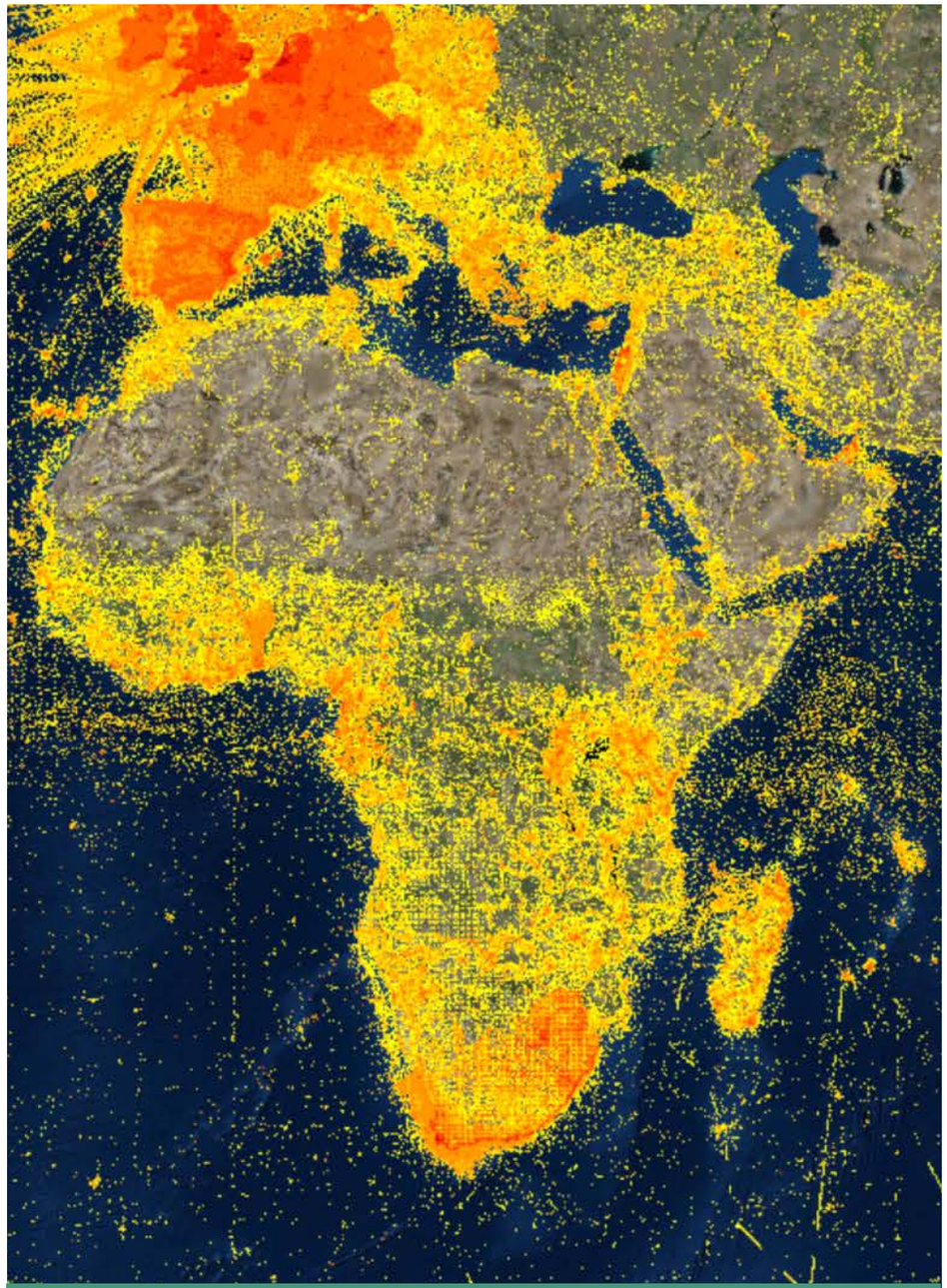
Monitoring ABS: From Permit Systems to Machine Learning

Paul Oldham
One World Analytics &
Manchester Institute of Innovation Research &
ABS Initiative



GBIF Occurrence Records for Species in Africa

The Capacity to Know

- How to improve knowledge about research and innovation?
- How to attract international research collaborations & investments?
- How to ensure that countries & communities benefit from R & D involving biodiversity & TK under the Nagoya Protocol?

Approaches

- Link national research permit systems together to make ABS easier & enable monitoring

- Use large scale digital methods to enable monitoring.

Permit Systems

- Take existing permit systems:
  - create and link up IT systems

- Bahamas - multiple systems
- Kenya - multiple systems
- India - monitoring module

The Nagoya Protocol: A Model Online Research Permit and Monitoring System

This is the project site for a model Online Research Permit and Monitoring System to support national implementation of the Nagoya Protocol.

The idea behind the model is to assist Parties to the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization of the United Nations Convention on Biological Diversity with implementing the Nagoya Protocol.

The model focuses on the creation of an online permit and monitoring system to make it easier for governments to administer research permit applications involving genetic resources and traditional knowledge and to monitor compliance under the Nagoya Protocol as well as making it easier to prepare national reports.

Download in Word and PDF

You can download Word versions of the sections in a .zip file here. For pdf versions go here.

You will also need to view the schematics which demonstrate the basic functions of the system. You can view them online from the Schematics menu or download them in powerpoint, keynote or pdf. The schematics are meant to be viewed as a slide show in presentation mode.

The draft workplan can be downloaded as headings to assist with project planning here.

Who Developed This?

The original model was written by Dr. Paul Oldham as part of work with Hartmut Meyer and Olivier Rukundo on implementation of the Nagoya Protocol in the Bahamas. The updated version is a joint work in progress and much better for it.

Financial Support

The model was developed with the support of The Bahamas Environment, Science & Technology Commission (BEST) of the Government of the Bahamas under the UNEP/GEF project "Strengthening Access and Benefit Sharing (ABS) in the Bahamas" as set out in *Oldham, P (2015) Concepts for an Electronic Monitoring Tool. UNEP/GEF project "Strengthening Access and Benefit Sharing (ABS) in the Bahamas"*. The present paper was written with the additional support of the multi-donor ABS Capacity Development Initiative hosted by the German Federal Ministry for

research-portal.nacosti.go.ke

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Sign Up

Login

Reseach Information Management System

Home Applications Guidelines Accredited Institutions Contact Us

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION (NACOSTI)

The RIMS Reseach Information Management System was developed by the Kenyan government to expedite permit application process for researchers at the same time improving the efficiency and transparency of the entire research permit application process.

GUIDELINES FOR ONLINE APPLICATIONS OF RESEARCH PERMIT

ORCID Login/Signup

research-portal.nacosti.go.ke/Login.html

Home Applications Guidelines Accredited Institutions Contact Us

RIMS

NACOSTI Research Information Management System

SIGN IN

NACOSTI is collecting your ORCID iD so we can link your ORCID iD with RIMS. When you click the "Authorize" button, we will ask you to share your iD using an authenticated process: either by registering for an ORCID iD or, if you already have one, by signing into your ORCID account, then granting us permission to get your ORCID iD. We do this to ensure that you are correctly identified and securely connecting your ORCID iD. Learn more about What's so special about signing in.

If you already have an ORCID account, you can use it with RIMS.
Login with your ORCID

Login

Not an ORCID user yet? Create an account, it is free and takes a few seconds. **We require basic profile information to be able to fill in your profile**

SignUp

ORCID

orcid.org/signin?oauth&client_id=APP-WCHNSELSEHON0HOY&response_type=code&scope=/authe...

ORCID uses cookies to improve your experience and to help us understand how you use our websites. Learn more about how we use cookies. Dismiss

ORCID

Sign into ORCID or Register now

Personal account | Institutional account

Sign in with your ORCID account

Email or ORCID iD

poldham@mac.com

ORCID password

Sign into ORCID

Forgotten your password? Reset it here

Sign in with a social media account

NACOSTI RIMS :: Registration

research-portal.nacosti.go.ke/Signup/profile/

Hi Paul, Please complete your Registration

ID/Passport Number *

ID/Passport Number

Title *

Choose option

First Name *

Paul

Middle Name

Middle Name

Last Name *

Oldham

Gender *

Male

Citizenship *

Choose option

Email *

Email

Mobile Phone Number *

Mobile

Attach Your ID/Passport
(.pdf only)

Choose file No file chosen

Attach Passport Photo (.jpgl.jpegl.png)
Max Width: 450px , Max Height: 450px

Choose file No file chosen

Next

Paul Oldham (0000-0002-101

orcid.org/0000-0002-1013-4390

ORCID
Connecting Research and Researchers

English

Registry | Search

FOR RESEARCHERS | FOR ORGANIZATIONS | ABOUT | HELP

7,218,074 ORCID iDs and counting. See more...

Paul Oldham

ORCID iD
https://orcid.org/0000-0002-1013-4390

Print view

Websites
Github site
One World Analytics website
Mendeley profile

Country
United Kingdom

Keywords
Social Anthropology, Science and Technology Studies, Biological Diversity, Biodiversity, Intellectual Property, Patents, Access and Benefit Sharing, Genetic Resources, Traditional Knowledge, Nagoya Protocol

Other IDs
Scopus Author ID: 26533255500

Employment (4)

Sort

One World Analytics: Lancaster, Lancashire, GB
2013-04-01 to present | Director
Employment
Source: Paul Oldham ★ Preferred source

United Nations University: Tokyo, JP
2016-01-01 to 2019-02-01 | Senior Visiting Research Fellow (Institute for the Advanced Study of Sustainability)
Employment
Source: Paul Oldham ★ Preferred source

Lancaster University: Lancaster, Lancashire, GB
2004-01-01 to 2013-02-01 | Research Fellow, Senior Research Associate, Research Associate (ESRC Centre for Economic and Social Aspects of Genomics)
Employment
Source: Paul Oldham ★ Preferred source

University of London Institute of Latin American Studies: London, London, GB
1996-09-01 to 1999-09-01 | Lecturer
Employment

Help

1013-4390

Works (39 of 39) Sort

The Scientific and Patent Landscape for Marine Genetic Resources in Southeast Asia

2019-04-03 | report

Source: Paul Oldham ★ Preferred source

Synthetic Biology - Mapping the Patent Landscape

2018-11-30 | other

DOI: 10.1101/483826

Source: Crossref ★ Preferred source

Synthetic Biology Patent Datasets

2018 | other

DOI: 10.17605/OSF.IO/73FMU

Source: Paul Oldham ★ Preferred source

An Online Research Permit and Monitoring System to Support Implementation of the Nagoya Protocol

2016-05 | online-resource

DOI: http://dx.doi.org/10.5281/zenodo.57420

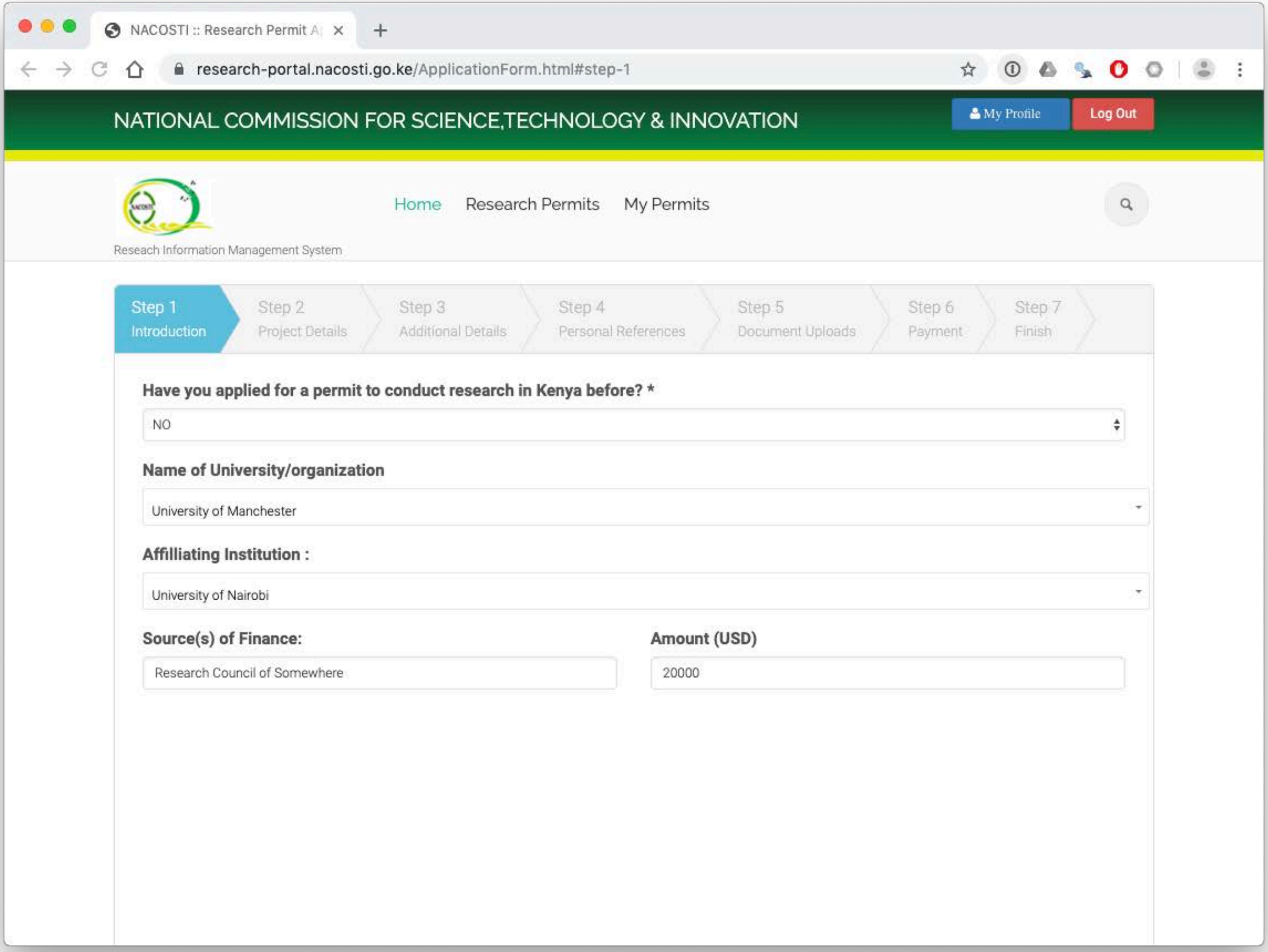
Source: Paul Oldham ★ Preferred source

A Review of UK Patent Activity for Genetic Resources and associated Traditional Knowledge

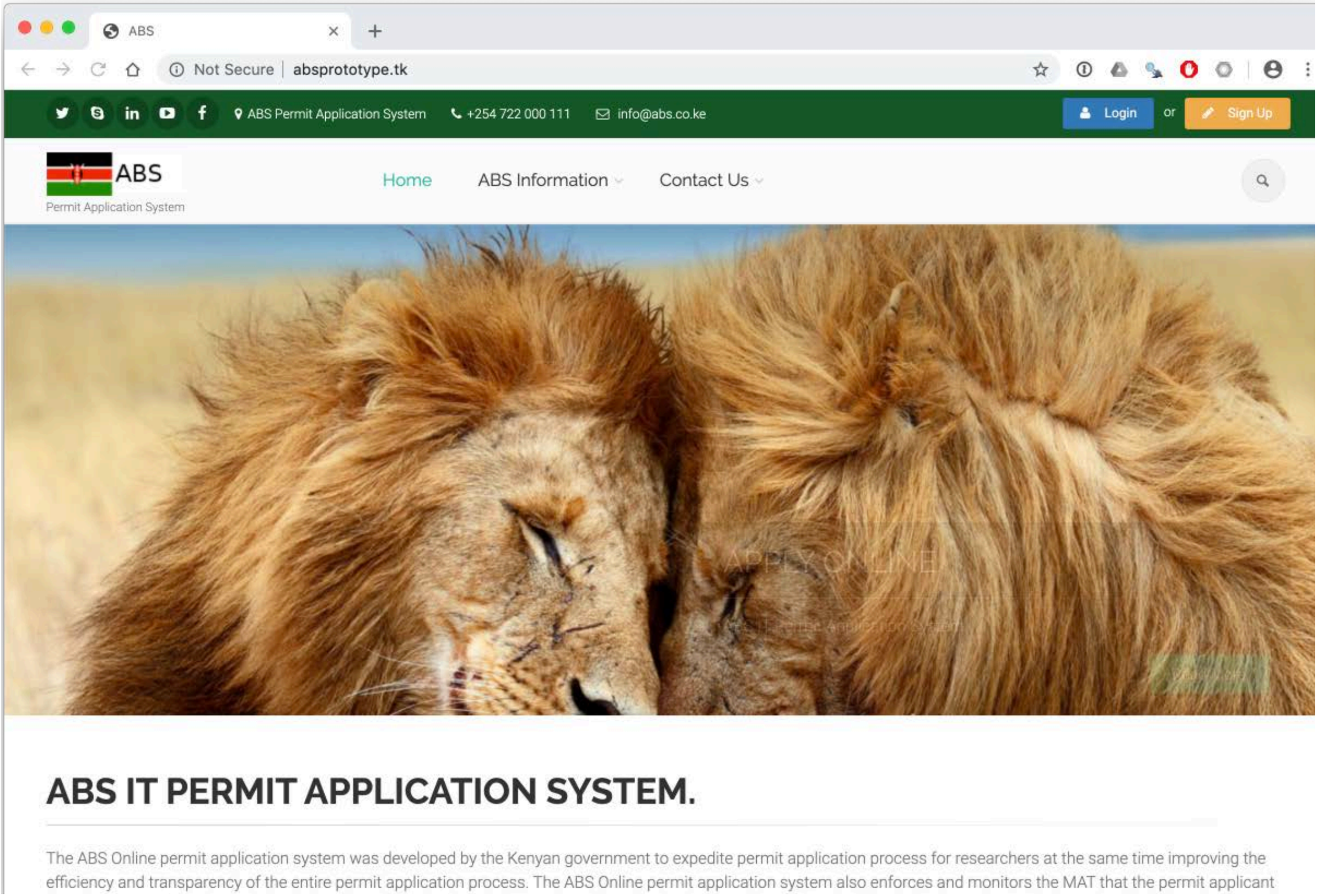
2016-01 | report

DOI: 10.13140/RG.2.1.1682.9849

Source: Paul Oldham ★ Preferred source



Filling out standard questions. Only ask a question once and use tidy data principles.



An ABS portal links to the system. Code for the prototype can be adapted to serve the needs of others.

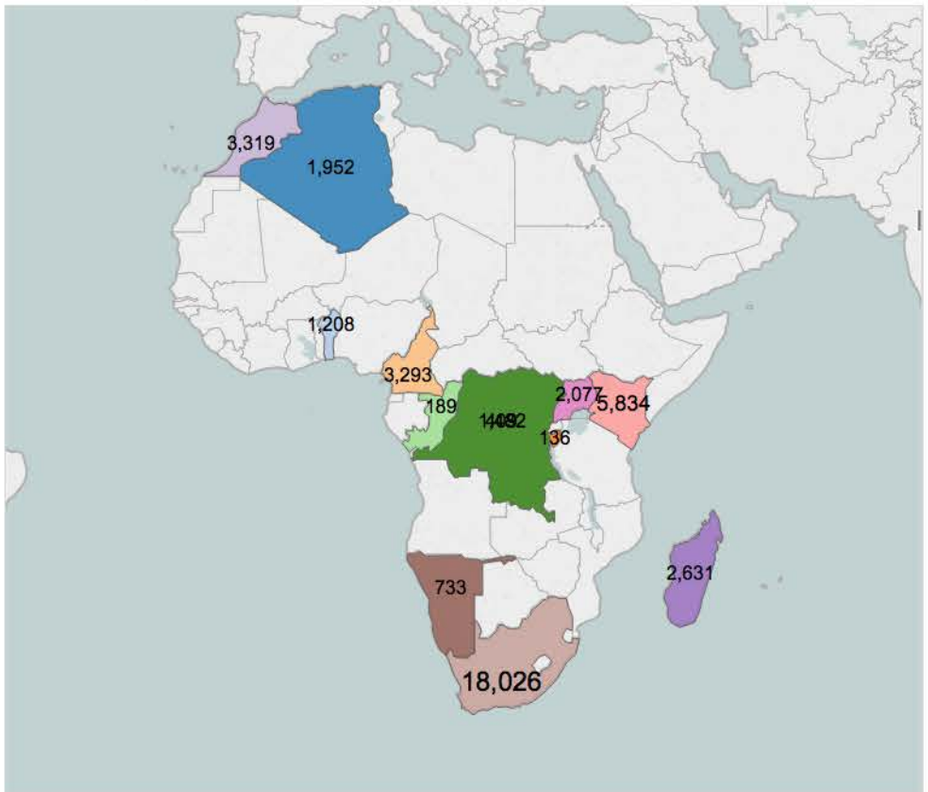
Monitoring

Addressing Issues of Scale:

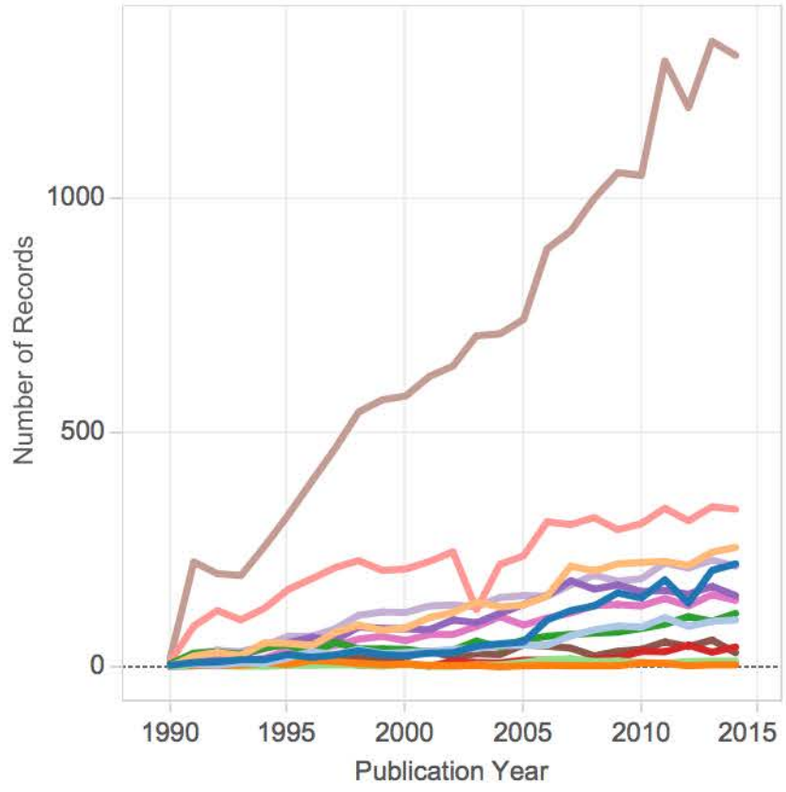
- 1 billion GBIF taxonomic occurrence records
- 211 million scientific publications
- 110+ million patent documents
- The internet (hundreds of millions)

Research Landscape for Genetic Resources & Traditional Knowledge in Africa

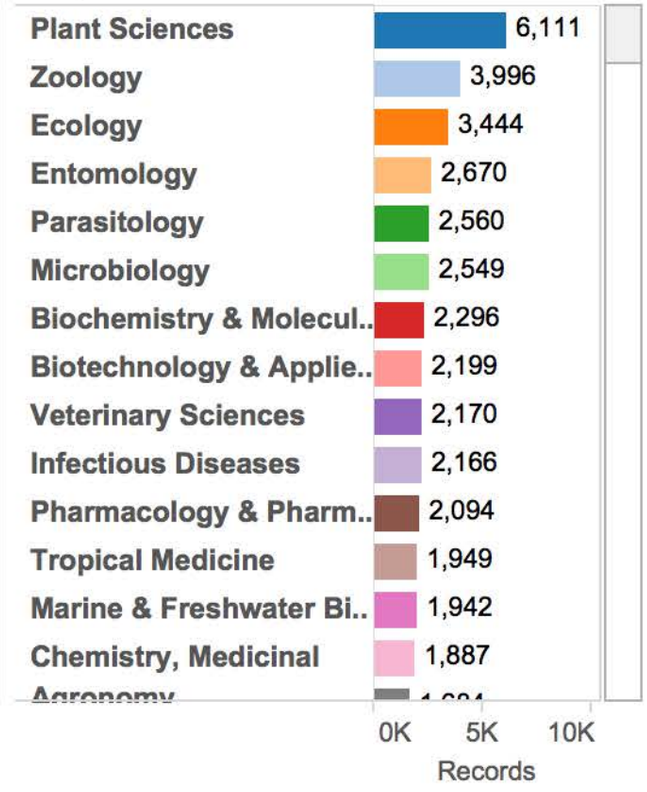
landscape countries



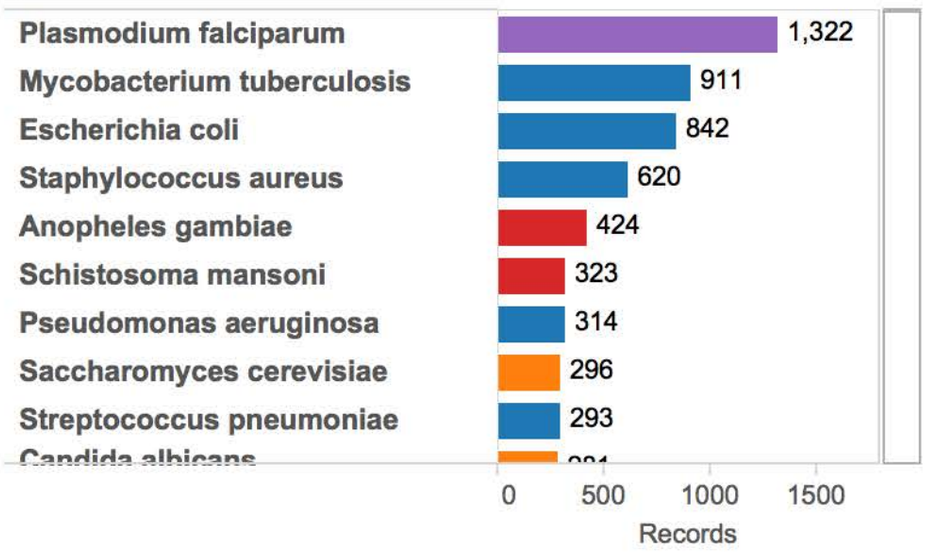
country trends



subject areas



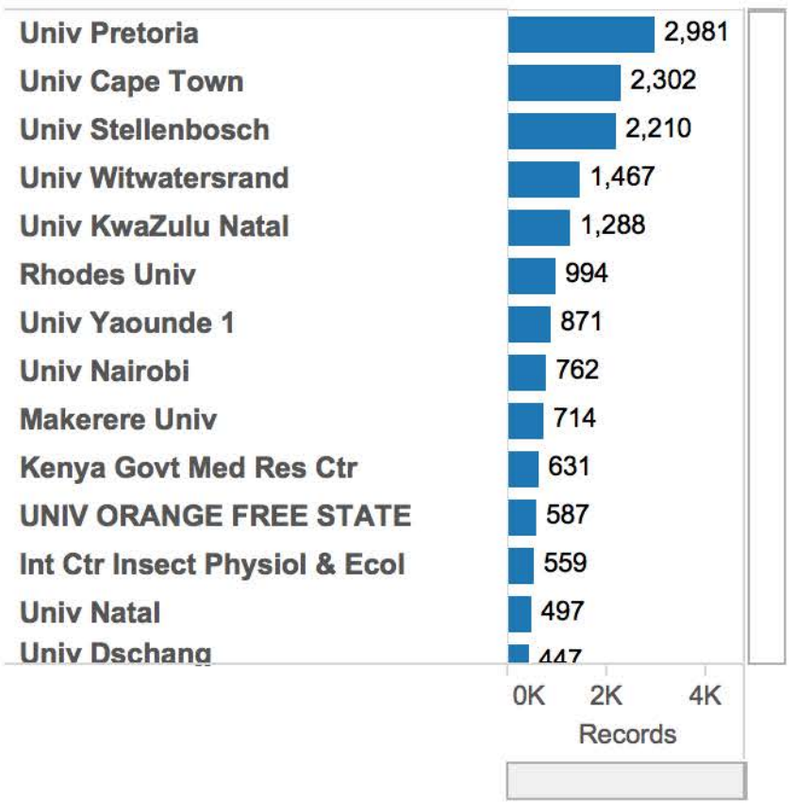
species



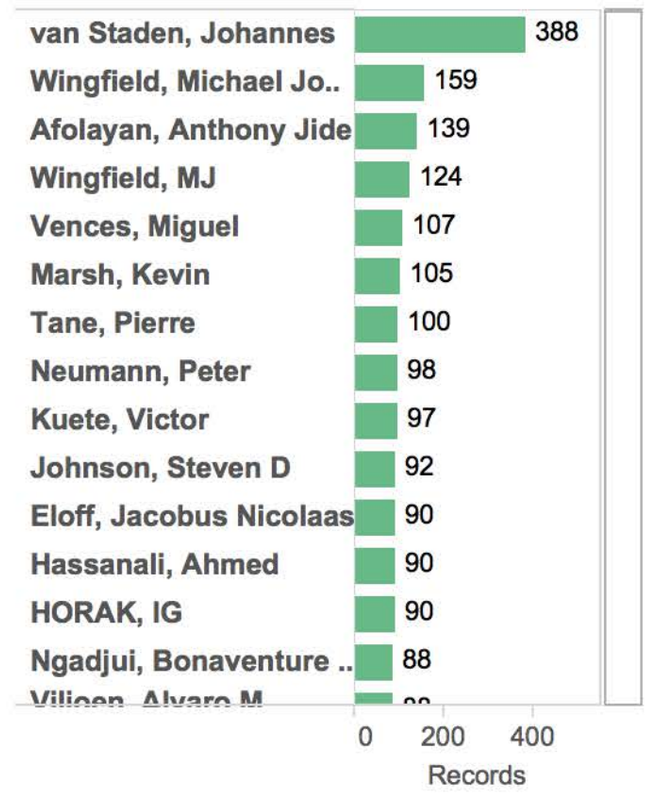
Kingdom

- Animalia
- Archaea
- Bacteria
- Chromista
- Fossil
- Fungi
- Plantae
- Protozoa
- Unknown

organisations



authors

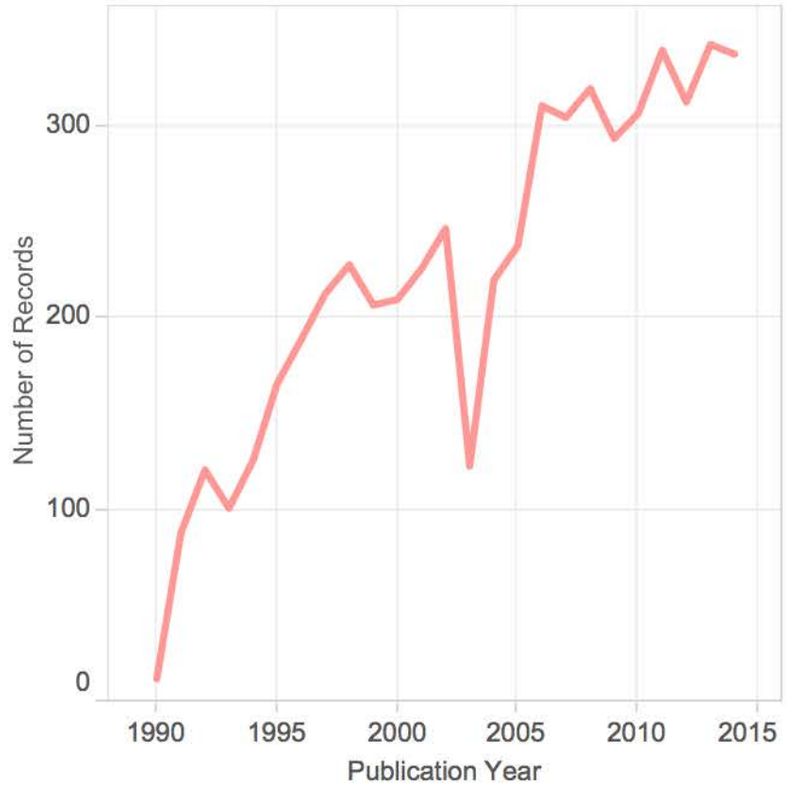


Research publications involving an author from selected African countries or referencing those countries that also include a species. Data source: Web of Science.

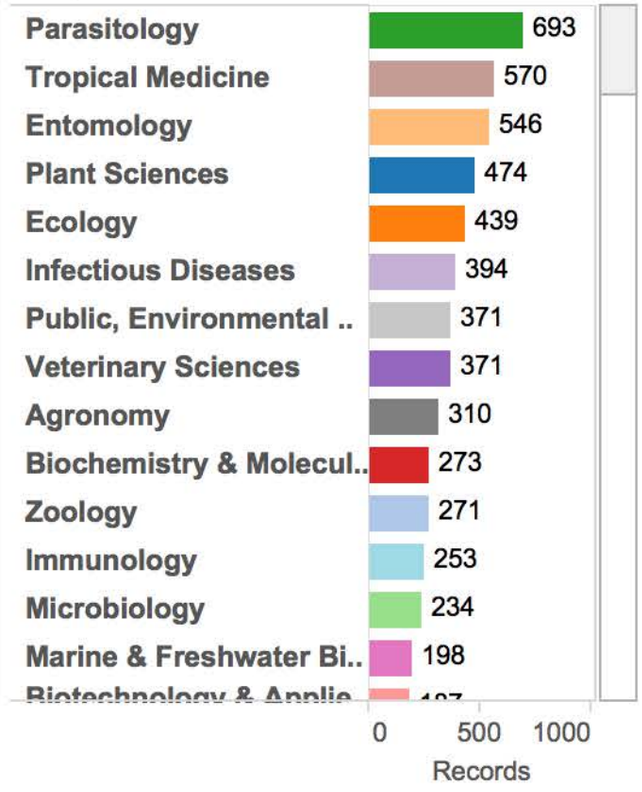
Research Landscape for Genetic Resources & Traditional Knowledge in Kenya

landscape countries

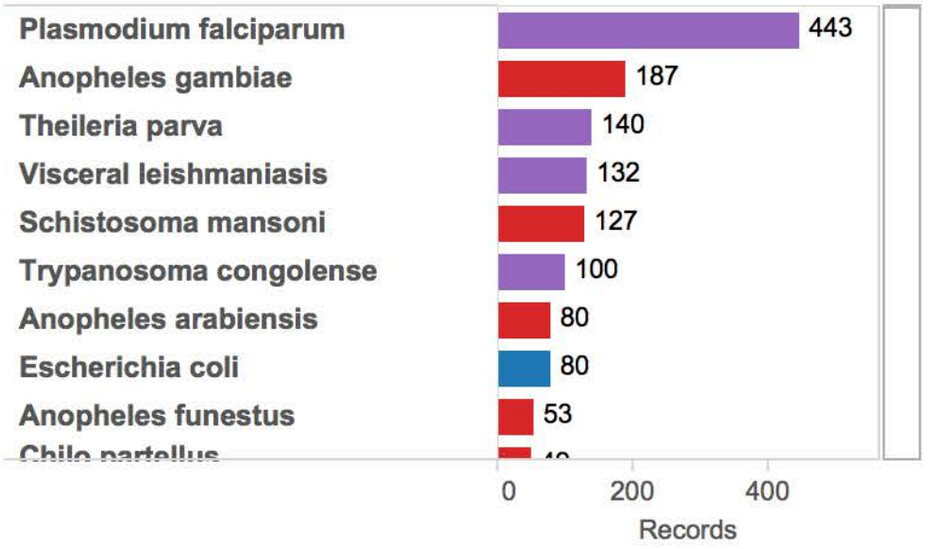
country trends



subject areas



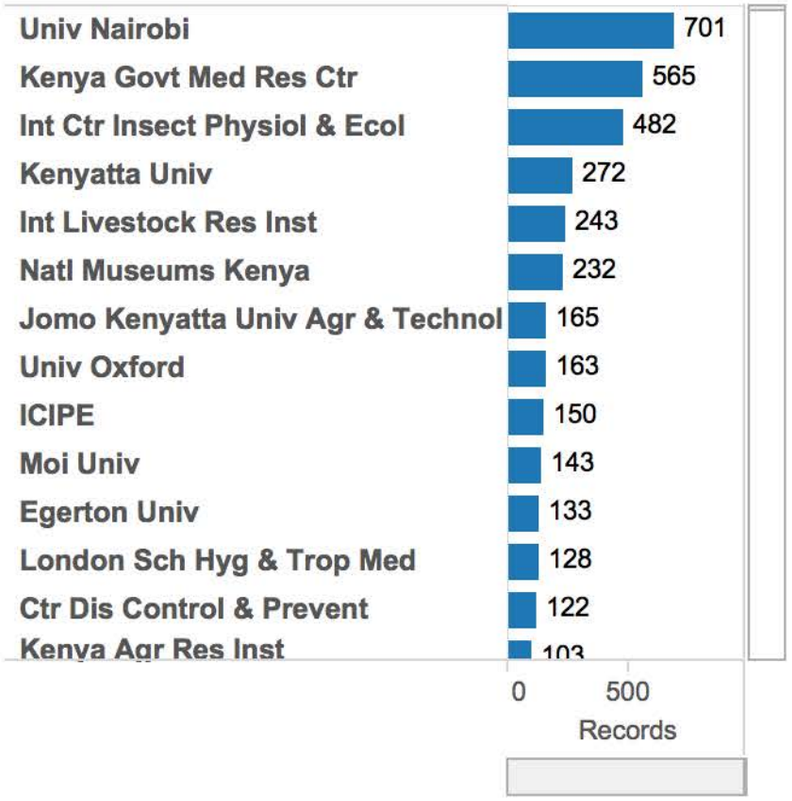
species



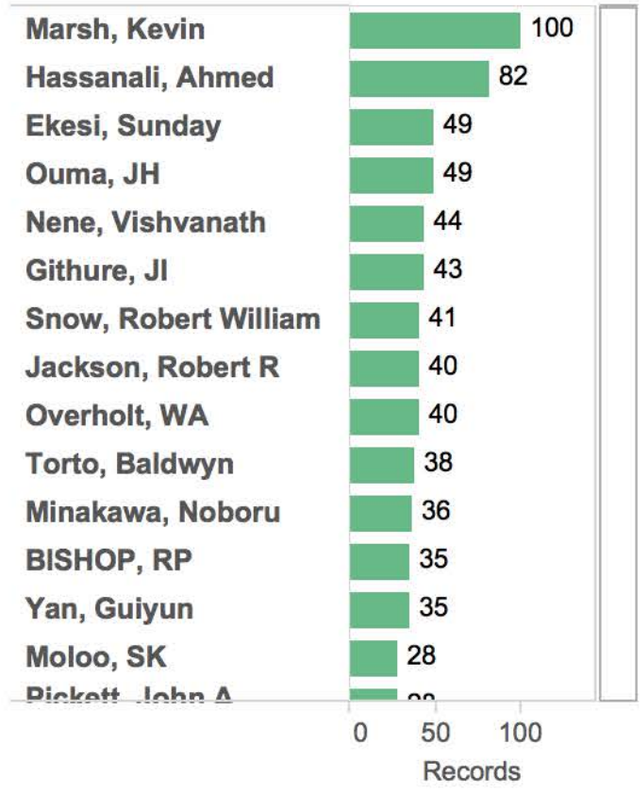
Kingdom

Animalia
Bacteria
Chromista
Fossil
Fungi
Plantae
Protozoa
Viruses

organisations

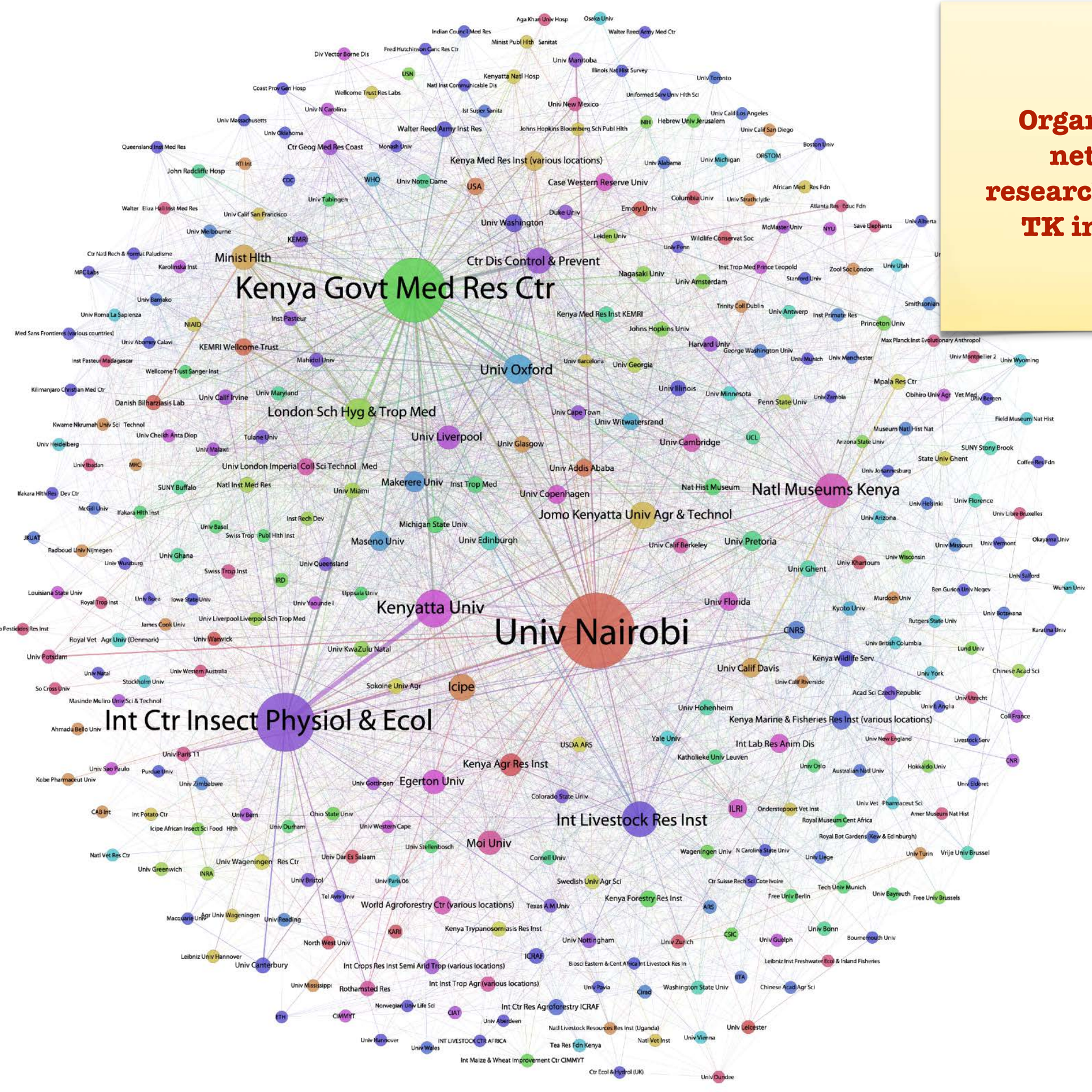


authors

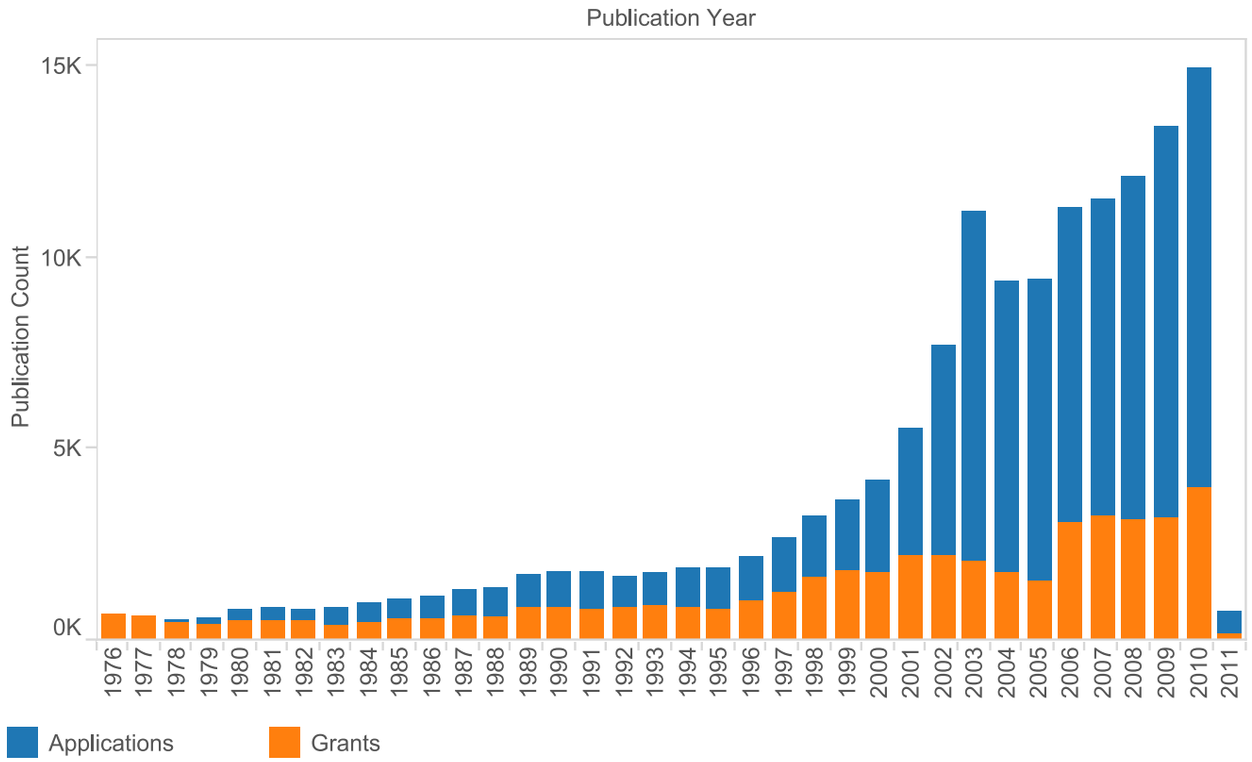


Research publications involving an author from Kenya or referencing Kenya that also include a species. Data source: Web of Science.

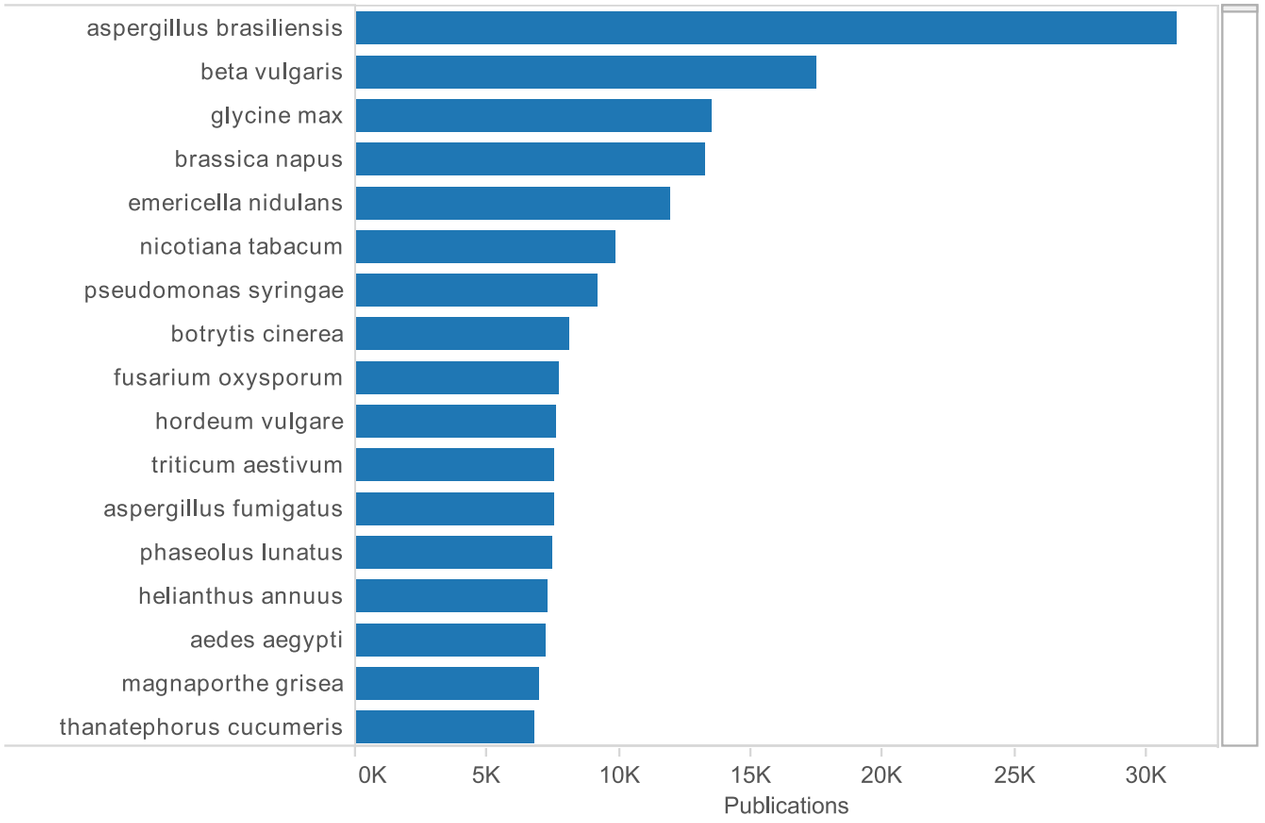
Organisation network researching GR & TK in Kenya



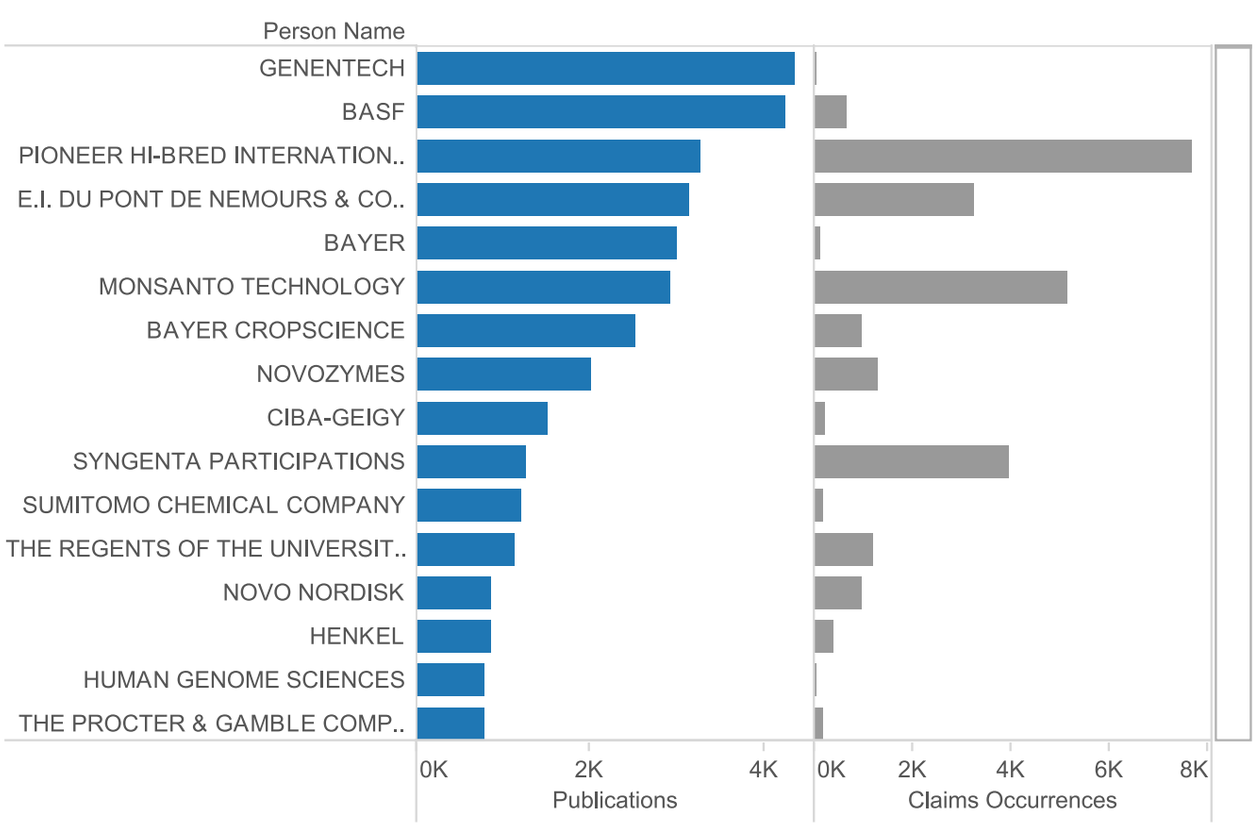
Global Trends



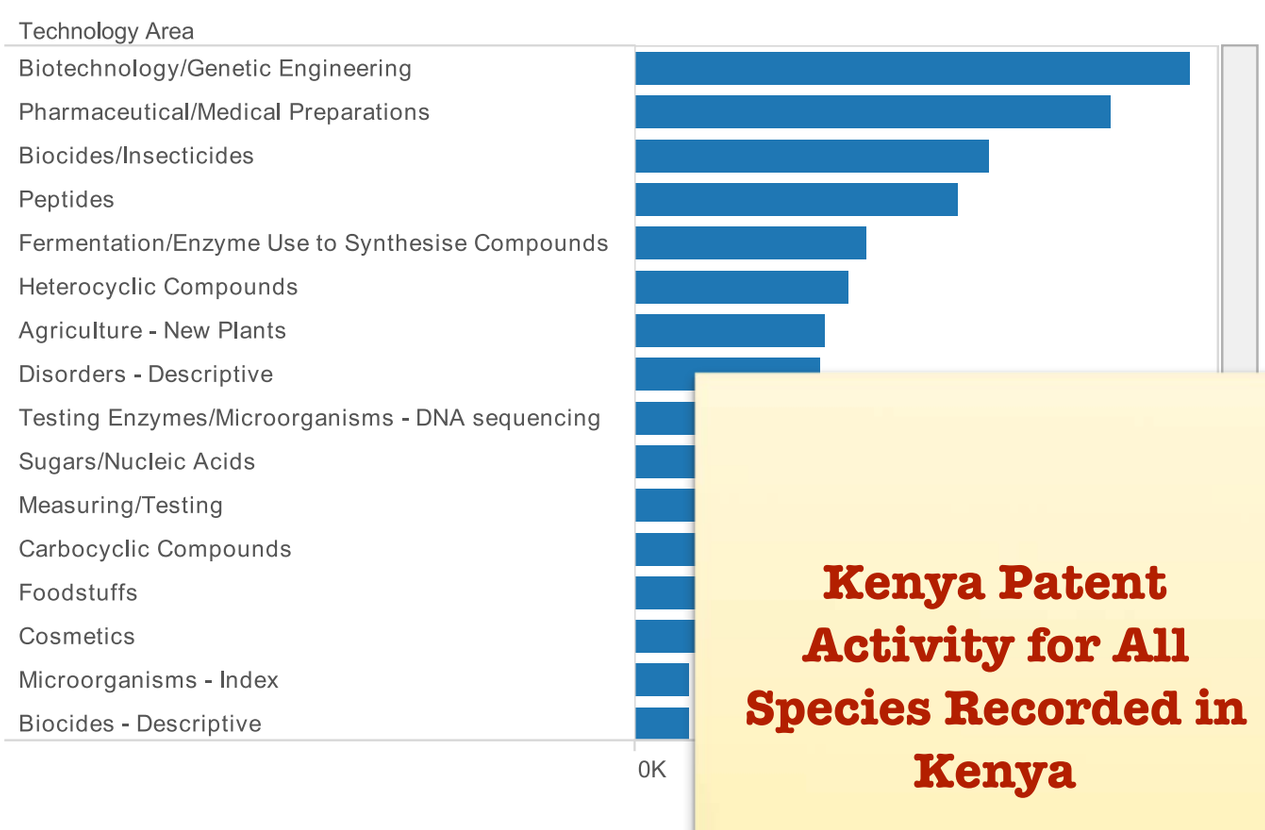
Global Species



Assignees



Global Technology Areas



Kenya Patent Activity for All Species Recorded in Kenya

This is hard and time consuming to do:

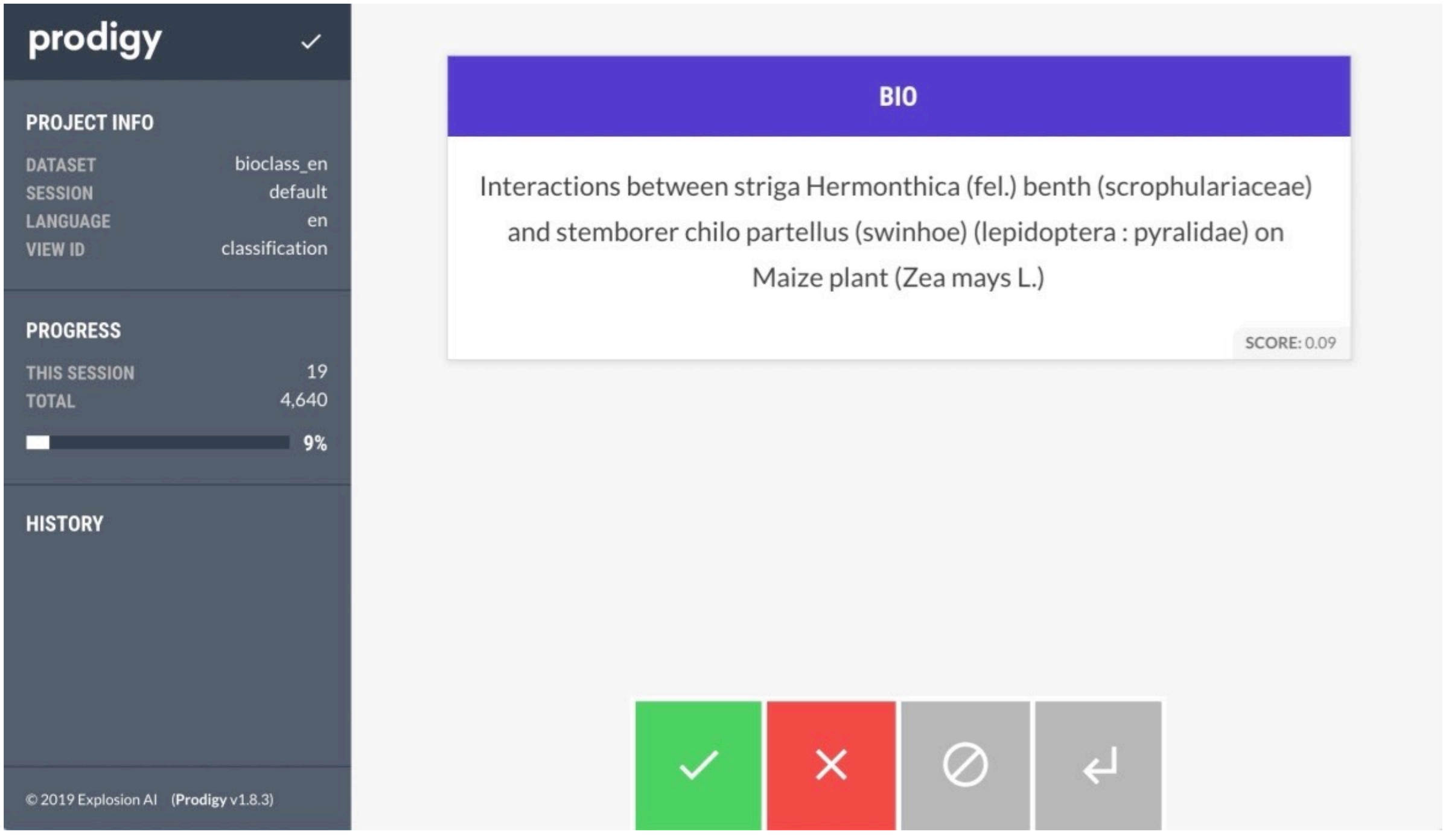
How to automate at scale?

Use machine learning to train a mathematical model to recognise patterns and predict what they are (e.g. is this a dog or a cat)

We are using spaCy from Explosion AI in Germany. spaCy is free, open source and transparent

Step 1: Classification

Is this document about biodiversity (YES/NO)



Prodigy from Explosion AI in Germany is a tool for annotating data to train machine learning models. Here we want to create a general classification model to identify documents that are biological. The more examples used to train models the better. Here we use a few thousand.

prodigy

PROJECT INFO

DATASET	bioclass_en
SESSION	default
LANGUAGE	en
VIEW ID	classification

PROGRESS

THIS SESSION	26
TOTAL	4,647

9%

HISTORY

- Herbs such as tongkat ali (eury... ✓
- Use of trichoderma and bacillu... ✓
- The pharmacokinetics of furaz... ✓
- For best results smaller quanti... ✓
- Extracts and compounds from ... ✓
- Relative dental maturity and a... ✓
- Interactions between striga H... ✓

BIO

The North Korean Conundrum and the Deficiencies of Western-Rational Social Theory

SCORE: 0.60

prodigy

PROJECT INFO

DATASET	bioclass_en
SESSION	default
LANGUAGE	en
VIEW ID	classification

PROGRESS

THIS SESSION	21
TOTAL	4,642

9%

HISTORY

Relative dental maturity and a... ✓

Interactions between striga H... ✓

BIO

Extracts and compounds from agapanthus africanus and their use as biological plant protecting agents

SCORE: 0.10

prodigy

PROJECT INFO

DATASET	bioclass_en
SESSION	default
LANGUAGE	en
VIEW ID	classification

PROGRESS

THIS SESSION	28
TOTAL	4,649

9%

HISTORY

- The invention provides an indu... ✓
- The North Korean Conundrum... ✗
- Herbs such as tongkat ali (eury... ✓
- Use of trichoderma and bacillu... ✓
- The pharmacokinetics of furaz... ✓
- For best results smaller quanti... ✓
- Extracts and compounds from ... ✓

BIO

A Perceptual Study of Relative Effectiveness of Tools and Techniques Used in Sales Promotion

SCORE: 0.98

#	LOSS	F-SCORE	ACCURACY
01	0.811	0.832	0.783
02	0.439	0.908	0.885
03	0.234	0.922	0.904
04	0.122	0.930	0.914
05	0.112	0.932	0.918
06	0.100	0.935	0.921
07	0.091	0.937	0.923
08	0.083	0.934	0.921
09	0.063	0.940	0.927
10	0.089	0.928	0.913

accept	accept	505
accept	reject	36
reject	reject	325
reject	accept	29

Correct	830
Incorrect	65

Baseline	0.60
Precision	0.93
Recall	0.95
F-score	0.94
Accuracy	0.93

Train a Model

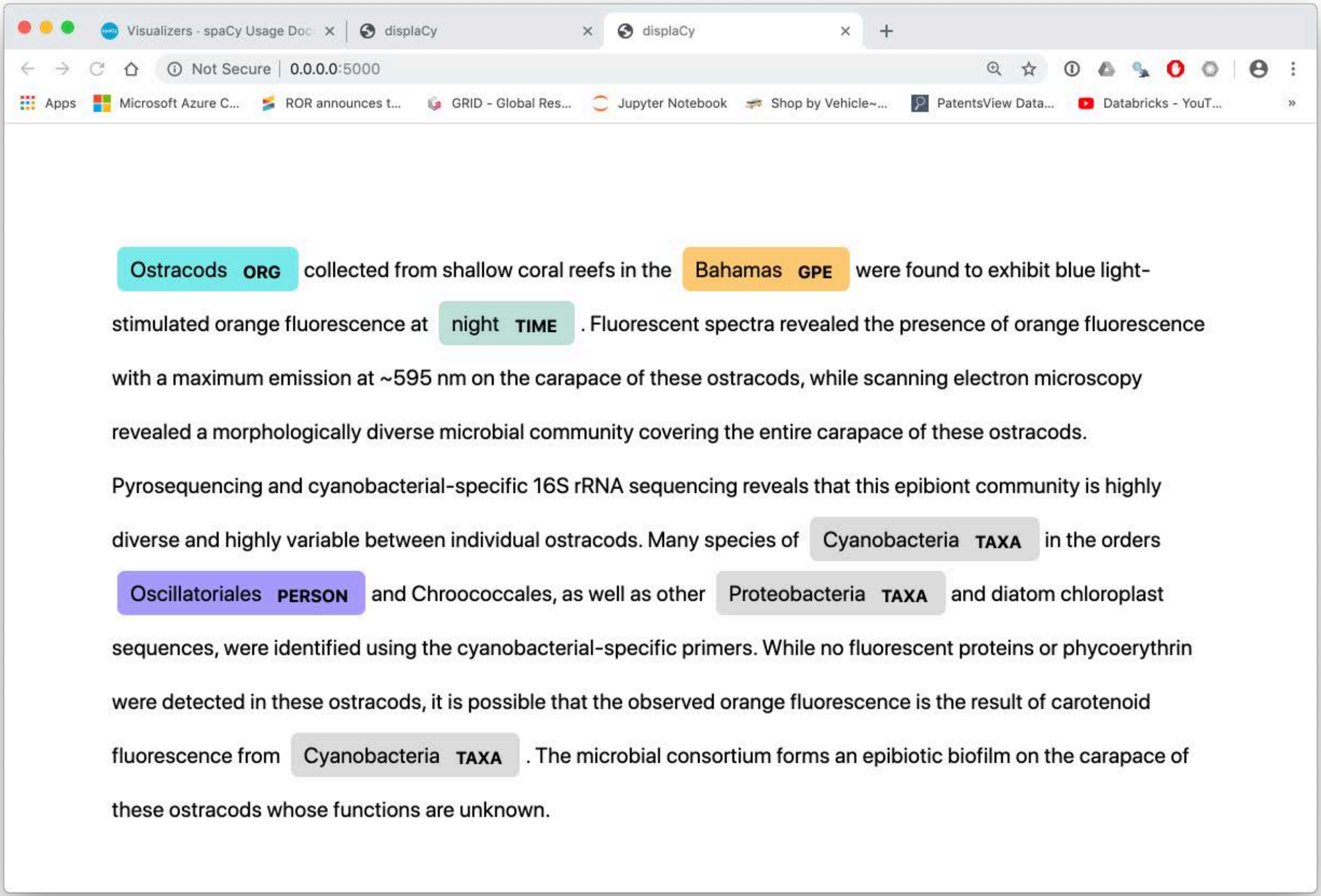
Here we use 3713 80% of the annotations we just made to train the model and we then test the results on an evaluation set with 978 (20%) examples the model has not seen. This gives us scores on how good the model is with unseen data (beware of over fitting - where a model becomes great at predicting data it has seen but rubbish with unseen data!)

Step 2:
Name Entity Recognition

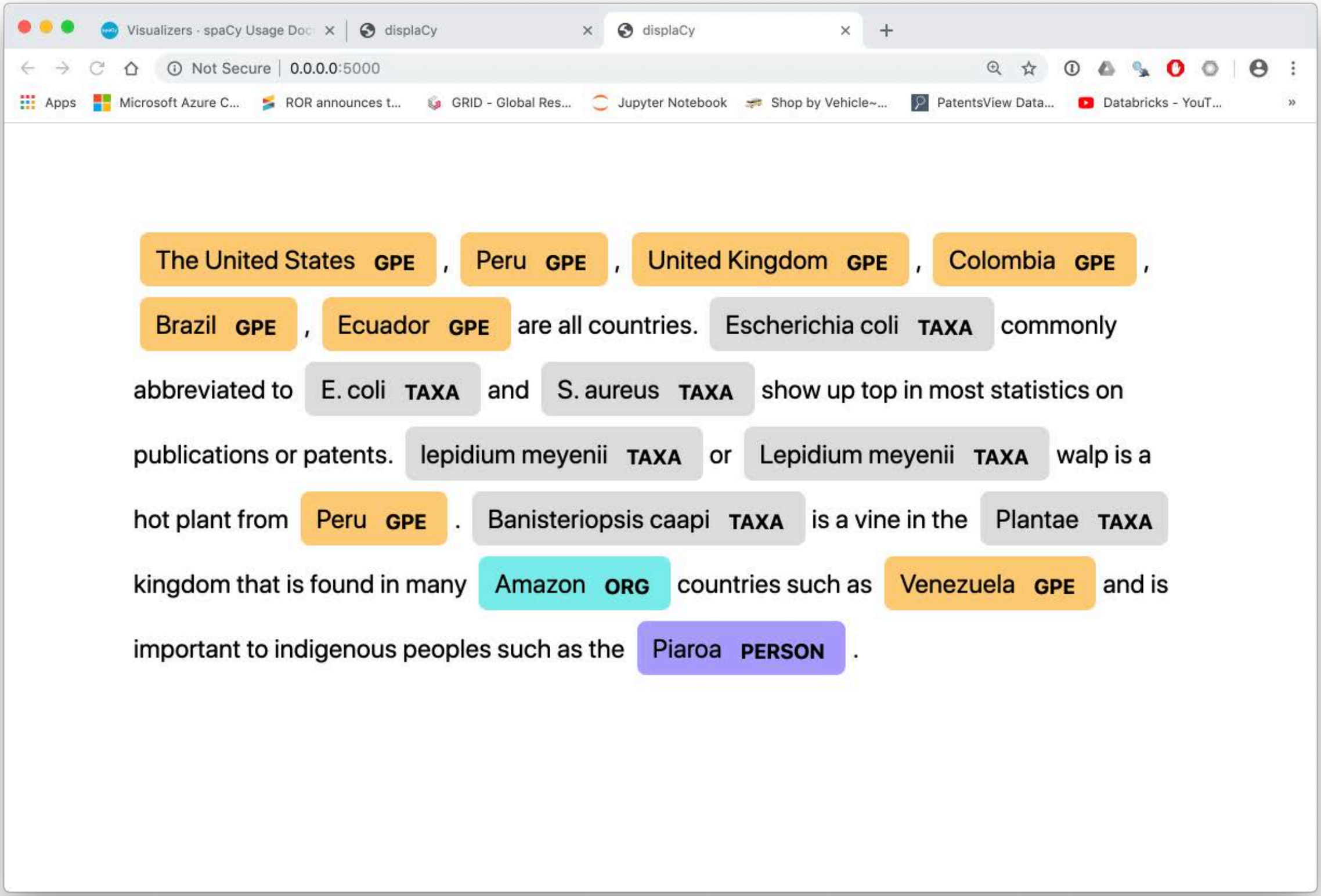
> Show me species, place, country, person, organisation or TK related terms in a document

Ostracods collected from shallow coral reefs in the Bahamas were found to exhibit blue light-stimulated orange fluorescence at night. Fluorescent spectra revealed the presence of orange fluorescence with a maximum emission at ~595 nm on the carapace of these ostracods, while scanning electron microscopy revealed a morphologically diverse microbial community covering the entire carapace of these ostracods. Pyrosequencing and cyanobacterial-specific 16S rRNA sequencing reveals that this epibiont community is highly diverse and highly variable between individual ostracods. Many species of Cyanobacteria in the orders Oscillatoriales and Chroococcales, as well as other Proteobacteria and diatom chloroplast sequences, were identified using the cyanobacterial-specific primers. While no fluorescent proteins or phycoerythrin were detected in these ostracods, it is possible that the observed orange fluorescence is the result of carotenoid fluorescence from Cyanobacteria. The microbial consortium forms an epibiotic biofilm on the carapace of these ostracods whose functions are unknown.

This is hard to read. We need to find a way to mark the interesting bits.



The model is identifying some of the entities but is getting some of the labels wrong - because the base model has not been trained on this type of data. I have added the Taxa category to the model but can you see where more work is needed to pick up biodiversity related terms?



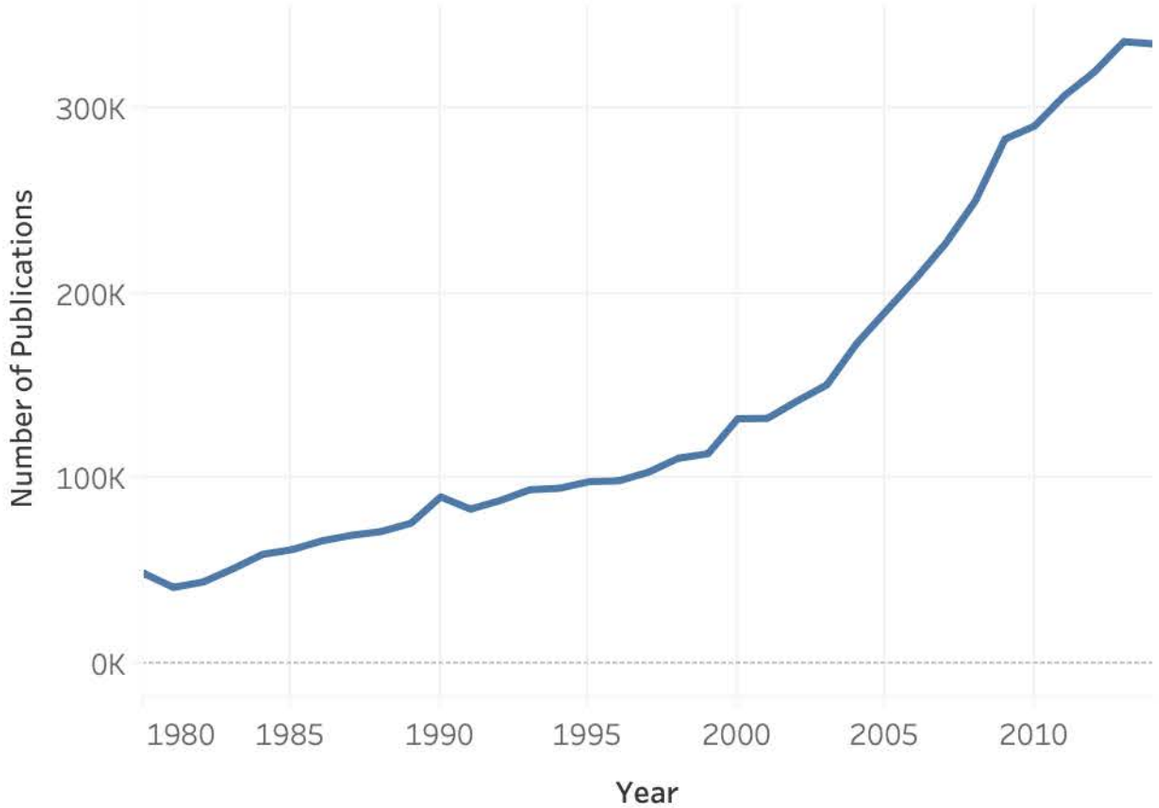
This example is better and picks up abbreviated species names and does well on country names. But it thinks Amazon is an organisation and that the Piaroa are a person. This highlights a need to train up the named entity recogniser but will work well once trained.

Step 3: Scale Up

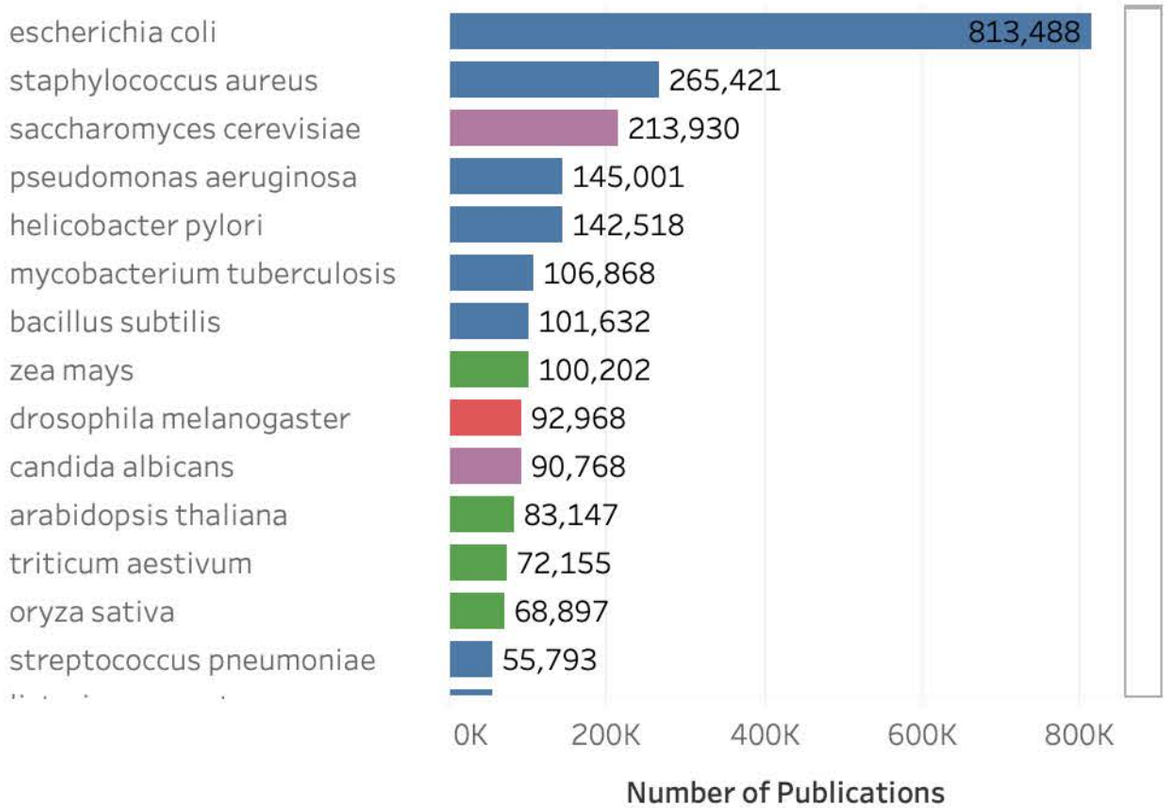
Apply to 211 million publications

Here a simpler dictionary approach was used to text mine 166 million publications from Microsoft Academic Graph. It allows us to visualise trends and actors and to draw maps. This is very much a first draft leading to the 2020 COP in China. This data is now being used to train models to work at scale.

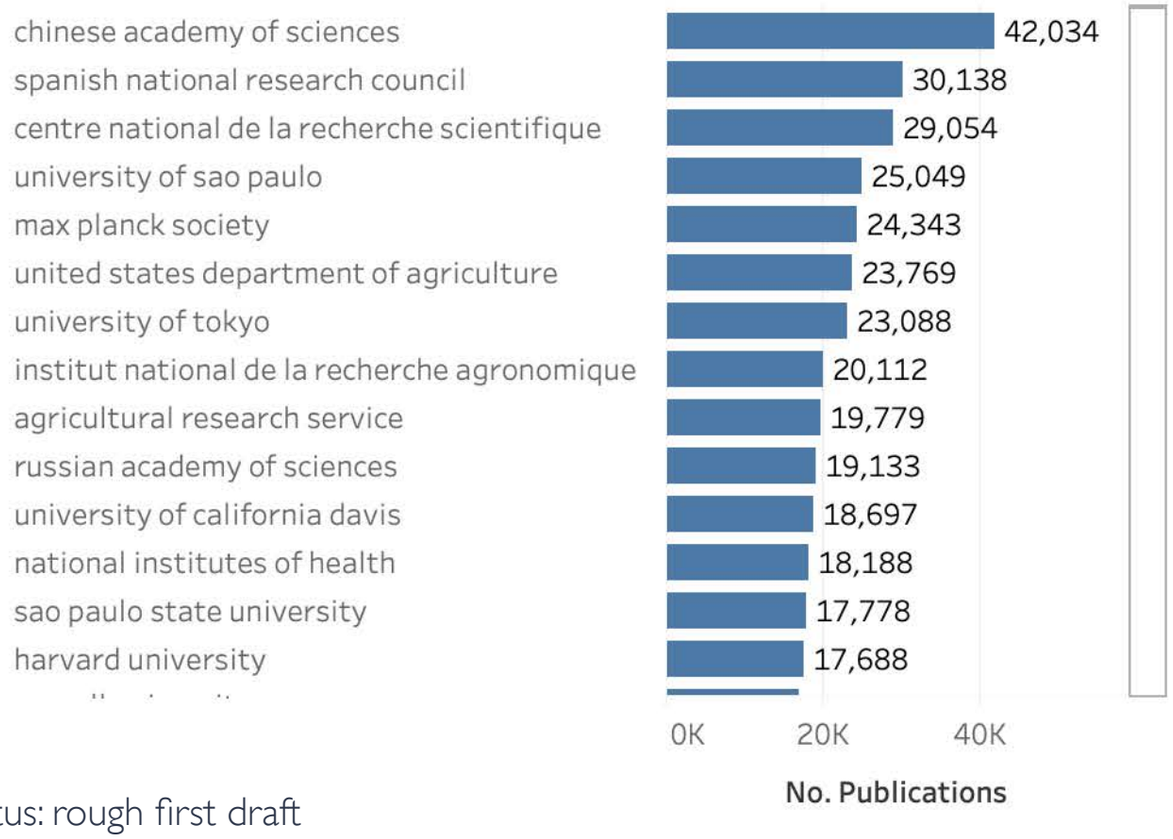
Trends



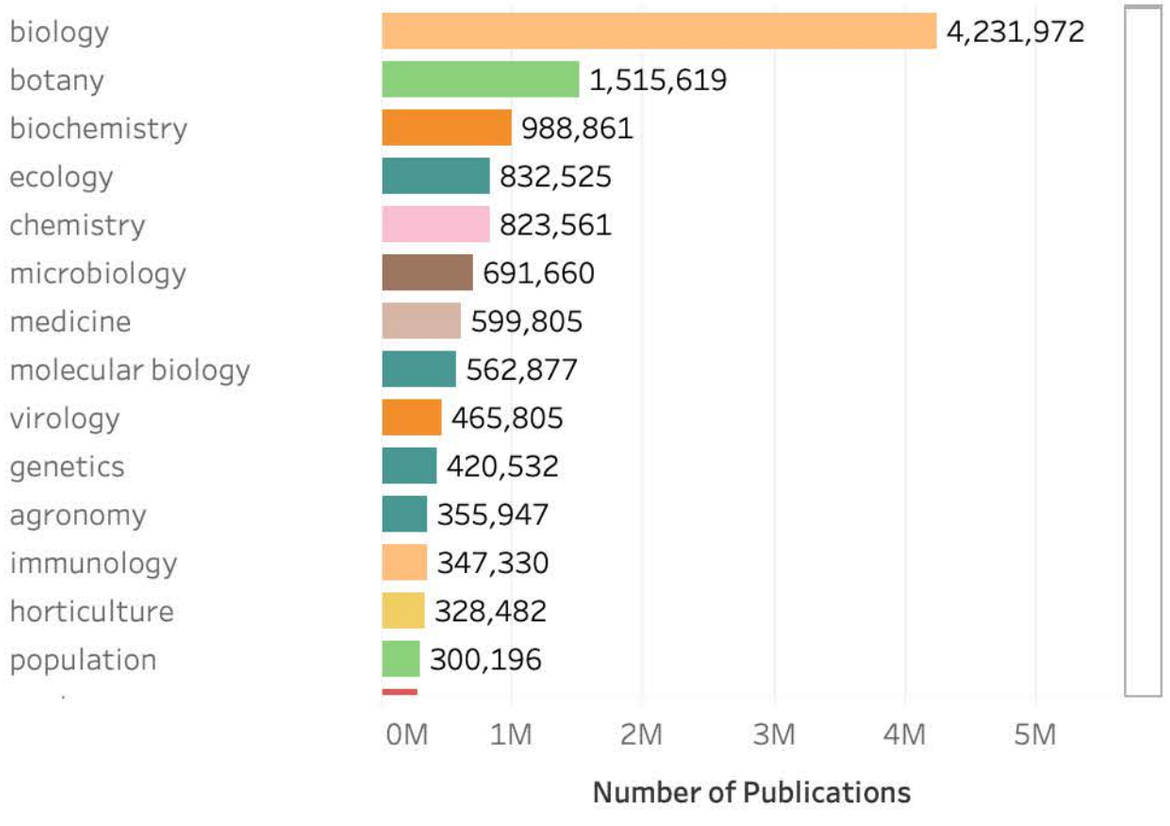
Species



Organisations

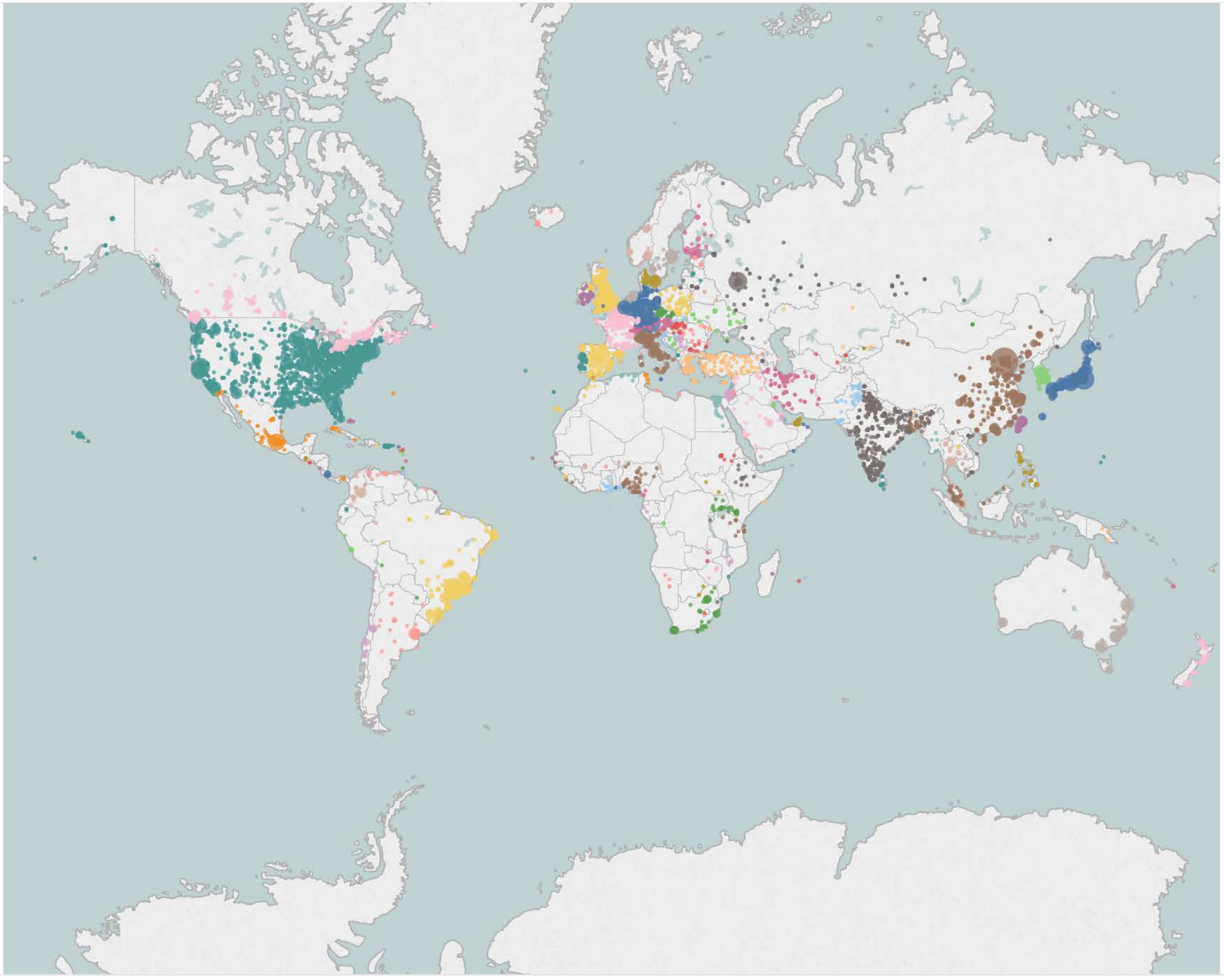


Fields of Study

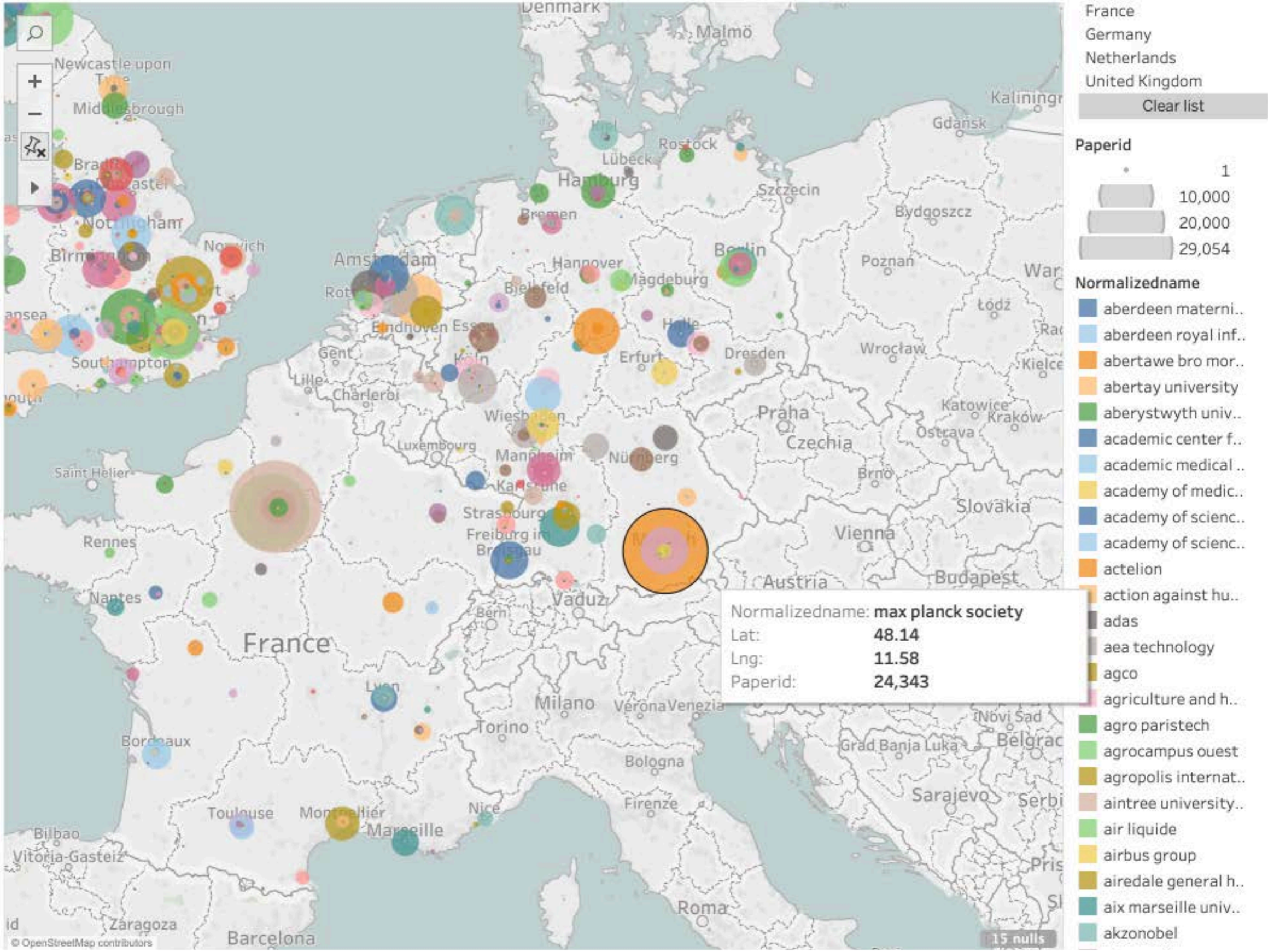


Status: rough first draft

Biodiversity Related Research Worldwide (First Draft)

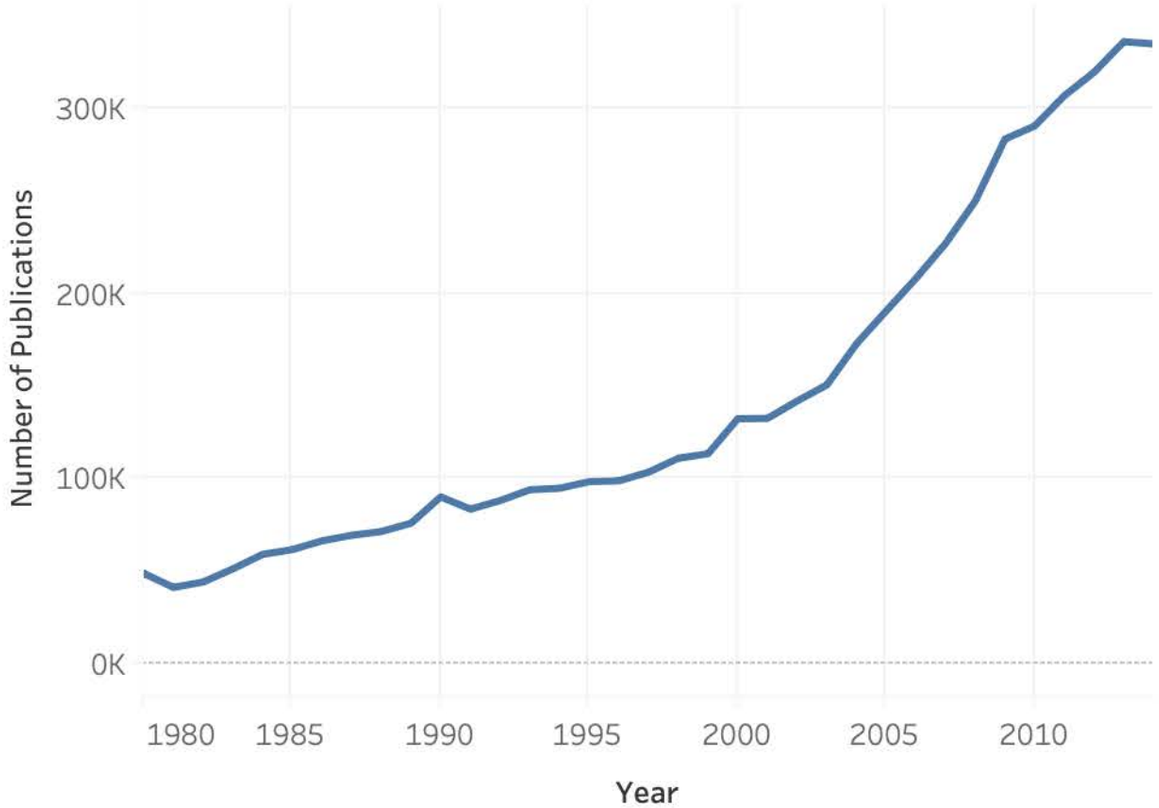


Binomial species names appearing in publication metadata. Source: Microsoft Academic Graph 1800 - 2017.

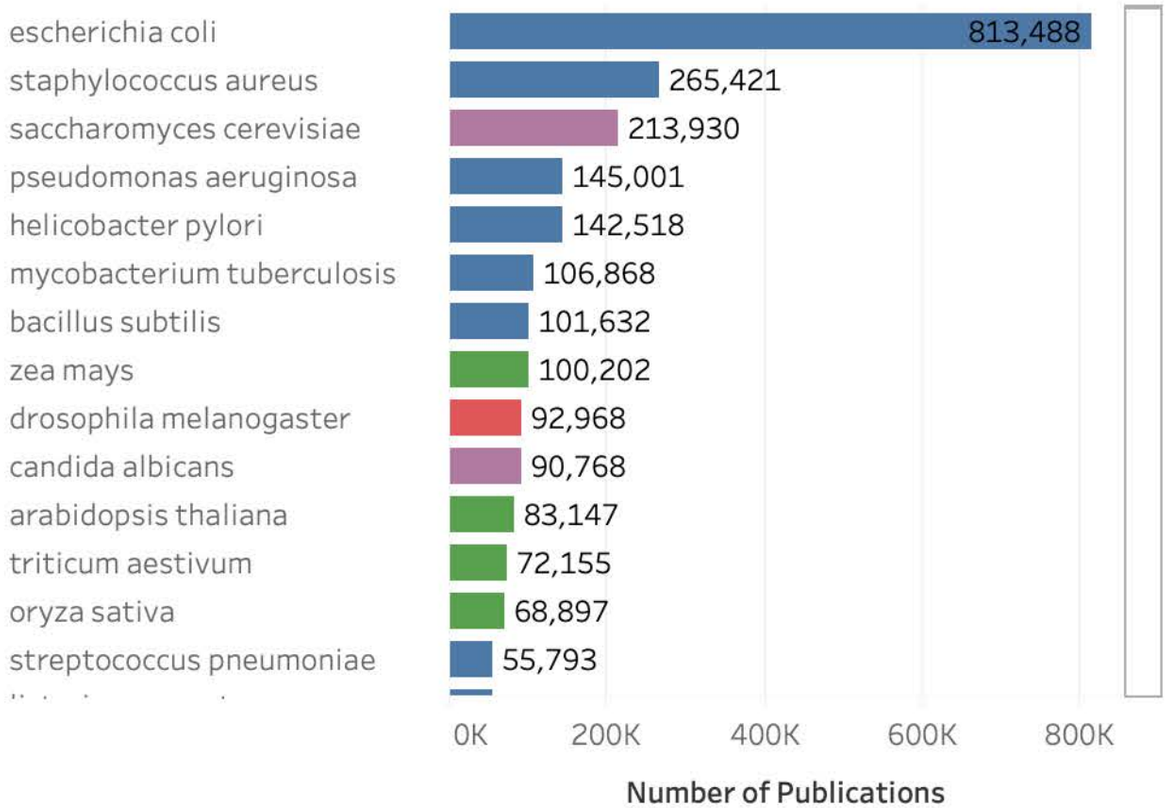


Using the open access GRID dataset we can map organisations around the world publishing on species.

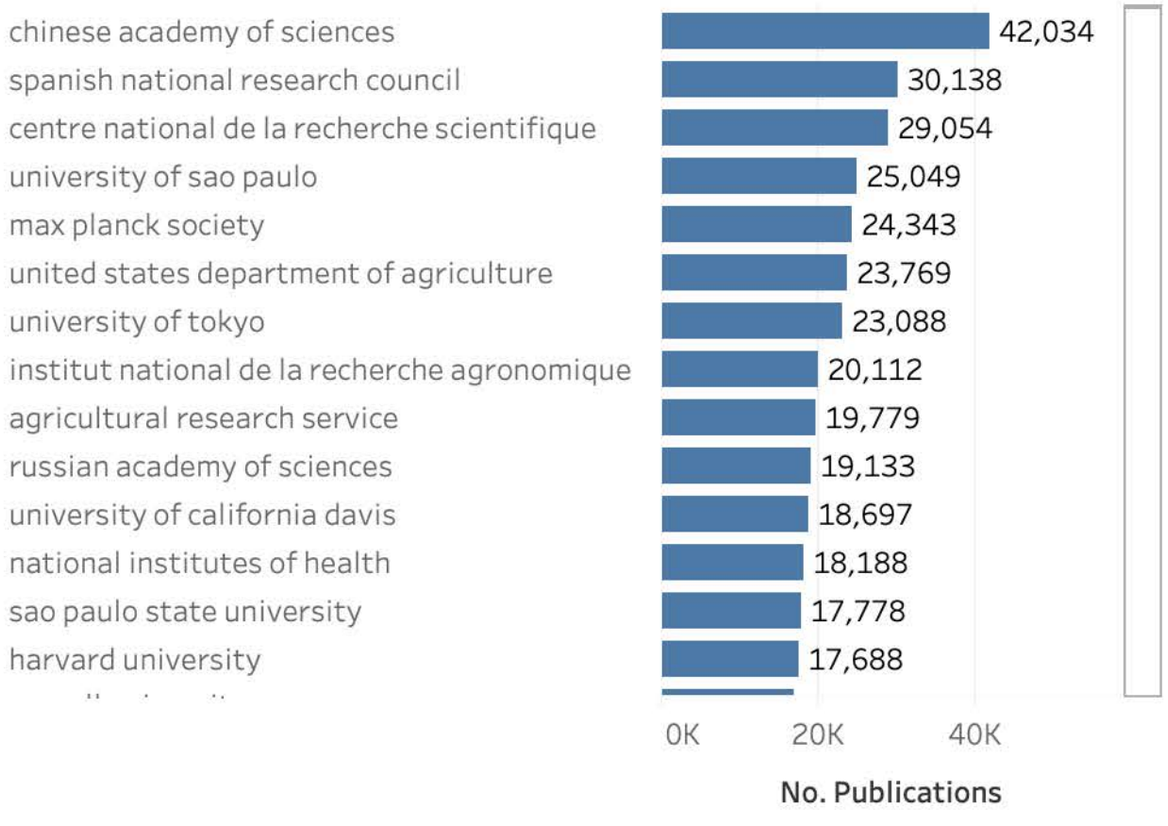
Trends



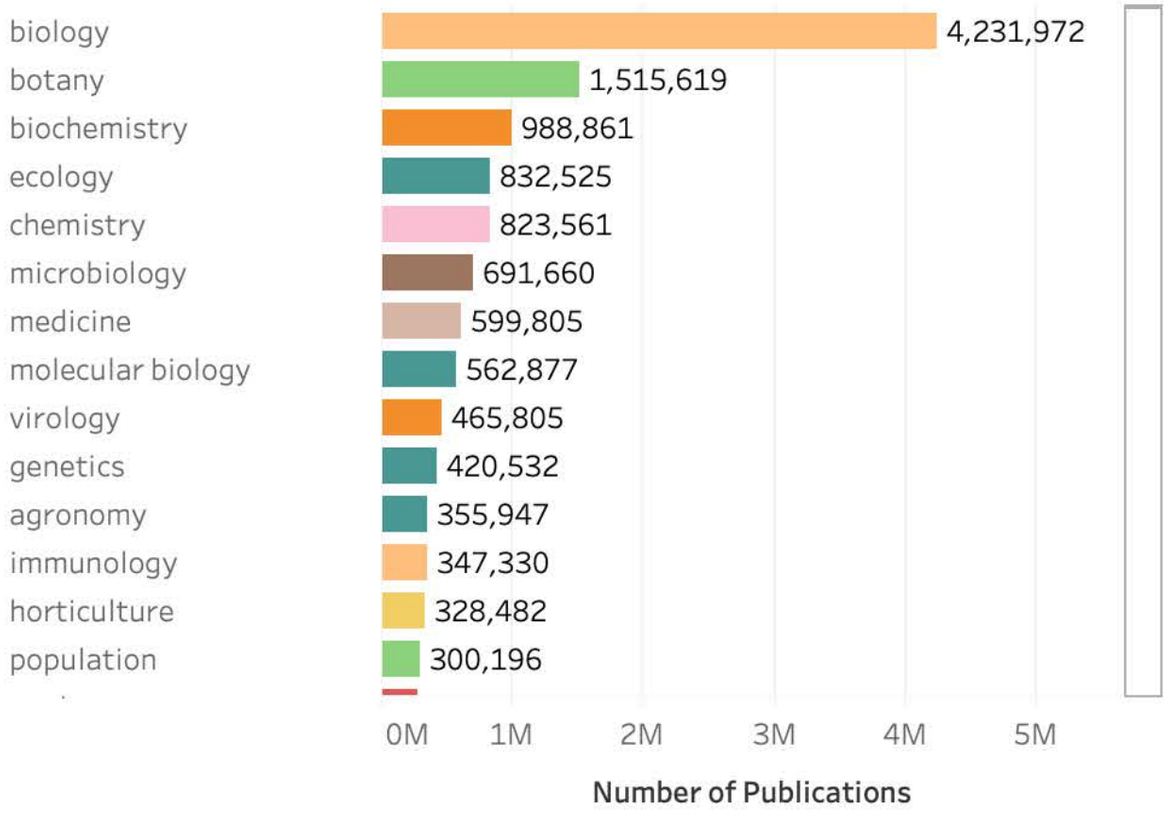
Species



Organisations



Fields of Study



Closing

Monitoring under the Nagoya Protocol allows us to do more than focus on compliance. It also allows us to improve our knowledge and understanding of biodiversity research and innovation on a global scale. That is a goal worth pursuing in thinking about the post-2020 vision to address the biodiversity and climate crisis.

Credits

- GBIF & Global Names Index
- Microsoft Academic Graph & Open Academic Graph
- Explosion AI for spaCy and Prodigy.