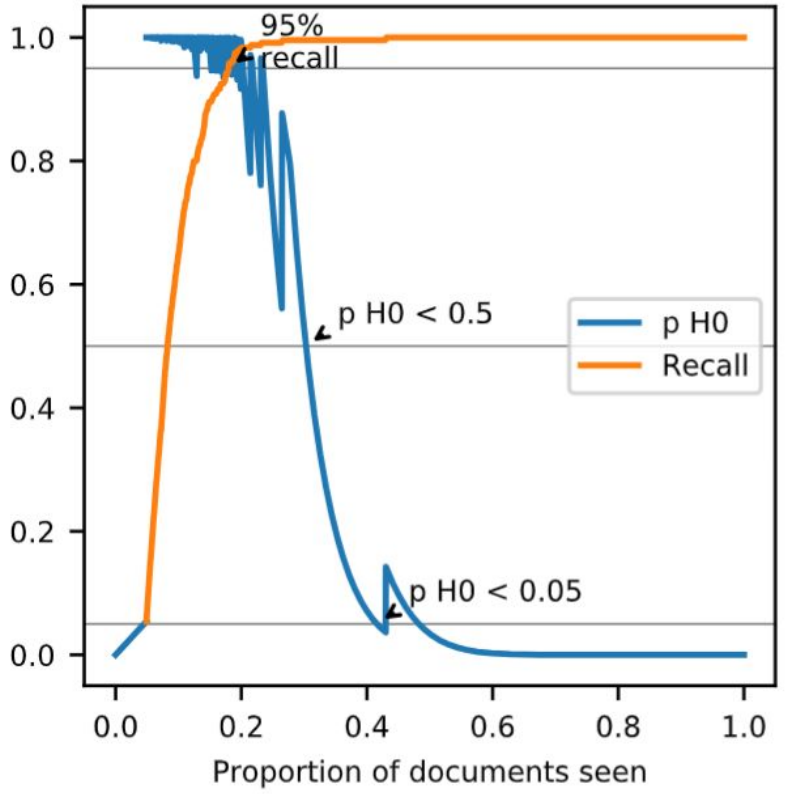


Unleashing AI;
But when do we stop?

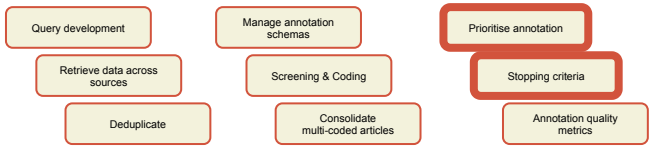


MCC

- 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 ***H0*** that a given recall target (e.g. 95% of relevant documents) has been missed.
- We calculate a p-score for ***H0*** and stop screening if this falls below a selected threshold.



Unleashing AI;
But when do we stop?



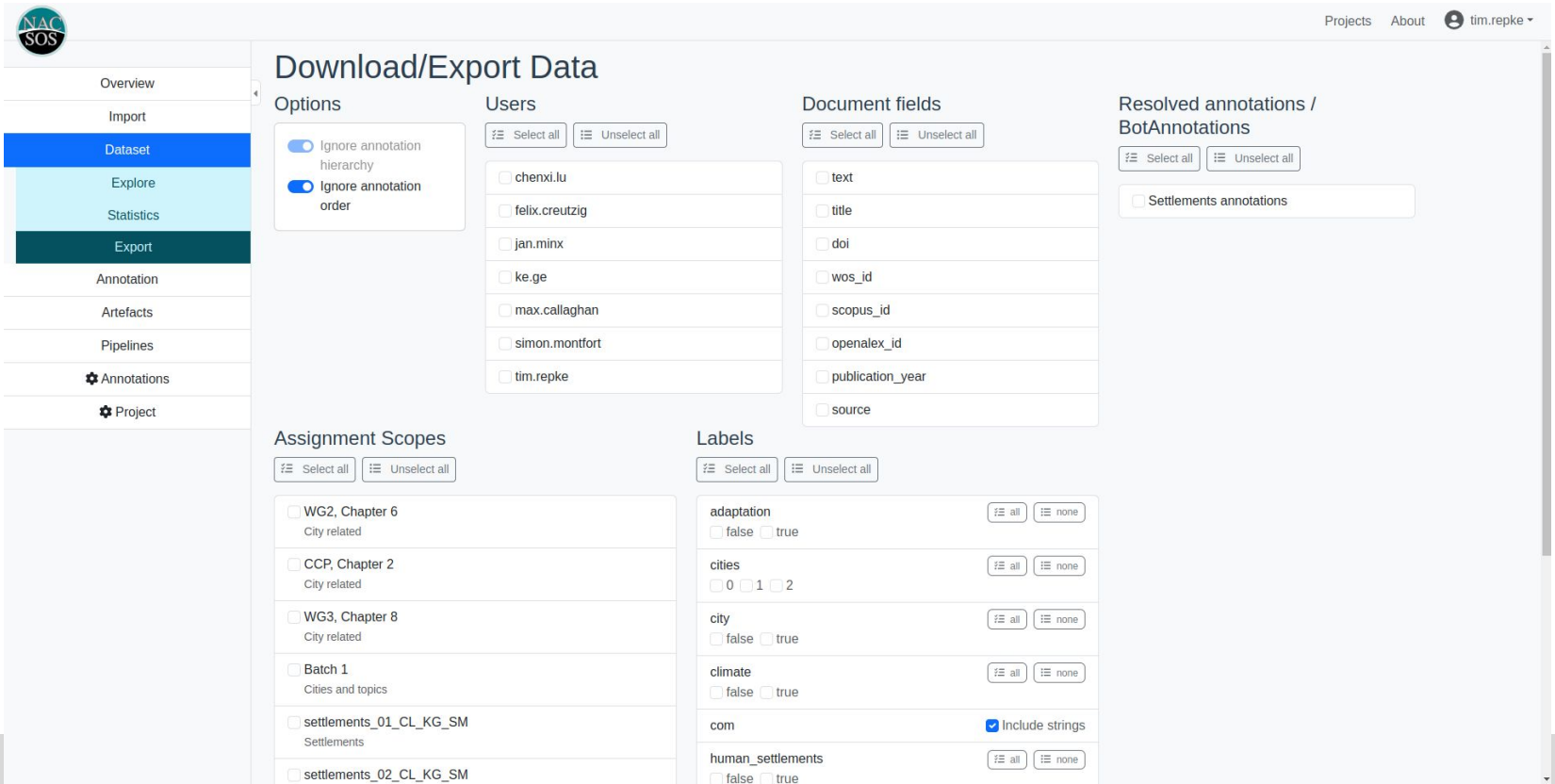
MCC

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

NAC SOS

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

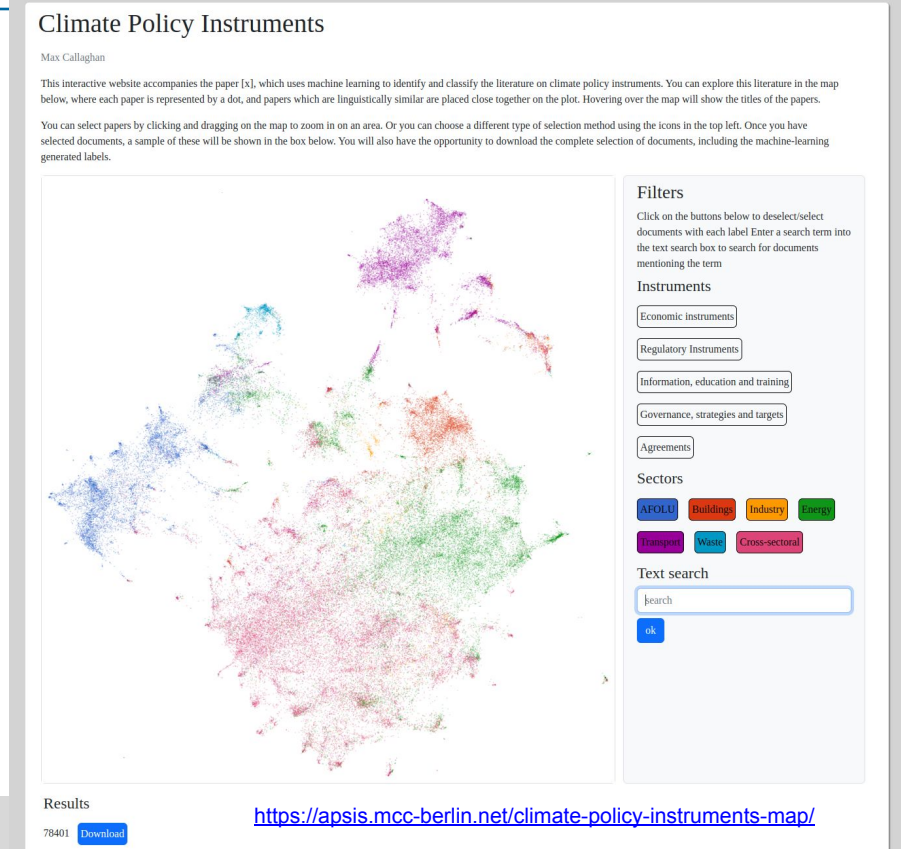
Beyond the platform



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



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/
- ...