



NHM data portal & big(ish) open data: new frontiers for natural history collections

Vince Smith, Natural History Museum, London
Digitisation of Biological Collections, 14 April 2014

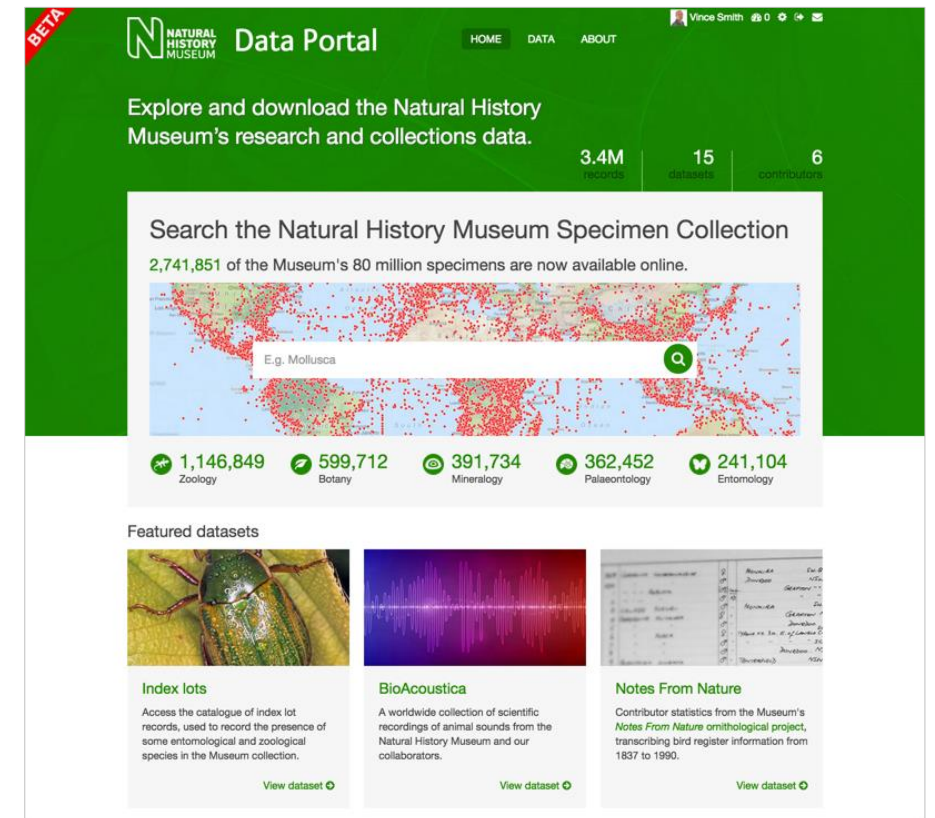
Why digitise natural history collections?

- Make collections accessible to global audiences
- Develop new ways to engage society with the natural world
- Create novel resources for education
- *Tackle major challenges for science and society*
 - *Origins and evolution of our solar system, earth and life*
 - *Environmental change*
 - *Health and disease*
 - *Food and agriculture*
 - *New sources of scarce minerals*
 - *Modelling the biosphere*

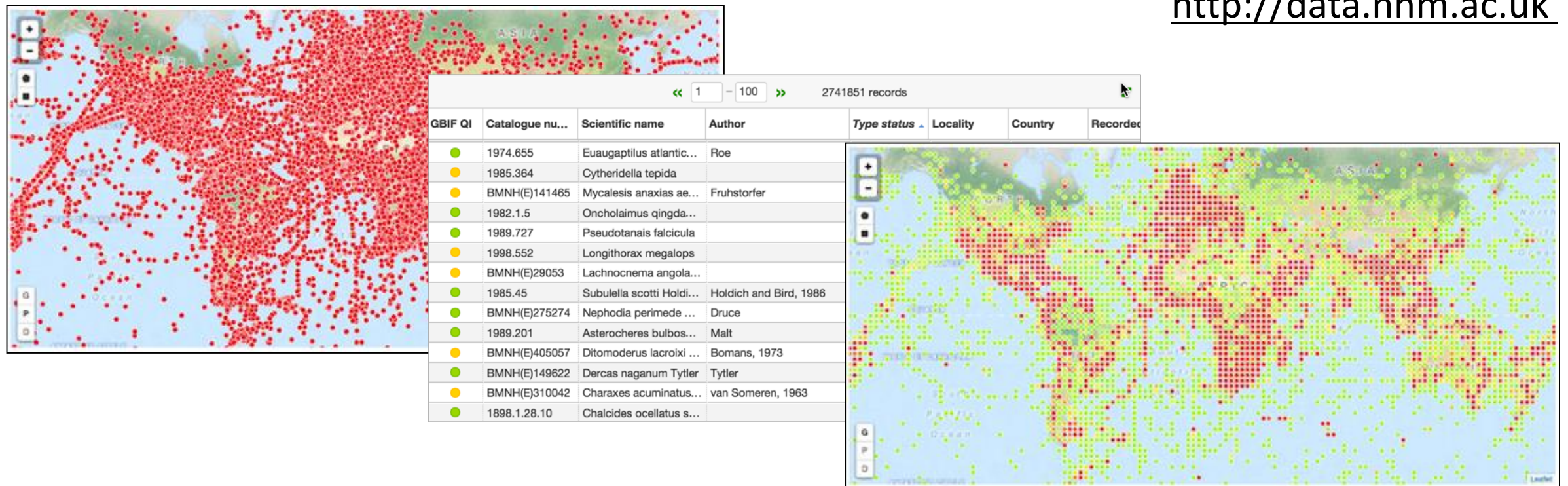


The NHM data portal

- A platform for deposition and discovery of NHM collections & data
- Promote innovation & collaboration through easy access & reuse of NHM data
- Integrates with our collection management system
- Handles heterogeneous datasets of NHM scientists
- Stable, citable (DataCite) identifiers on datasets & records to measure impact
- Technically sustainable



<http://data.nhm.ac.uk>



CKAN – the foundation for the the NHM data portal

- Enterprise, open source data portal platform
- Developed by Open Knowledge Foundation
- Widely used by governments, large organisations & academic communities
- Key features
 - Publish & find datasets
 - Store & manage data
 - Engage with users & others
 - Customise & extend



(Data)set discovery, metadata, usage & activity

NHM Portal

- Planned throughout 2013
- Build time circa 6 months
- Internal alpha Aug. 14
- Public beta Dec. 2014
- Handles multiple datasets
- Explore, download & cite
- Initial focus on collections data
- 3.6M records & weekly updates

Discovery

Metadata

Usage

Activity stream

Grid Map Statistics

Enter search text Configure fields & filters

Classification Location Collection event Identification Specimen Mineralogy Stratigraphy Meteorites Botany Zoology Palaeontology

Scientific name: Author: Kingdom:

Phylum: Class: Order:

Family: Genus: Subgenus:

Species: Subspecies: Higher classification:

Taxon rank:

☐ Has image ☐ Exclude centroids ☐ Has type ☐ Has lat/long

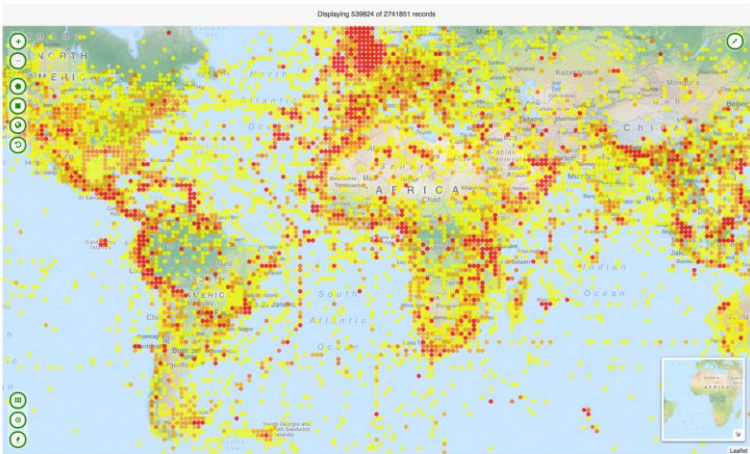
Display all classification fields: ☐

Extensive & fast search filters

Primary data views for each dataset



Point map



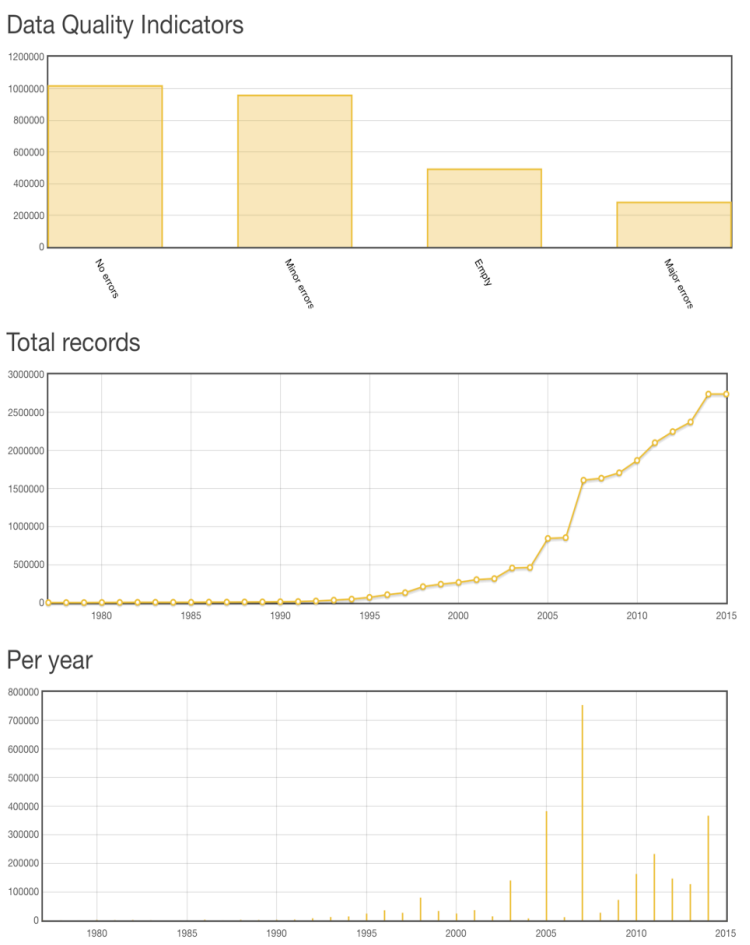
Grid map



Heat map

<div><div><div><div><<</div><div>1</div><div>-</div><div>100</div><div>>></div></div><div>2741851 records</div></div></div>														
	GBIF QI	Catalogue n...	Scientific name	Author	Type status	Locality	Country	Recorded ...	Collection...	Class	Family	Genus	Species	Subspecies
View		1980.604	Notasellus sarsi ...	Pfeffer		Hut Cove, ...			ZOO			Notasellus	sarsi	
View		BM000557703	Pteridophyte				Belize	Angela Hu...	BOT					
View		1944.1.8.395	Pustulopora pro...	Smitt					ZOO	Stenolaem...		Pustulopora	proboscid...	
View		BM000928531	Scorzoneroides ...	Moench				Ipse	BOT		Asteraceae	Scorzoner...	autumnalis	
View		PV M 94345	Morganucodon ...	Kühne, 1949		Pant 2 Qua...	Wales		PAL	Mammalia	Morganuc...	Morganuc...		
View		1969.1.6.5	Hippopetraliella ...	Cook, 1967			?Angola		ZOO	Gymnolae...		Hippopetr...	africana	
View		BM000509436	Appendicula co...	Ridl.			Indonesia	Collaborati...	BOT		Orchidaceae	Appendicula	congesta	
View		1852.10.4.76	Scatophagus sp						ZOO	Actinopter...	Scatophag...	Scatophagus	sp	
View		BM000528699	Leptactina Hook.f.	Hook.f.				George Lat...	BOT		Rubiaceae	Leptactina		
View		BM000036588	Microsorium me...	(D.Don) Ching			Philippines	Elmer Dre...	BOT		Polypodiac...	Microsorium	membrana...	
View		PM P 46148 (1)	Dentalina comm...			Charing	United Kin...		PAL			Dentalina	communis	
View		1965.6.22.90	Lycodes reticula...	Reinhardt, 1835		Barents Sea	Norway		ZOO	Actinopter...	Zoarcidae	Lycodes	reticulatus	
View		BM000778798	Aspidosperma				Brazil	Richard Sp...	BOT		Apocynac...	Aspidospe...		
View		1994.11.1.17...	Nemichthys sco...	Richardson, 1848		Cape Verd...	Cape Verd...		ZOO	Actinopter...	Nemichthy...	Nemichthys	scolopaceus	
View		BM000525861	Eulophiella elisa...	Linden & Rolfe			United Kin...	H.T. Pitt	BOT		Orchidaceae	Eulophiella	elisabethae	
View		1897.3.25.14	Dendoricella sc...	(Ridley)					ZOO			Dendoricella	schmidtii	
View		BM000077603	Goodyera proce...	(Ker Gawl.) Hook.			Nepal	Nathaniel ...	BOT		Orchidaceae	Goodyera	procera	
View		19601339	Phyllidia varicos...	Lamarck, 1801			India		ZOO			Phyllidia	varicosa	
View		PM PF 63641	Globorotalia (Tu...	Blow, 1979				Deep Sea ...	PAL		Globorotali...	Globorotalia	archeoco...	
View		BM000638359	Procris					Adolph Da...	BOT			Procris		
View		1935.4.27.200	Diurella bidens						ZOO			Diurella	bidens	
View		1940.3.4.14-15	Trachischium m...			Abor Coun...	India		ZOO		Colubridae	Trachischium	monitcola	
View		1972.11.9.8	Porcellidium bre...	Thompson and A...		eastern ed...		A G. Hume...	ZOO			Porcellidium	brevicauda...	
View		BM000026559	Paphiopedilum ...	Lindl.				M.T. Masters	BOT		Orchidaceae	Paphioped...	barbatum	
View		PV P 63743	Pneumatophorus			Selsey; Bra...	British Isles	R Fowler; F...	PAL			Pneumato...		
View		BM0005345...	Orchidaceae					Friedrich C...	BOT		Orchidaceae			
View		BMNH(E)403...	Metopodontus ...	Boileau, 1905			Laos		BMNH(E)	Insecta	Lucanidae	Metopodo...	biplagiatus	nigripes
View		1898.5.20.67...	Clausilia brunnea				Syria		ZOO			Clausilia	brunnea	
View		BM000561314	Polysiphonia mi...	Kütz.	Possible type		Italy	Friedrich Tr...	BOT		Rhodome...	Polysiphonia	minutula	
View		BM000926805	Ateramnus					Olof (Peter)...	BOT		Euphorbia...	Ateramnus		
View		RT Lowe 200...	Bromus diandru...	Roth			Portugal	Master - R...	BOT		Gramineae	Bromus	diandrus	

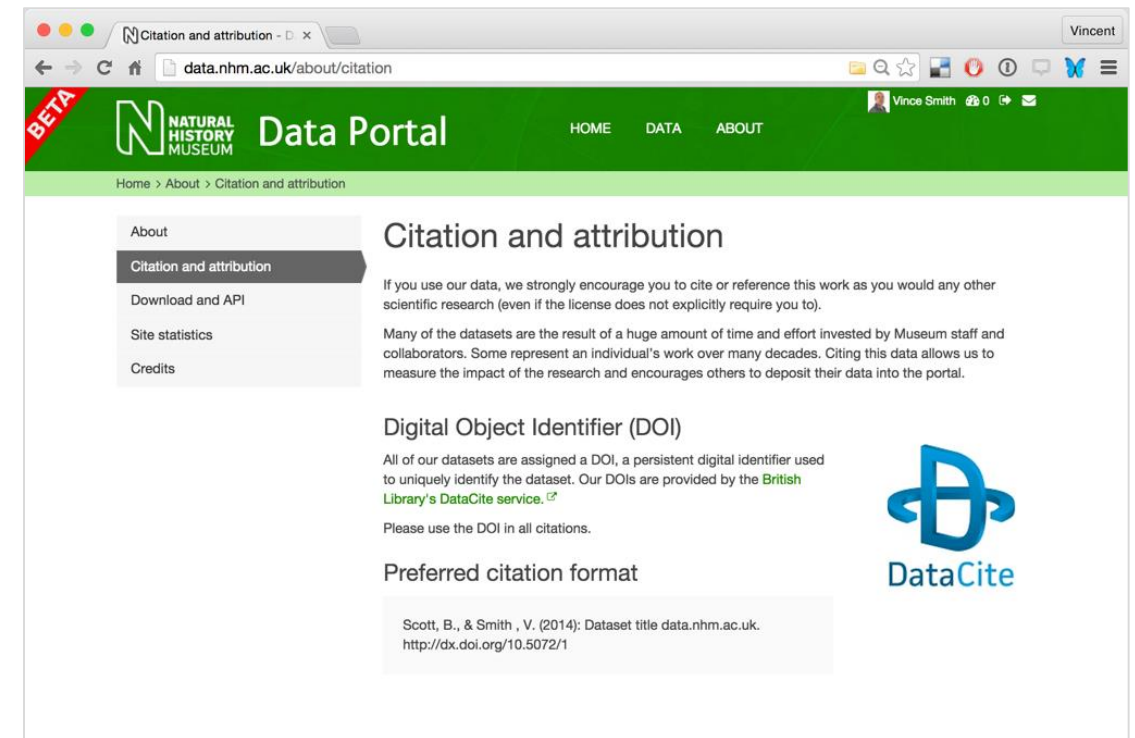
Filterable table



Statistical overview

Dataset & record citation

- DataCite DOIs on every dataset
- Stable URI (UUID) on every record
- Prior identifiers aliased & disambiguated
- Citation encouraged with clear statements at dataset & record level
- Allows us to track cited usage
- Dynamic DOI's on subsets coming soon



📄 Cite this dataset

Natural History Museum (2014):
Collection specimens
data.nhm.ac.uk.

<http://dx.doi.org/10.5519/0002965>

Retrieved: 16:06 08 Apr 2015 (GMT)

Additional Information

Cite as	http://data.nhm.ac.uk/specimen/9c96cbfd-8fb6-4420-9c34-3a670c77dda7
Created	June 20, 2003
Last updated	November 11, 2008
Format	dwc
License	Creative Commons CCZero

Data access & feedback

CKAN Data API

Access resource data via a web API with powerful query support. Further information in the [main CKAN Data API and DataStore documentation](#).

[Endpoints »](#)

The Data API can be accessed via the following actions of the CKAN action API.


Create	http://data.nhm.ac.uk/api/action/datastore_create
Update / Insert	http://data.nhm.ac.uk/api/action/datastore_upsert
Query	http://data.nhm.ac.uk/api/action/datastore_search
Query (via SQL)	http://data.nhm.ac.uk/api/action/datastore_search_sql


[Querying »](#)


Query example (first 5 results)

```
http://data.nhm.ac.uk/api/action/datastore_search?resource_id=05ff2255-c38a-40c9-b657-4ccb55ab2feb&limit=5
```

Extensive API

 Download

 Data API

 Contact dataset curator

The resource will be extracted, with current filters applied, and an email will be sent to your registered address shortly.

Send

Cancel

DwCA Downloads

Contact form

* Your name:

* Your email:

Contact type:

Dataset: Collection specimens

Resource: Specimens

* Your request:

You can use [Markdown formatting here](#)

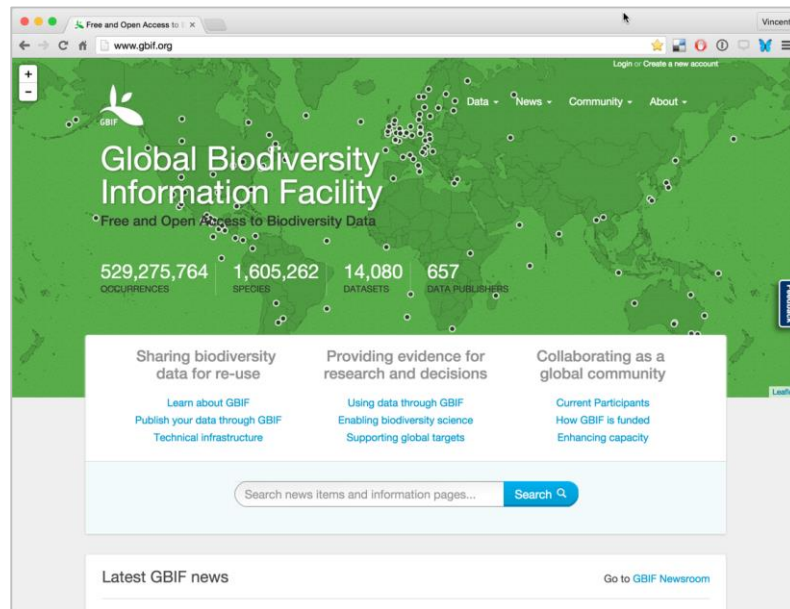
* Required field

Link to data curator team

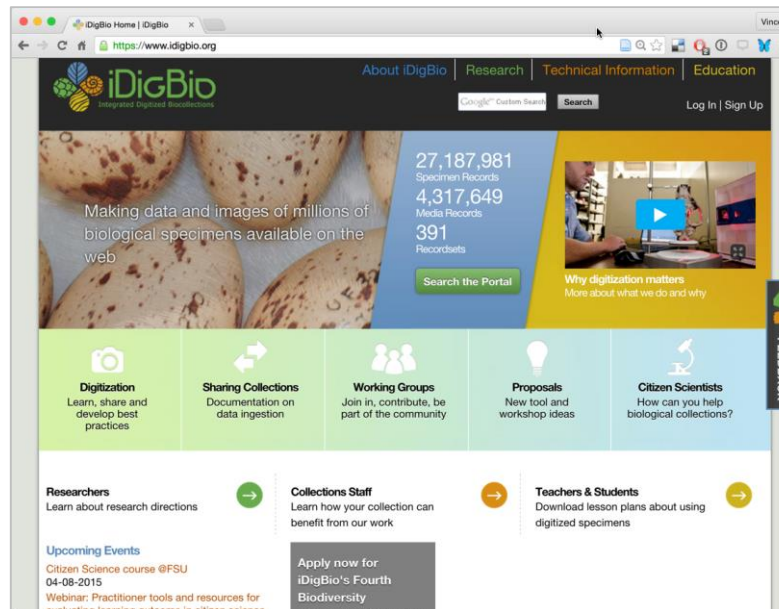


R integration

Serving external data portals



GBIF



iDigBio



EOL

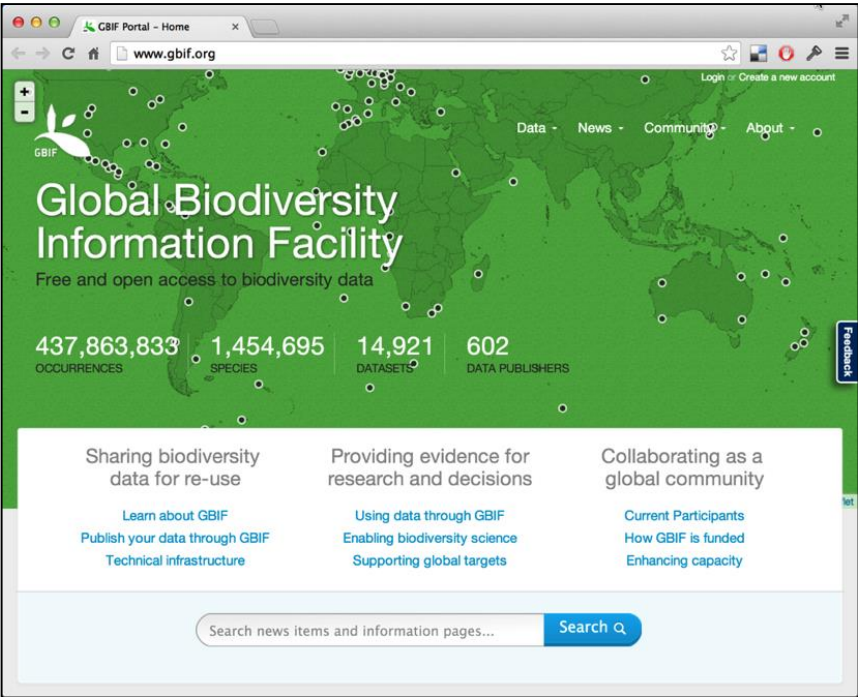


Vertnet



CRIA

Traffic light data quality indicators

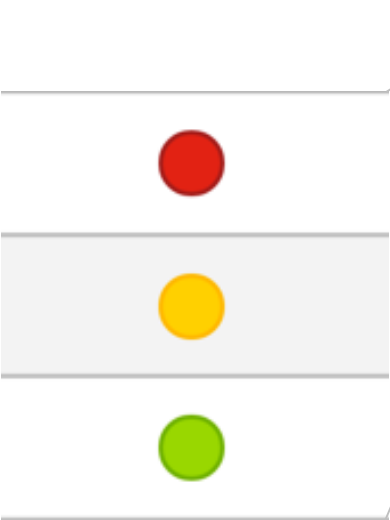


Via GBIF API

Major errors

Minor errors

No errors



GBIF QI	Catalogue nu...	Scientific name	Author	Type status
Yellow	RT Lowe 2000 ...	Bromus diandrus Roth	Roth	
Yellow	BM001147086	Frullania microphylla (...)	(Gottsche) Pearson	
Green		Daphnusa ocellaris W...	Walker, 1856	
Yellow	BM000559415	Capsicum annum (D...	(Dunal) Heiser & Pickersgill	
Green	1949.1.19.30	Crossaster papposus...	(Linnaeus, 1767)	
Yellow	PM P 43052 (2)	Orbitolina birmanica ...	Sahni, 1937	
Yellow	BM000798867	Chasalia kolly (K.Sch...	(K.Schum.) Hepper	Isotype
Green	BM000003217a	Salix arctica Pall.	Pall.	
Yellow	Carlos Types - ...	Polystichum viviparu...	Fée	Isotype
Yellow	1998.3.12.1-50	Neolepidapedon sp.		
Green	50021	Solanum galapagens...	S.C.Darwin & Peralta	
Red	PM OS 16045			
Yellow	BMNH(E)70713	Mellicta athalia		
Green	1974.1.25.142	Rhodeus suigensis M...	Mori, 1914	

Potential errors highlighted & “corrected”

Normal view

Darwin Core view

GBIF view

Back to resource

Download

Contact record curator

BM000925118

Classification

Scientific name:

Wahlenbergia parvifolia (P.J.Bergius) Lammers

Author:

(P.J.Bergius) Lammers

Family:

Campanulaceae

Genus:

Wahlenbergia

Species:

parvifolia

Higher classification:

Campanulaceae

Location

Label locality:

Hort. Kew.

Country:

United Kingdom

Continent:

Europe

Higher geography:

Europe; United Kingdom

Decimal latitude:

54.155348

Decimal longitude:

-2.109

Verbatim latitude:

54 09 19.25 N

Verbatim longitude:

002 06 32.40 W

Centroid:

True

Collection event

Recorded by:

Anon.

Record number:

S.N.

Last parsed

February 16, 2015. 13:55:47

GBIF quality indicators: MINOR ERRORS

- Coordinate rounded
- Geodetic datum assumed WGS84

GBIF quality indicators: MINOR ERRORS

- Coordinate rounded
- Geodetic datum assumed WGS84

Potential errors highlighted & “corrected”

Normal view

Darwin Core view

GBIF view

Back to resource

Download

Contact record curator

GBIF interpreted record

Identification details	According to GBIF Backbone Taxonomy
Identified as species	Wahlenbergia parvifolia (P.J.Bergius) Lammers
Taxonomic classification	Plantae > Magnoliophyta > Magnoliopsida > Asterales > Campanulaceae > Wahlenbergia

Occurrence details

Recorded by	Anon.
Record number	S.N.

Location

Country	United Kingdom
Coordinates	-2.109, 54.15535
Geographic classification	EUROPE > United Kingdom

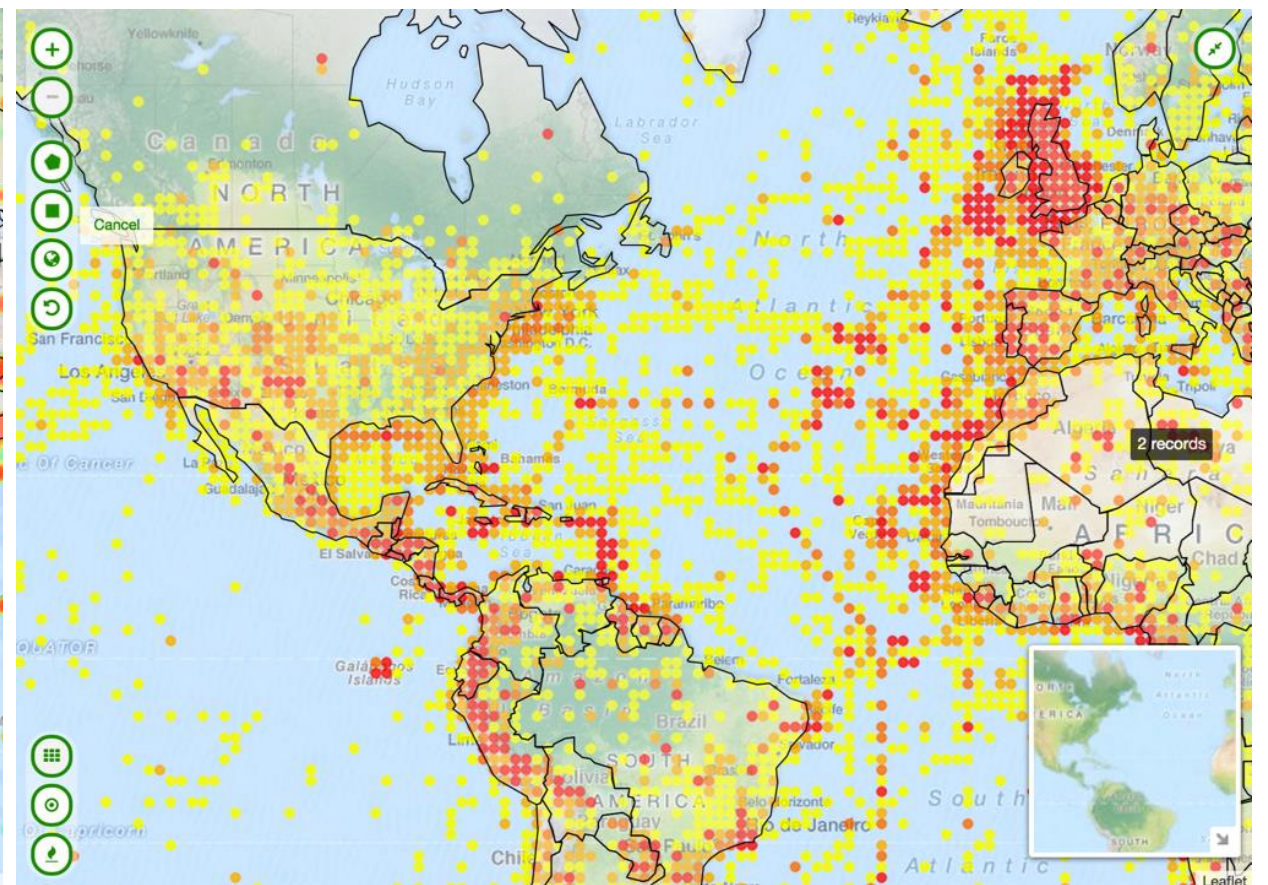
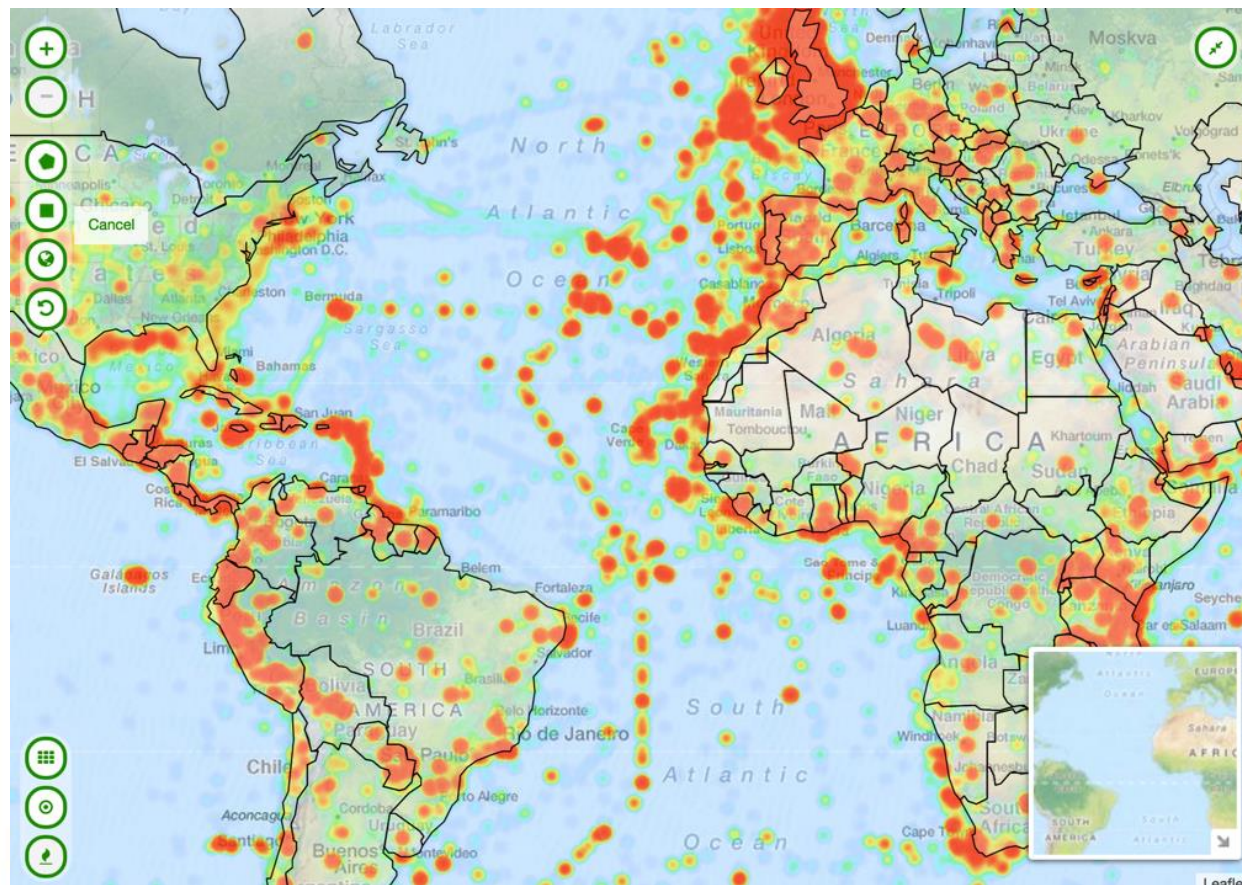
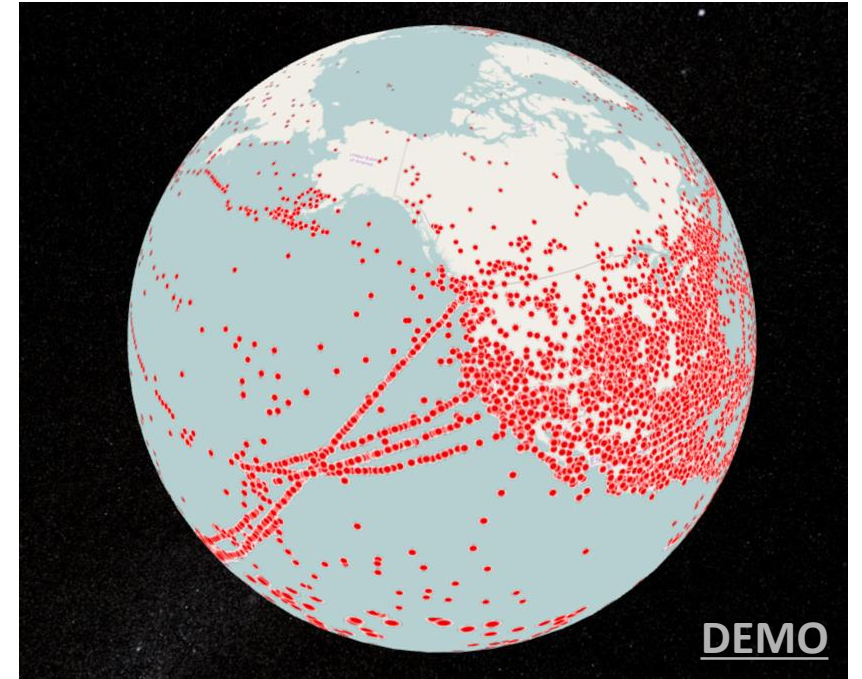
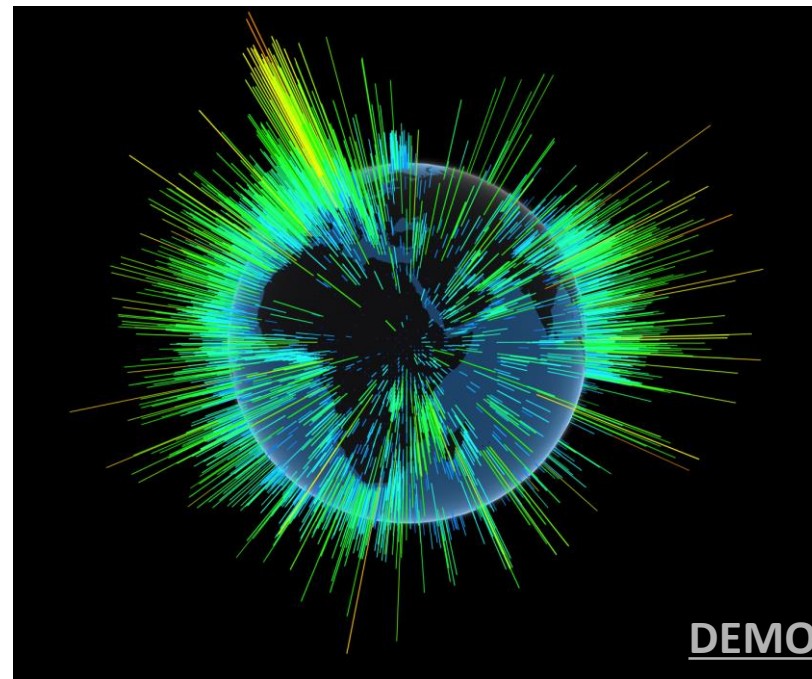
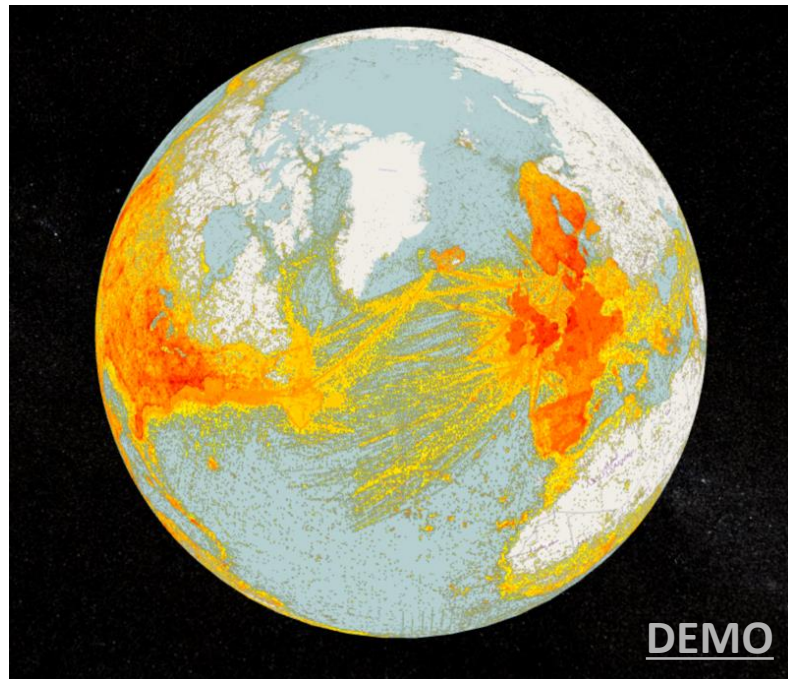
Data Quality Issues

Coordinate rounded	Original coordinate modified by rounding to 5 decimals.
Geodetic datum assumed WGS84	Indicating that the interpreted coordinates assume they are based on WGS84 datum as the datum was either not indicated or interpretable.

GBIF Publication details

Last interpreted	February 16, 2015. 20:42:22
Last parsed	February 16, 2015. 13:55:47

Data visualisations (embedded & via API)



Easy addition of new datasets

Quick & semi-automated workflow

1. Name the dataset

Name

Please name your data record.

Name:

Unique identifier for this data record:

3. Describe the data file

Name

Data Files

Description

Please add a description of the data set. What does the data contain/show? Why was it produced? (It is often displayed with the package title. In particular, it should start with a short sentence that describes the data set succinctly, because the first few words alone may be used in some views of the data sets. Here is the place to state if there are any limitations or deficiencies to the data in order to enable users to evaluate the information; even incomplete data may be adequate for some users).

Municipal waste generation in England from 2000/01 to 2009/10 tables 1-5

Licence

Publisher

Theme & tags

Additional resources

Temporal coverage

Geographic coverage

Extras

5. Add additional resources

Name

Data Files

Description

Licence

Publisher

Theme & tags

Additional resources

You can add links to additional resources here. Use a link for each resource. This could be PDFs of files explaining the data or links to a website with more information.

Description: Link: Format:

Mandate:

An internet link to the enabling legislation that serves as the mandate for the collection or creation of this data, if appropriate. For example Public Record Act s.2 would be: <http://www.legislation.gov.uk/id/ukpga/Eli2/6/7/51/section/2> This should be taken from The National Archives' Legislation website, and where possible be a link directly to the relevant section of the Act.

7. Geographic coverage

Name

Data Files

Description

Licence

Publisher

Theme & tags

Additional resources

Temporal coverage

Geographic coverage

☒ England ☐ Scotland ☐ Wales ☐ Northern Ireland ☐ Overseas ☐ Global

Where a dataset covers multiple areas, the system will automatically group these (e.g. 'England', 'Scotland' and 'Wales' all being selected would be shown as 'Great Britain').

2. Upload / link the data file

Name

Data Files

In this tab you enter the links to the files containing the actual data at the heart of this record.

Your record may be creating a single file record (a record which points to just one file) or you may be creating a timeseries record (where multiple files are released over time, for example the £25k spend data, where a new data file is added to the record monthly).

This record is:

☒ A single file record ☐ A timeseries record

If your data comes in multiple formats, use a line for each.

NB: Do not put links to web pages here. These and other useful documents that are not actually the data should go in the 'Additional Resources' section.

File Title: URL: Format:

Format: This should give the file format in which the data is supplied. You may supply the data in a form not listed here, but still constrained by the Public Sector Transparency Board's principles that require that all data is available in an 'open and standardised format' that can be read by a machine. Data can also be released in formats that are not machine-processable (e.g. PDF) alongside this.

Format choices: CSV | RDF | XML | XBRL | SDMX | HTML+RDFa | Other as appropriate

4. Theme & tag

Name

Data Files

Description

Licence

Publisher

Theme & tags

Please select a primary theme...

2. If there are other themes in this data, select them here:

☐ Health ☐ Environment ☐ Education ☐ Finance ☐ Society ☐ Defence ☐ Transportation ☐ Spending data ☐ Government

3. Please provide additional tags for this data record to help users find and browse between related data records. Just type and click enter to add more tags:

Tags:

6. Temporal coverage

Name

Data Files

Description

Licence

Publisher

Theme & tags

Additional resources

Temporal coverage

Temporal coverage: - or 07:45 31/03/2006

If available, please indicate the time as well as the date. Where data covers only a single day, the second box can be left blank.

Geographic coverage

Extras

8. Save & finish

Important: By submitting content, you agree to release your contributions under the terms & conditions of the site. Please refrain from editing this page if you are not happy to do this.

Heterogeneous non-collection datasets

ZooKeys 481: 131–138 (2015)
doi: 10.3897/zookeys.481.8788
<http://zookeys.pensoft.net>

RESEARCH ARTICLE

Imp: The customizable LEGO® Pinned Insect Manipulator

Steen Dupont¹, Benjamin Price¹, Vladimir Blagoderov²

¹ Department of Life Sciences, Natural History Museum, London, SW7 5BD, U.K. ² Department of Science Facilities, Natural History Museum, London, SW7 5BD, U.K.

Corresponding author: Steen Dupont (steen.dupont@nhm.ac.uk)

Academic editor: S. Winterton | Received 21 October 2014 | Accepted 19 January 2015 | Published 4 February 2015
<http://zoobank.org/EE66CFD1-239C-4316-840F-9B4B85BF750B>

Citation: Dupont S, Price B, Blagoderov V (2015) Imp: The customizable LEGO® Pinned Insect Manipulator. ZooKeys 481: 131–138. doi: 10.3897/zookeys.481.8788

Abstract
We present a pinned insect manipulator (IMp) constructed of LEGO® building bricks with two axes of movement and two axes of rotation. In addition we present three variants of the IMp to emphasise the modular design, which facilitates resizing to meet the full range of pinned insect specimens, is fully customizable, collapsible, affordable and does not require specialist tools or knowledge to assemble.

Keywords
Specimen Manipulator, Entomology, Stage, Digitization, Imaging

Introduction
Natural history collections are one of the most fundamentally important institutions in science, where voucher specimens are housed in perpetuity, embodying the research of generations of scientists. Within entomology collections insects were historically preserved by drying on an appropriately sized pin. Today insect groups are preserved in ethanol, mounted on microscope slides or in paper or plastic envelopes, but the dry pinned method has not changed for the majority of insect orders since its development in the early 18th century. This has resulted in millions of pinned insect specimens housed in natural history collections globally, for example the pinned insect collection

Copyright Steen Dupont et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

doi: 10.3897/zookeys.481.8788

Imp: The customizable LEGO® Pinned Insect Manipulator

Steen Dupont, Benjamin Price, Vladimir Blagoderov (2015): Imp: The customizable LEGO® Pinned Insect Manipulator (Annotated building instructions) [data.nhm.ac.uk](http://dx.doi.org/10.5519/0038449)
<http://dx.doi.org/10.5519/0038449>
Retrieved: 16:08 08 Apr 2015 (GMT)

License
Creative Commons Attribution
[OPEN DATA](#)

Follow dataset
Followers 0
[Follow](#)
Follow this dataset to receive update notifications.

Social
[Google+](#)
[Twitter](#)
[Facebook](#)

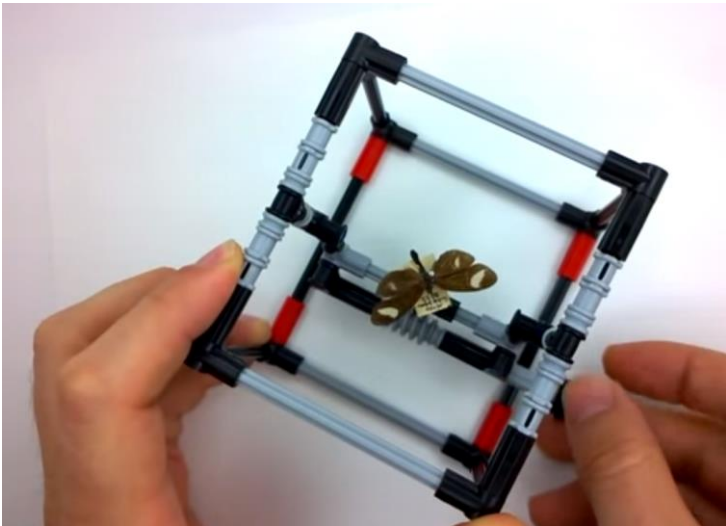
Dataset Metrics Activity Stream [Contact dataset curator](#)

Imp: The customizable LEGO® Pinned Insect Manipulator (Annotated building instructions)
Dupont et al. 2015. Imp: The customizable LEGO® Pinned Insect Manipulator. Supplement.

Data and Resources
[Supplement 1: Annotated building instructions](#) [View](#) [Download](#)
[Supplement 2: Video demonstration](#) [View](#) [Go to resource](#)

Digitization Entomology Imaging Specimen Manipulator Stage

Additional Info
Author(s) Steen Dupont, Benjamin Price, Vladimir Blagoderov
Dataset category Research
Last updated February 5, 2015
Created January 15, 2015
License Creative Commons Attribution



Assembly Video

Imp: The customizable LEGO® Pinned Insect Manipulator
Steen Dupont*, Benjamin Price*, Vladimir Blagoderov**
* Department of Life Sciences, Natural History Museum, London, SW7 5BD, U.K.
** Department of Science Facilities, Natural History Museum, London, SW7 5BD, U.K.

Supplement 1: Annotated building instructions

Table of Contents

Micro-IMP and parts list	2
Micro-IMP building instructions	3-14
IMP and parts list	15
IMP building instructions	16-30
Open-IMP and parts list	31
Open-IMP building instructions	33-46
Giant-IMP and parts list	47
Giant-IMP building instructions	59-70
Parts list for all IMP models	70-73

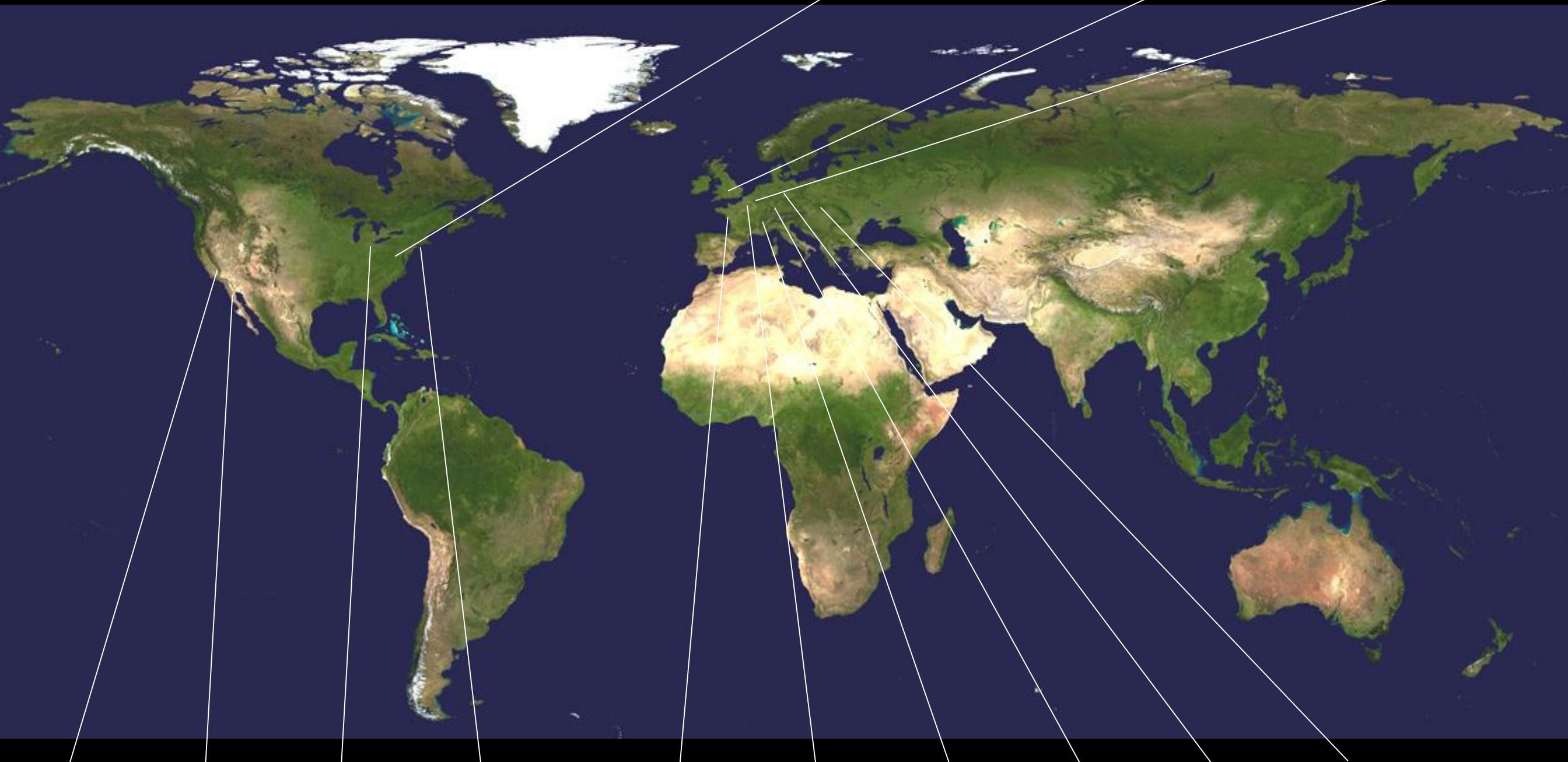
Step-by-step instructions

1.5-3 BILLION SPECIMENS
1.9 million species
300 years of collection

Washington
125M

London
80M

Paris
60M



San
Francisco
28M

Los
Angeles
35M

Chicago
25M

New
York
30M

Brussels
37M

Leiden
37M

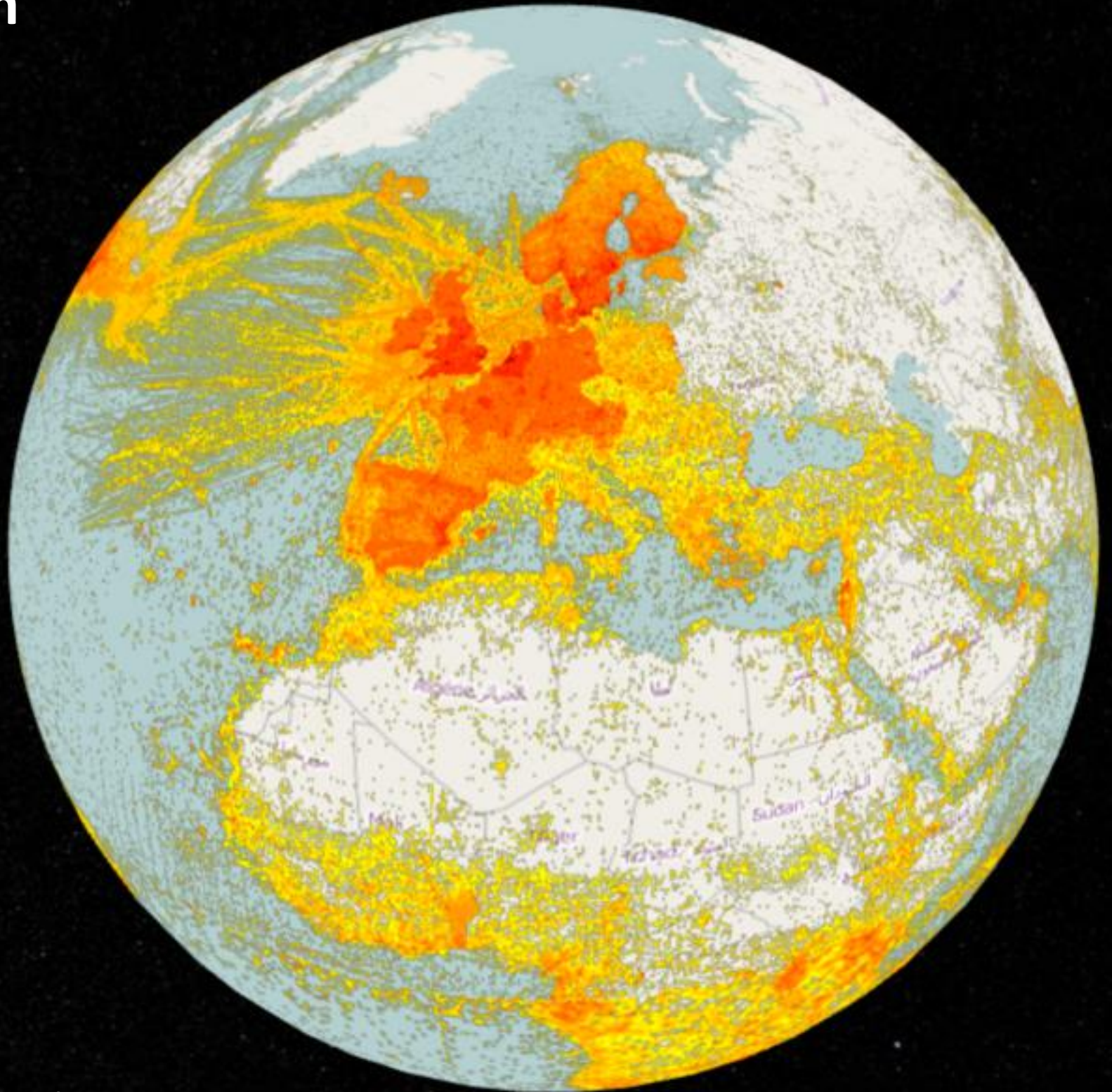
Vienna
35M

Frankfurt
40M

Berlin
30M

St
P'berg
32M

1.5-3 BILLION SPECIMENS
1.9 million species
300 years of collection



**Use NH collections to explore
changes over space & time
Correlate with land use changes**

- **Natural**
- **Semi-natural**
- **Human dominated**

Goals

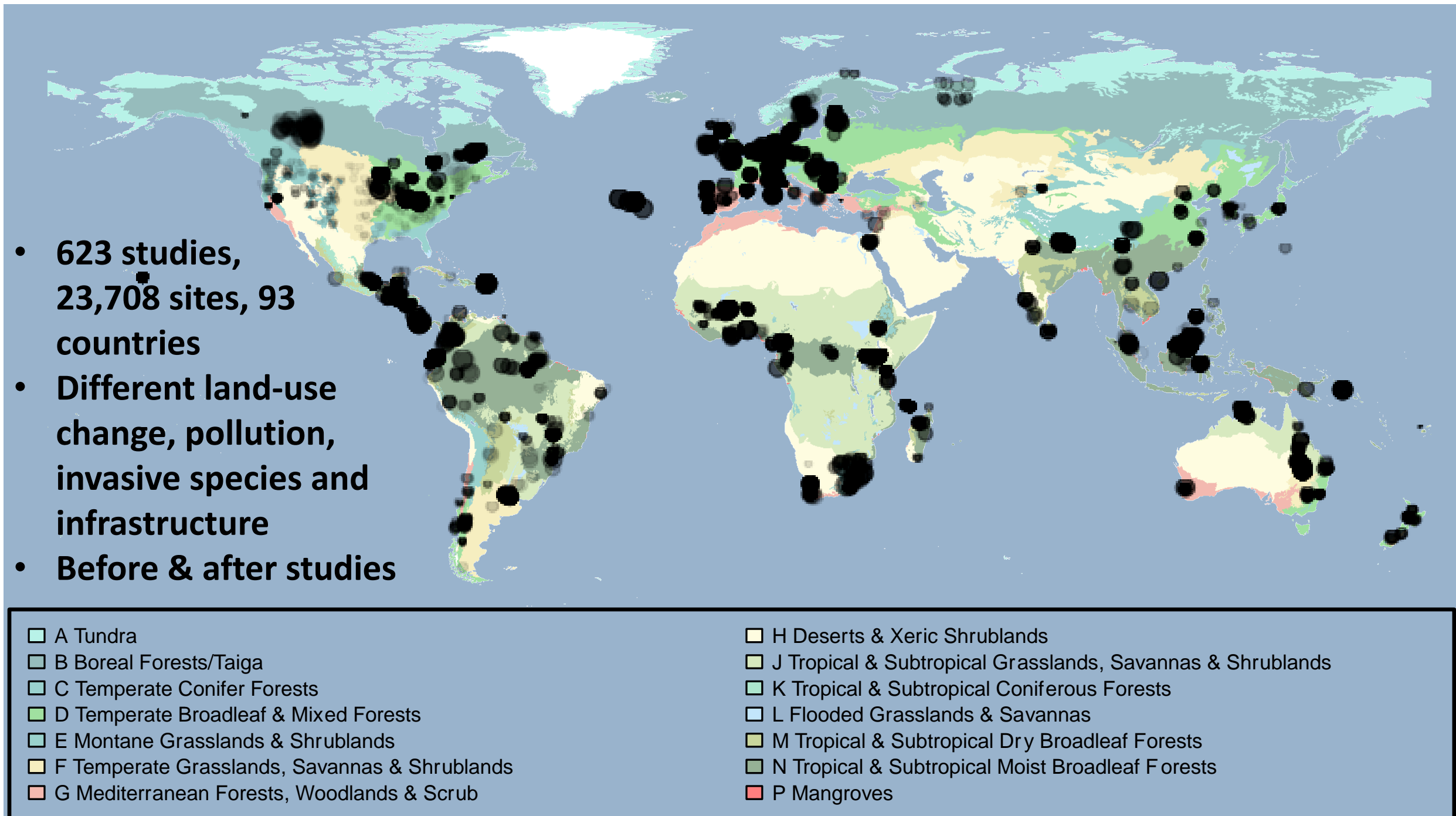
Quantify human impacts

Predict how to mitigate human impacts

PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems)

Meta-analytic approach to investigate how local biodiversity typically responds to human pressures
Improve our ability to predict future biodiversity changes

- 623 studies, 23,708 sites, 93 countries
- Different land-use change, pollution, invasive species and infrastructure
- Before & after studies

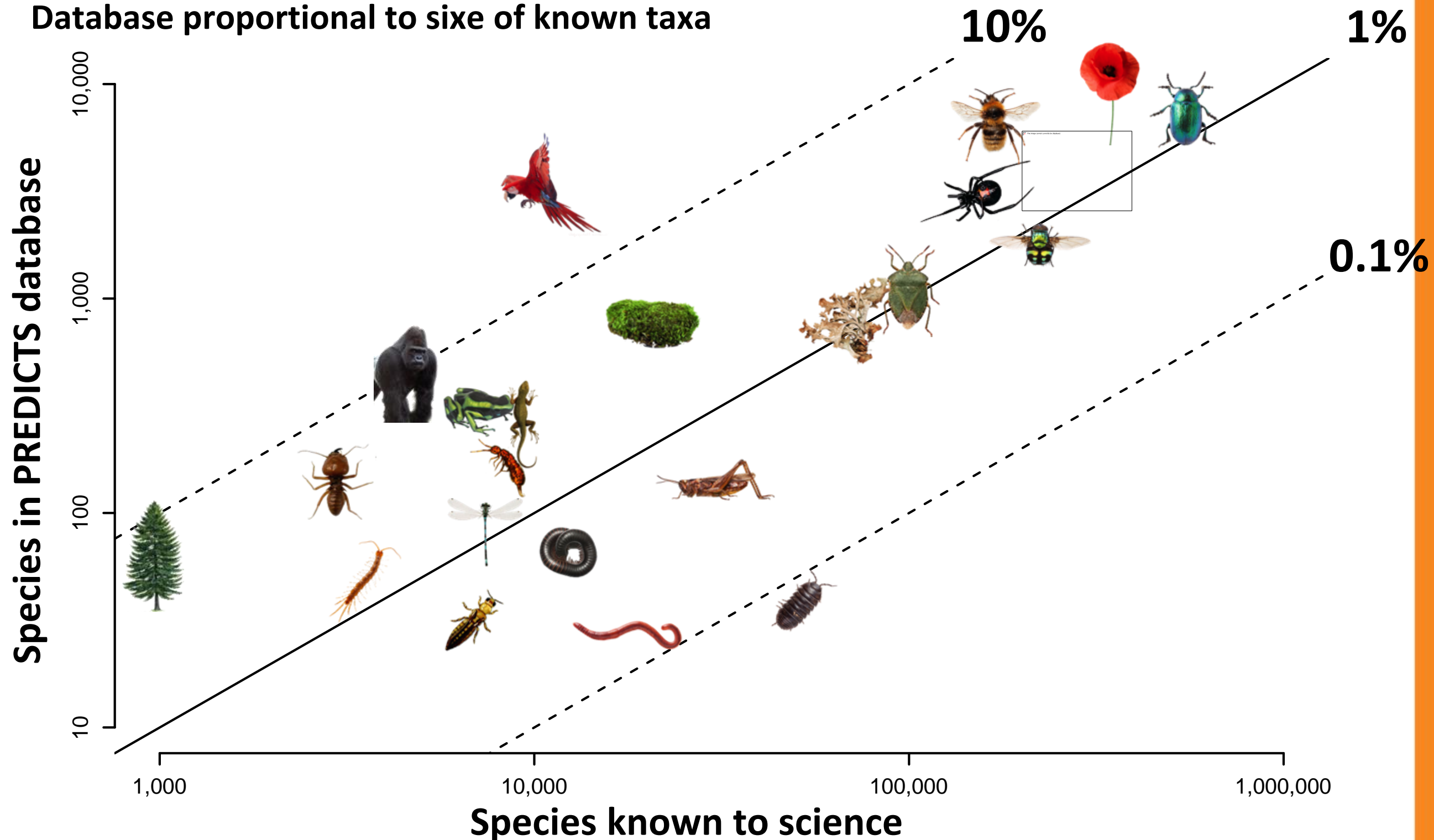


Hudson et al. doi: 10.1002/ece3.1303

PREDICTS taxon representation (a model of the natural world)

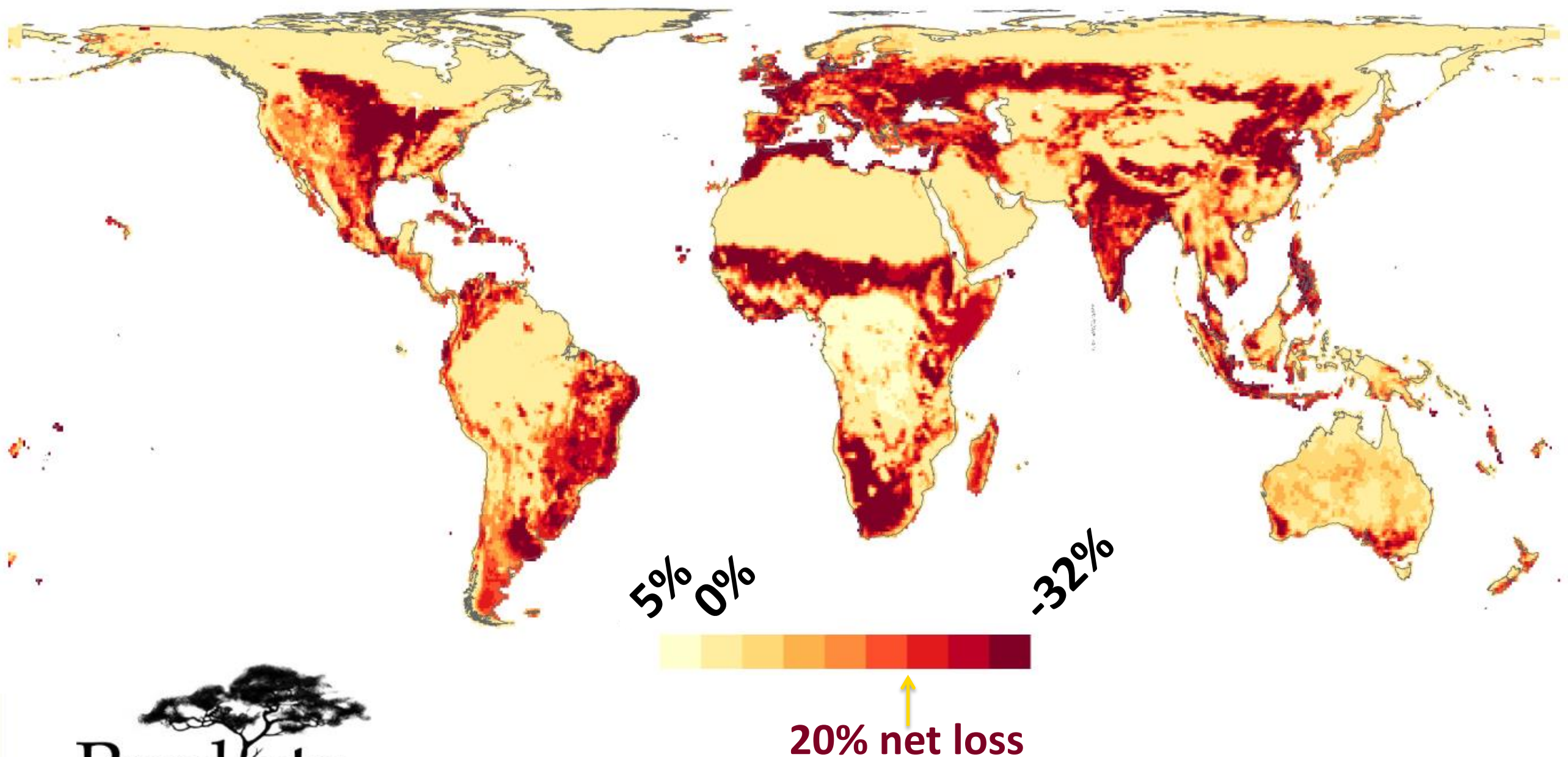
Coverage of 41,443 species

Database proportional to size of known taxa

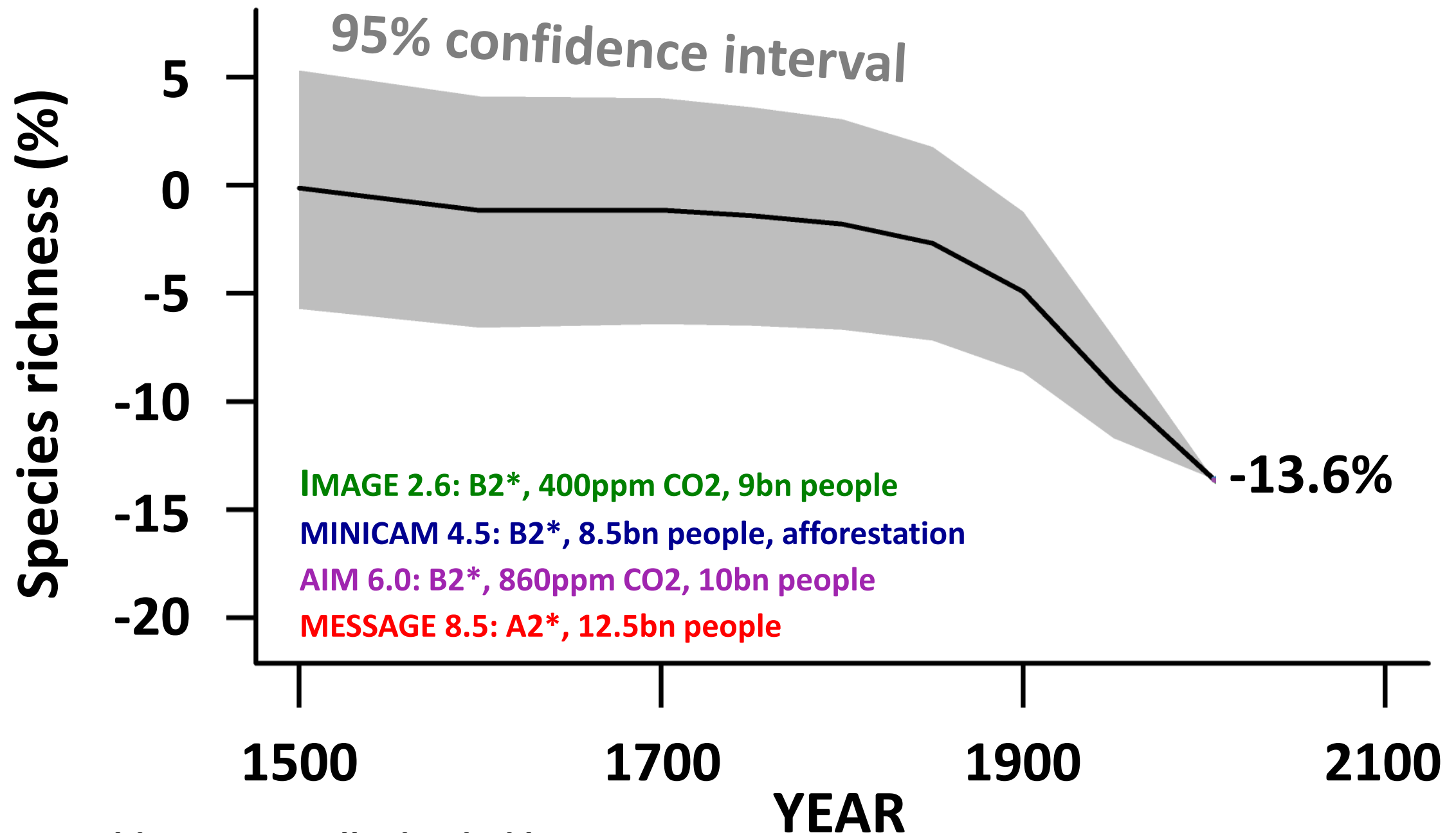


PREDICTS data modelling species richness

Species richness lost by 2005



Predicting trends in biodiversity using natural history collections



* B2 = world economically divided but ecologically friendly

A2 = still divided, not ecofriendly

Newbold, Hudson et al, Nature doi:10.1038/nature14324

