The changing intellectual model of natural history institutions

Vincent S. Smith
@NHM_Digitise

Helsinki, Finland
28th Sept. 2017

NATURAL HISTORY MUSEUM

BMNH, established *"for the learned & the curious"*

Collections with purpose

Collections that change how we think

Collections that inspire the next generation

Collections at scale

NHM Collection – 2015/16

Dept. / Division	Collection size	% specimens databased	# sci. visitor days	# loaned specimens
Earth Sciences				
Mineral & Planetary Sciences	212,600	95%	165	52
Envt & Economic Earth Sciences	212,000	95%		51
Invertebrate & Plant Palaeobiol.	7,000,000	4.9%	705	183
Vert Palaeobiol & Anthropology	1,064,000		702	1,432
Life Sciences				
Algae, Fungi & Plants	6,369,000	11%	415	1,298
Insects	32,013,000	<1%	1,565	28,493
Vertebrates	4,282,000	4.2%	2,932	716
Invertebrates	24,403,000		828	559
TOTAL	**76,101,600**	**4.50%**	**7,303**	**32,784**

New opportunities

Limited potential to grow

NHM Digital Collections Programme (2015-2025)

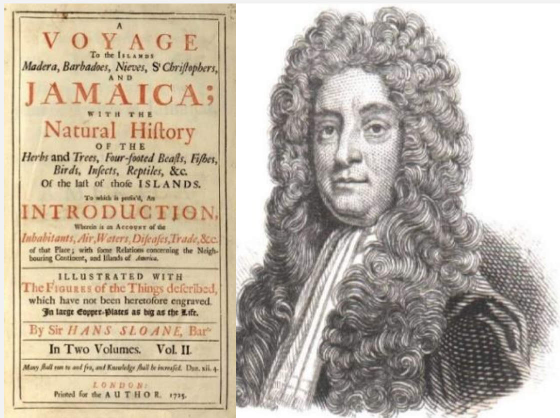
“To collate, organise and make available to global scientific & public audiences one of the world’s most important natural history collections”

‘Ambition’ to digitise 20 million specimens

Digitisation Objectives for 2, 5 & 10 yrs.

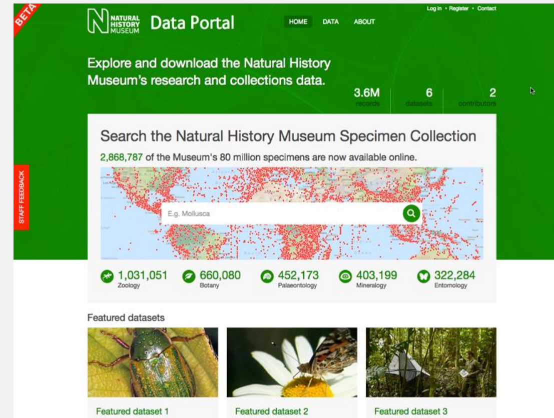
	2 Year	5 Year	10 Year
POLICY & PROTOCOL	Defined data policy & standards	Policies embedded in operating practices	World leaders in digital curation
DATA CAPTURE	Prioritised digitisation workflows piloted	Portfolio of mass digitisation projects	Some major collections digitised
PEOPLE & SKILLS	Task force formed and operating	Best practice processes integrated into training	Digital curation core part of BAU practice
INFRASTRUCTURE	Refined collections database, tools & hardware	Future collections database implemented	Broad connections to other large digital collections
STAKEHOLDERS & GOVERNANCE	Key user communities engaged	Peer-to-peer development	Proactive engagement of emerging audiences
PARTNERSHIPS	Partners involved in pilot projects	Fully-funded digitisation portfolio	Major international coalitions
RESEARCH	Research-orientated projects and initiatives	Collaborative research material published	Major contributions to grand challenges
ACCESS	Live NHM Data Portal	Tools, visualisations & analytics	Integrated global network of users

Pilot Digitisation Projects



Age of Enlightenment

Crop Wild Relatives



Informatics

Environmental Change

Micro Collections

Open Herbarium

Environmental Change

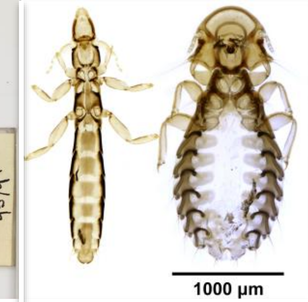
- UK butterflies & moths
- 800k specimens
- 2 mins per specimen
- £1 per specimen

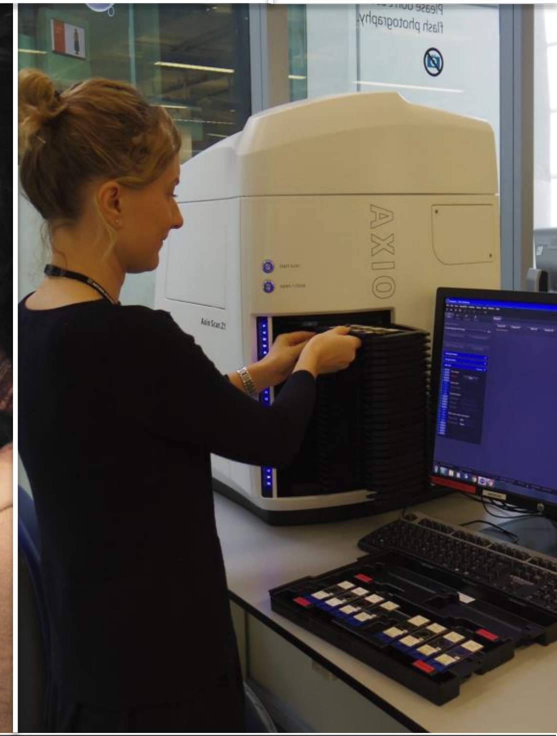
Large-scale digitisation:

- High-throughput digitisation workflows
- Informatic pipelines
- Computer-assisted object recognition

Parasites & Vectors:

- High-throughput microscope slide digitisation
- Low cost primary setup, high throughput
- A second attempt at slides (low tech.)



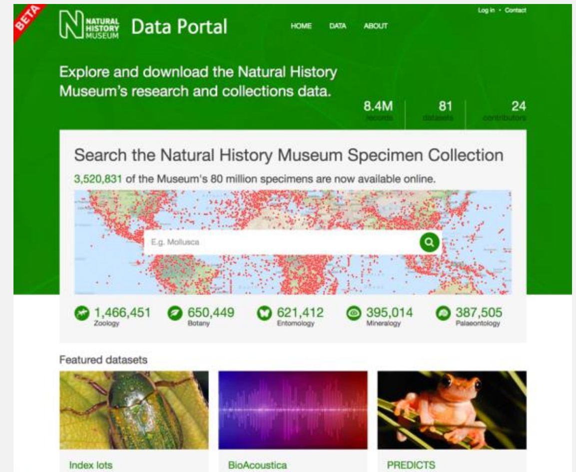


Parasitic lice (new spp. & cospeciation)

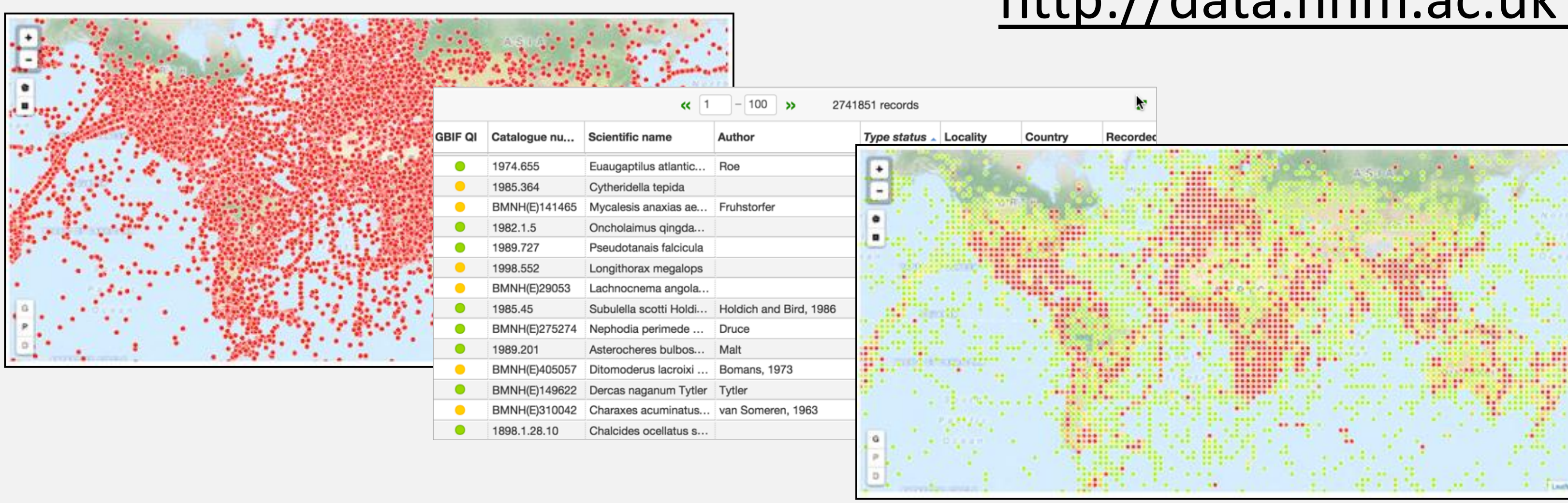
- Circa 70-80k slides (20k to date)
- 600-1k slides per day pp. (£0.07-0.2 ea.)
- Twin tracks (low & high resolution)
- Setup, imaging, transcription, Q.C. barcoding, databased (but not research)

Data Access: NHM Data Portal

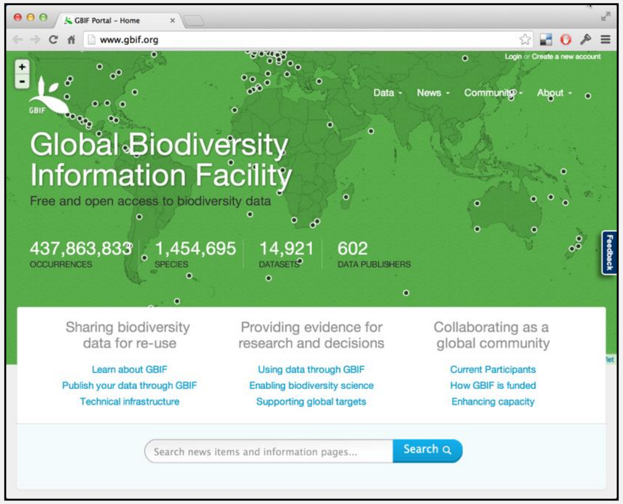
- Discovery of NHM collections & research data
- Easy access & reuse to promote collaboration (website, API, R-package, RDF & download)
- 8.3m data records (+ images, sound, video & 3D)
- Traffic light data quality indicators
- Stable, citable identifiers
- Default open licensing (CC-Zero, CC-BY)
- >2.3Billion records downloaded



http://data.nhm.ac.uk



Services supporting data quality indicators



Via GBIF API

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

Major errors

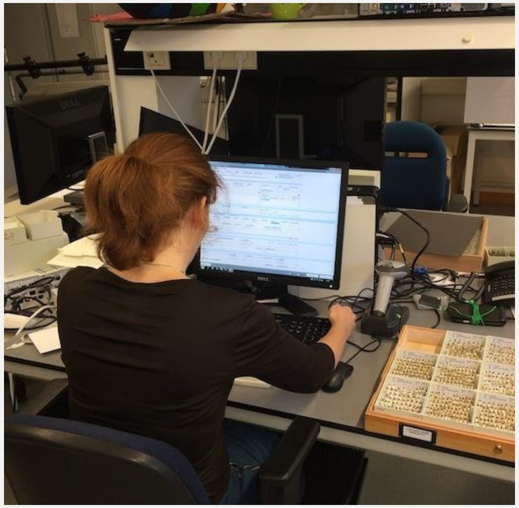
Minor errors

No errors

nb. similar services offered by CRIA for Brazilian data

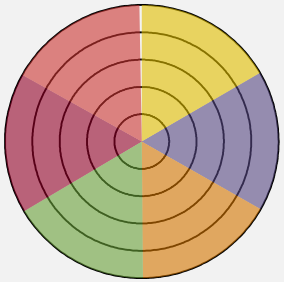
Industrial workflows

Possible for some collections, but just part of the story



Creating a sustained digital culture

Normalising ‘digital’ within NHM London



Prioritization

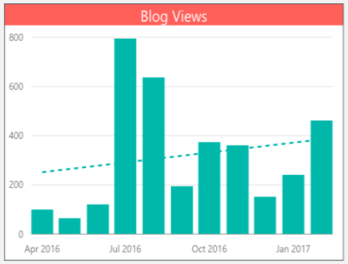
Outreach

Policy

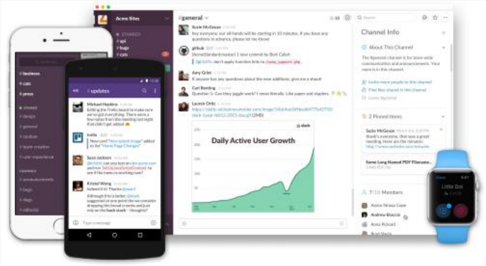
Government

Training

Events

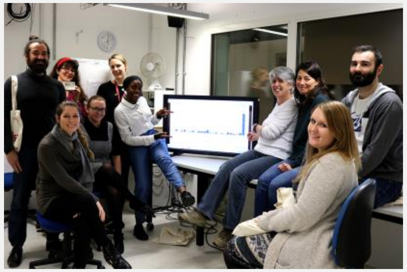


Monitoring & KPIs

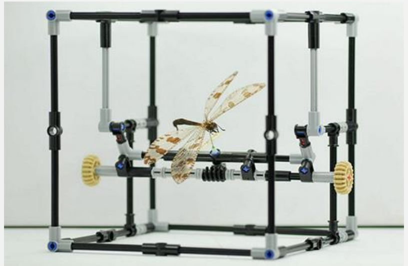


Team communication

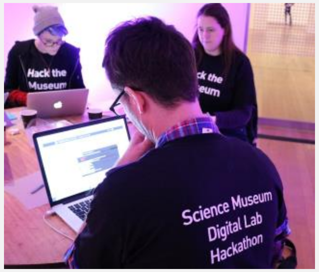
Team organization



Volunteers



Innovation



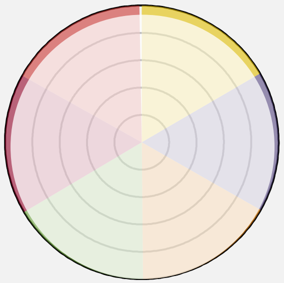
Citizen science

Big data

Peers

Creating a sustained digital culture

Normalising 'digital' within NHM London



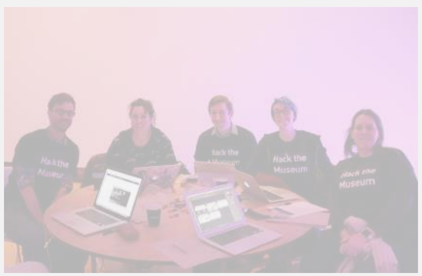
Prioritization

Outreach

Policy

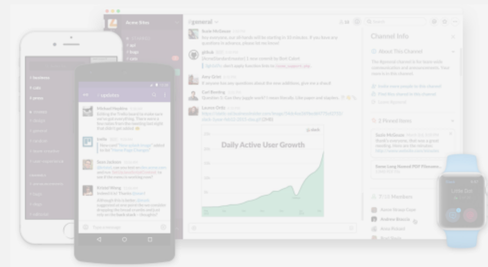
Government

Training

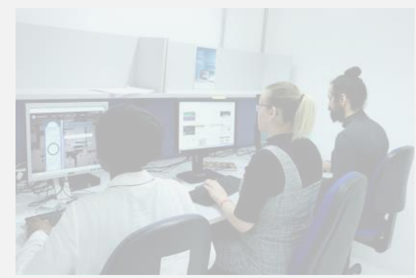


Events

Monitoring & KPIs



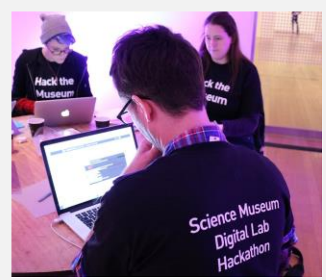
Team communication



Team organization

Volunteers

Innovation



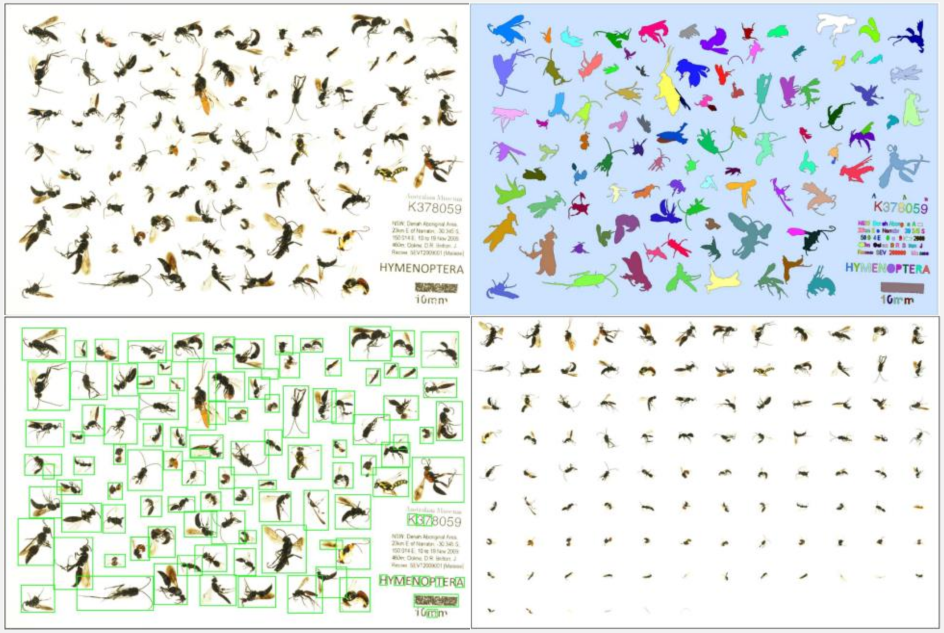
Citizen science

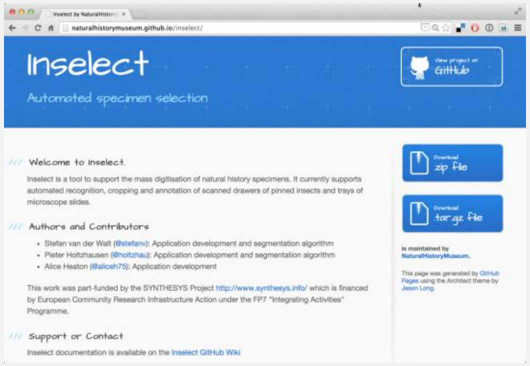
Big data

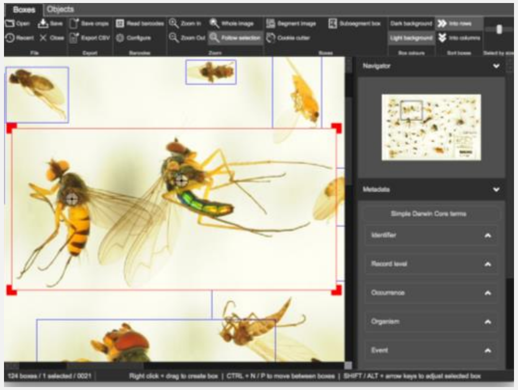
Peers

Innovation in image processing

Computer vision to detect, sort & process specimens







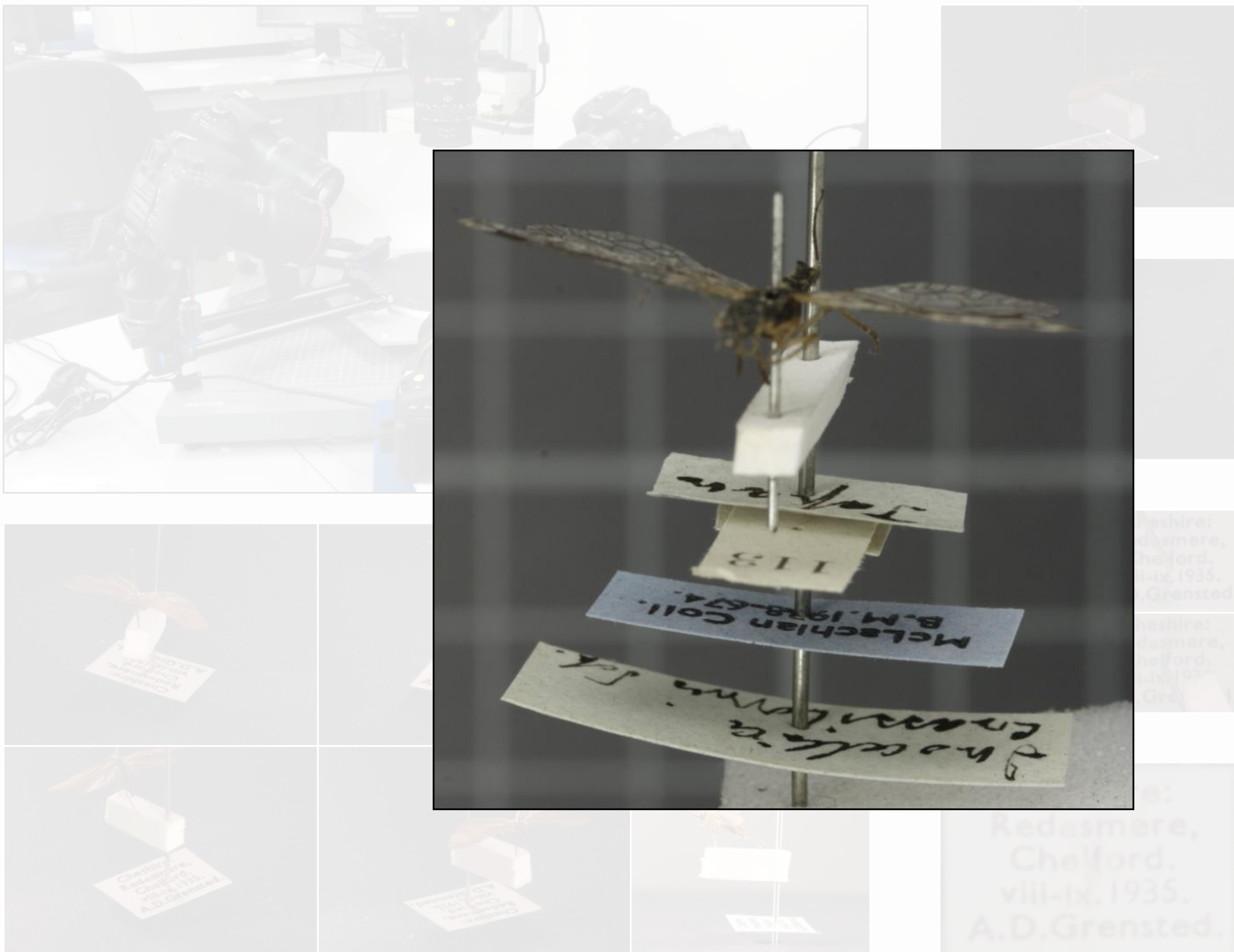
Inselect: edge detection software for collections

- Automatically detects specimens
- Batch processing (annotation & export)
- Multiple use cases
- Circa 10x faster than pilot insect digitisation

https://naturalhistorymuseum.github.io/inselect/

Innovation in image processing

Computer vision to extract label data



1. Label identification & perspective warp

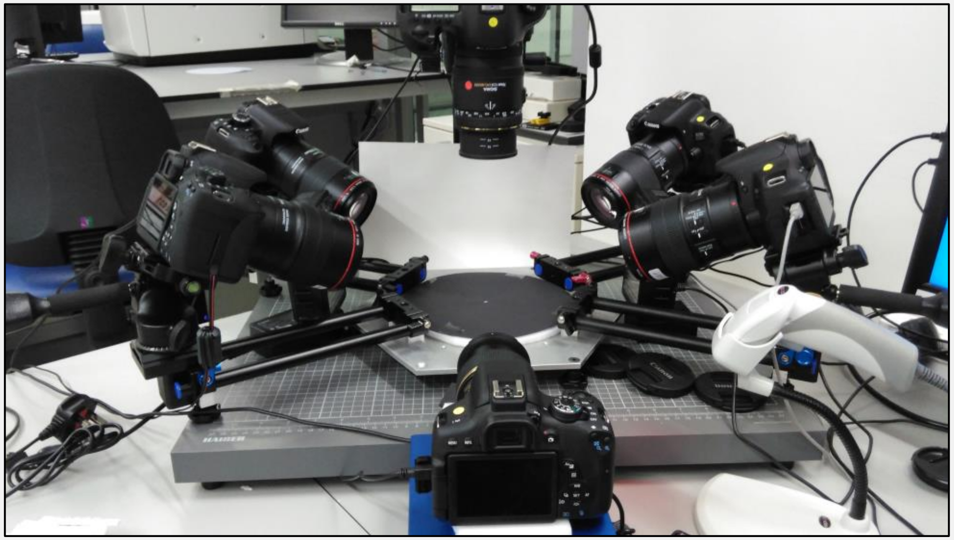
2. Label crop

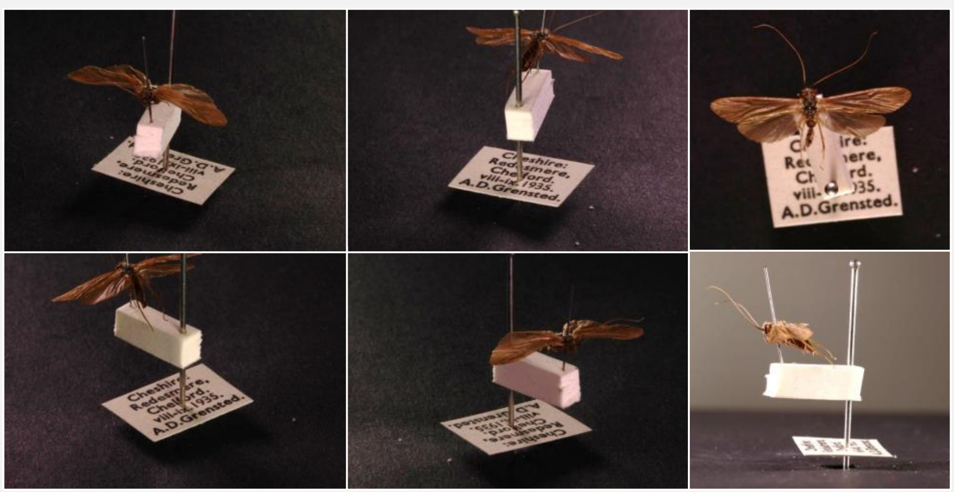
3. Label merge

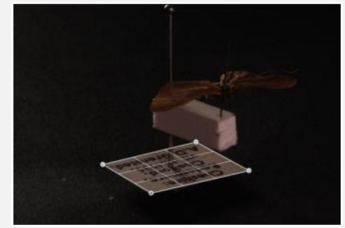
4. Finished for OCR or crowdsourcing

Innovation in image processing

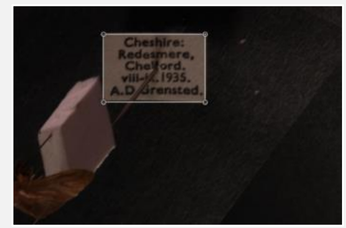
Computer vision to extract label data



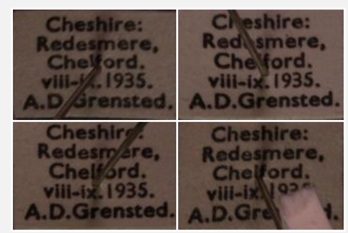




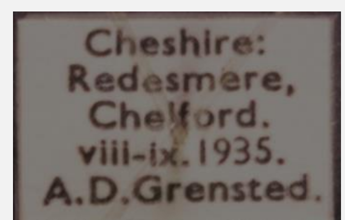
1. Label identification & perspective warp



2. Label crop



3. Label merge

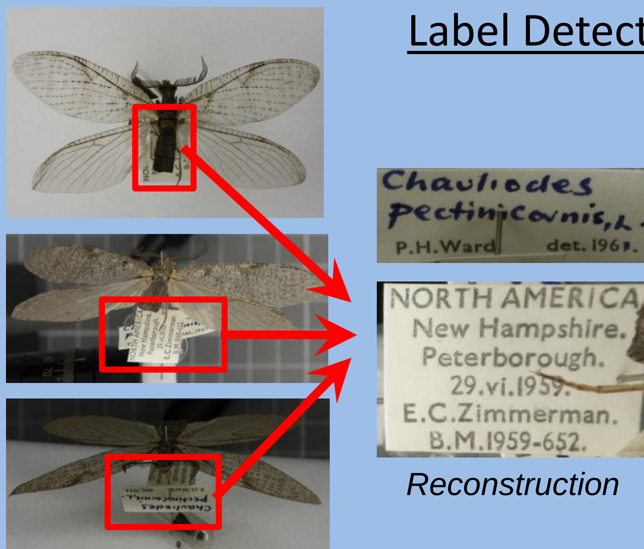


4. Finished for OCR or crowdsourcing

Computer vision research

High speed, automated data collection from large sets of images

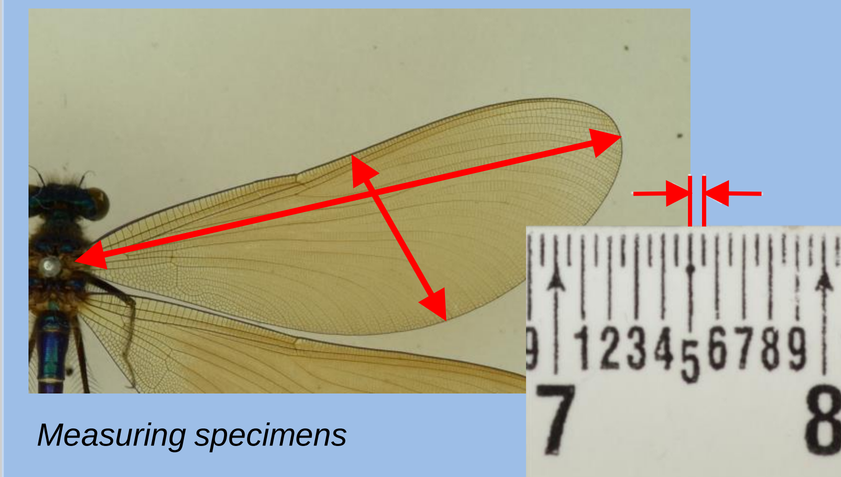
Label Detection



Reconstruction

Source

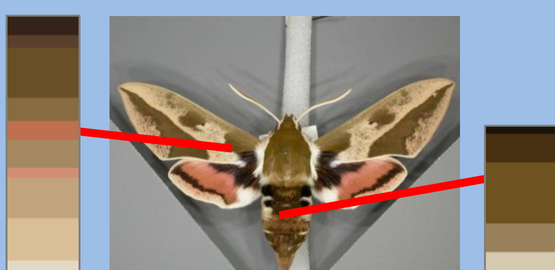
Morphometrics



Measuring specimens

Measuring rulers

Colour Palette Analysis



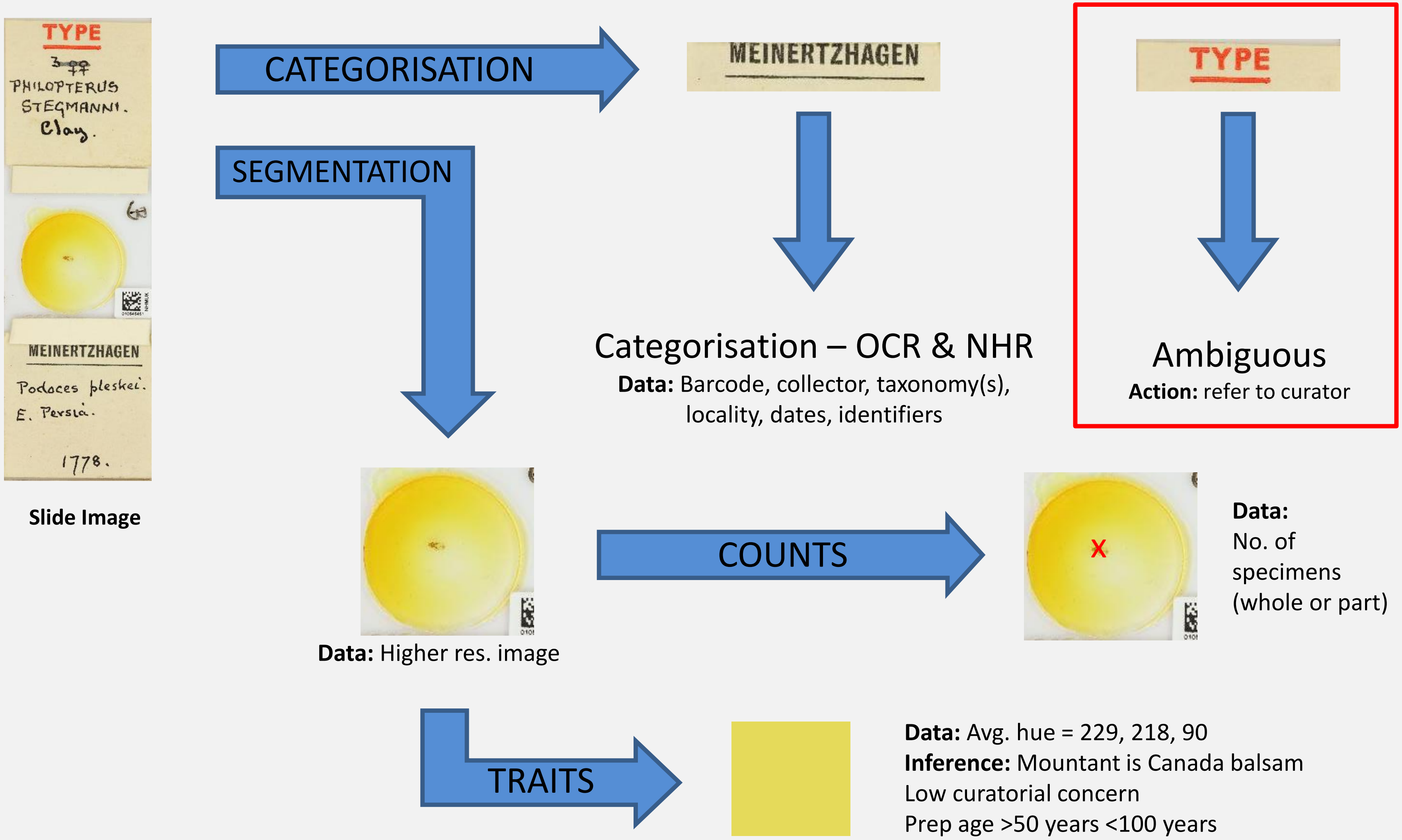
Hyles euphorbiae (Linnaeus, 1758)

Trait detection

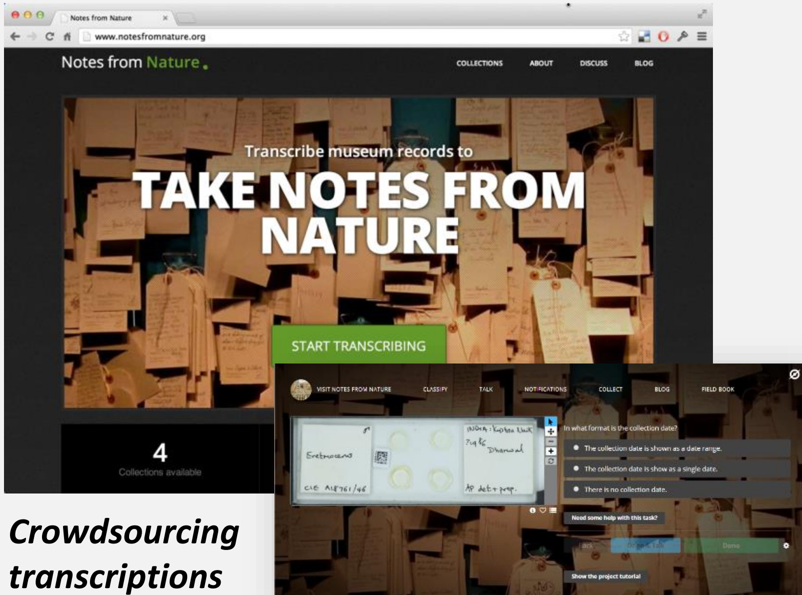
Sex brands in Hesperia comma

Specimen data refinery

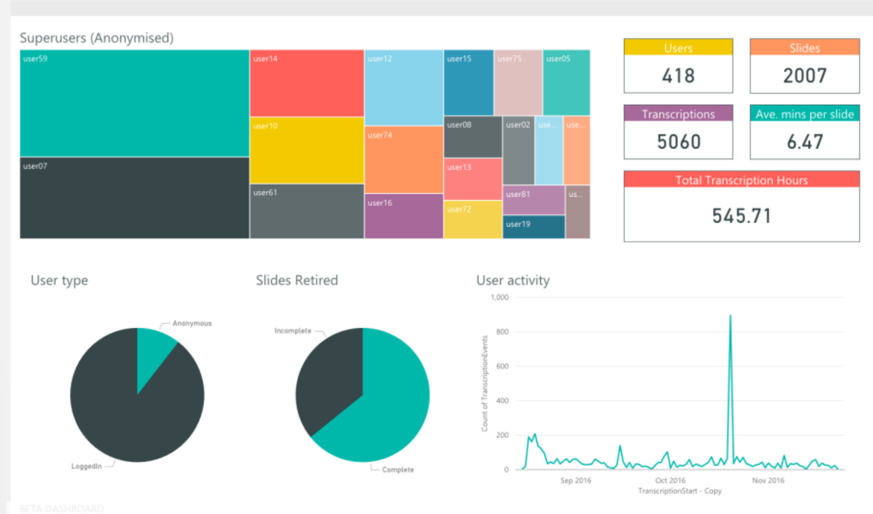
Linking processes, creating enhanced records & triaging tasks



Citizen science - Transcription

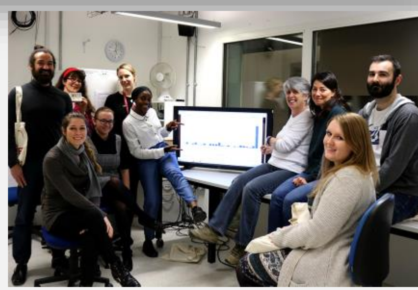


Crowdsourcing transcriptions

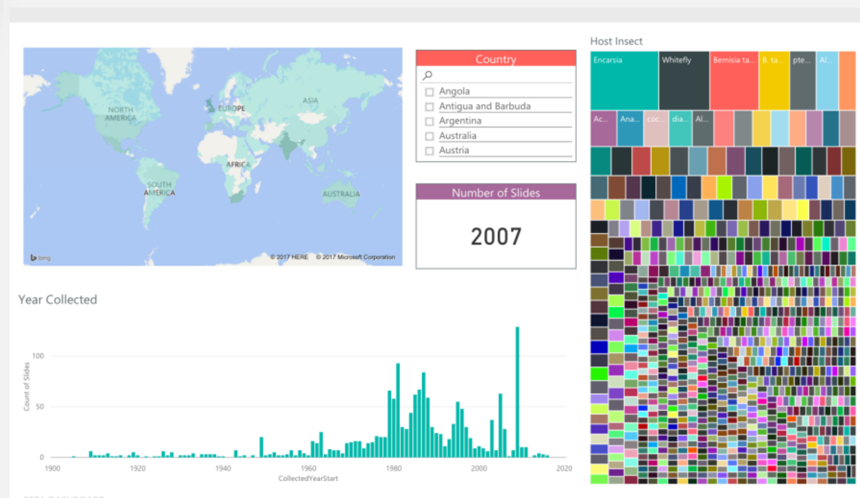


Data on transcribers

A partnership with Notes from Nature



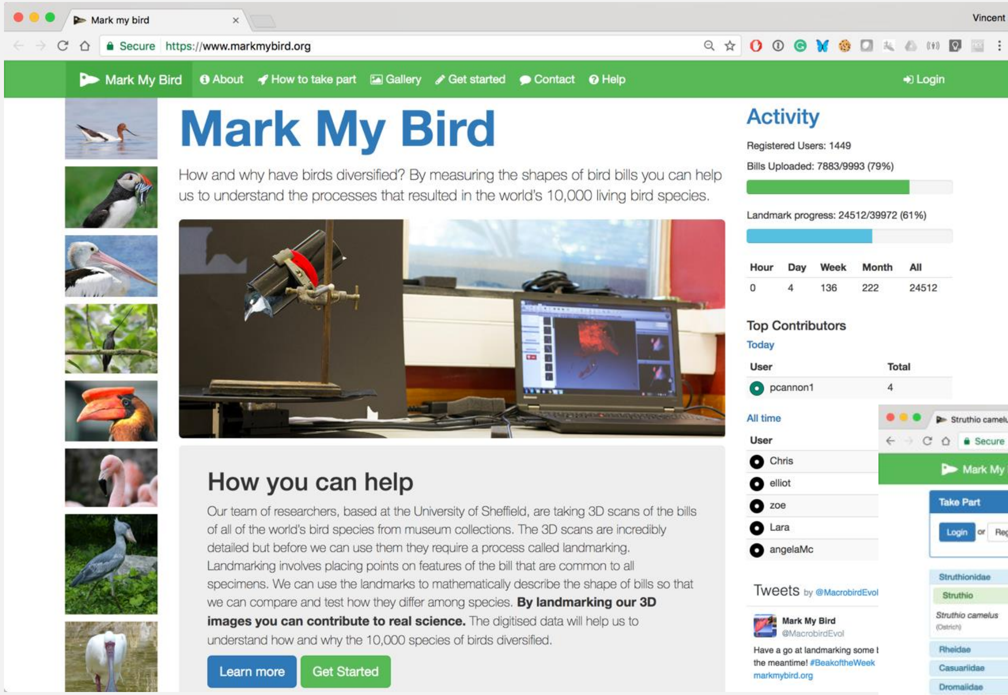
Part of a wider community of over 1 million users interested in citizen science projects



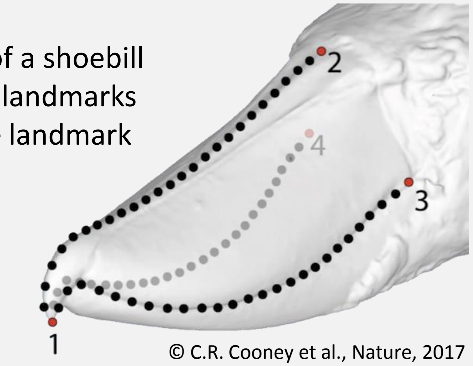
Data on transcriptions

3D Scanning of Bird Beaks

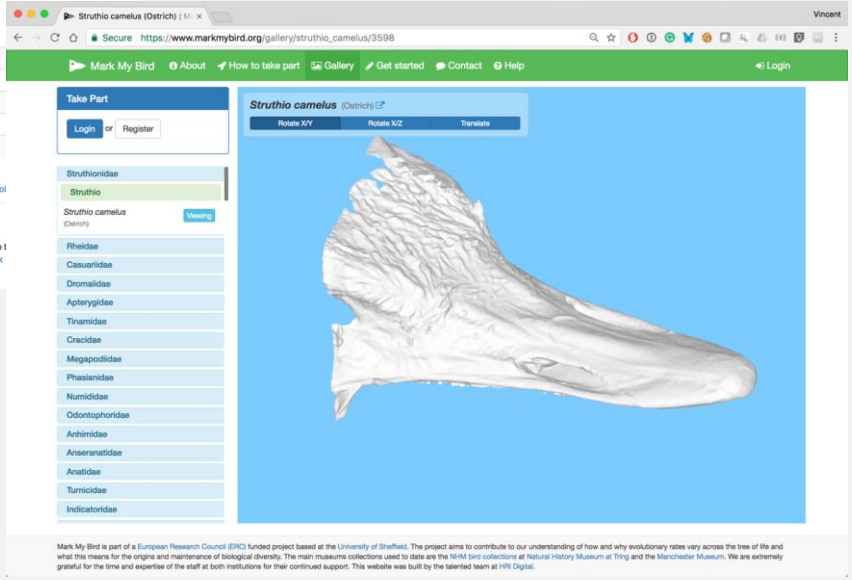
Citizen science to uncover the secrets of beak evolution



3D Scan of a shoebill with four landmarks and three landmark curves



© C.R. Cooney et al., Nature, 2017

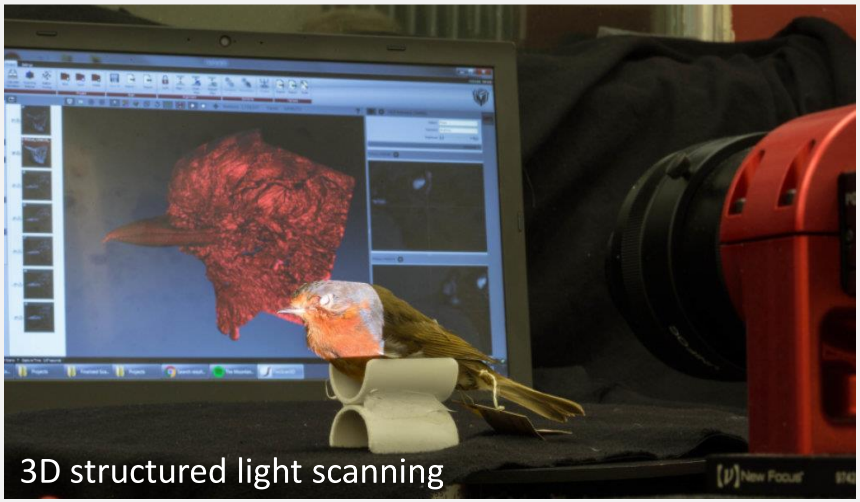


- Gavin Thomas' Team, Univ. of Sheffield
- NHM & Manchester Museum, 2k spp.
- Crowdsourcing website constructed
- Volunteers place landmark on bill scans
- NHM Portal (http://dx.doi.org/10.5519/0005413)

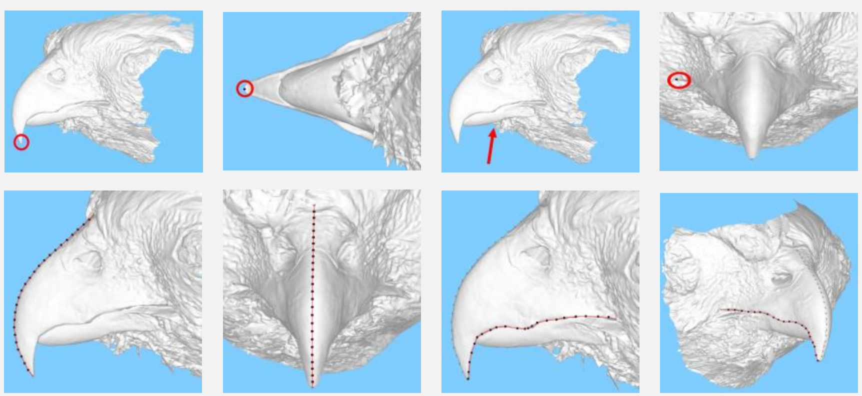
https://www.markmybird.org/

3D Scanning of Bird Beaks

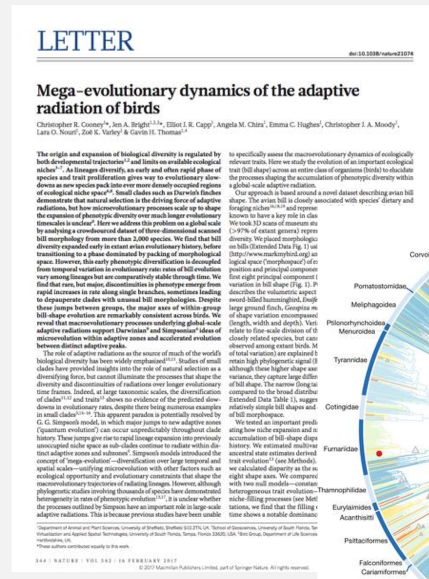
Citizen science to uncover the secrets of beak evolution



3D structured light scanning



Crowdsourced beak landmarks

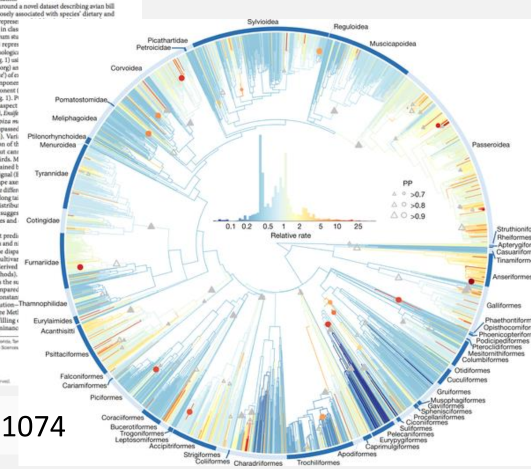


LETTER

doi:10.1038/nature21074

Mega-evolutionary dynamics of the adaptive radiation of birds

Christopher R. Cooney[1]*, Jen A. Bright[1,2,3]*, Elliot J. R. Capp[1], Angela M. Chira[1], Emma C. Hughes[1], Christopher J. A. Moody[1], Lara O. Nouri[1], Zoë K. Varley[1] & Gavin H. Thomas[1,4]

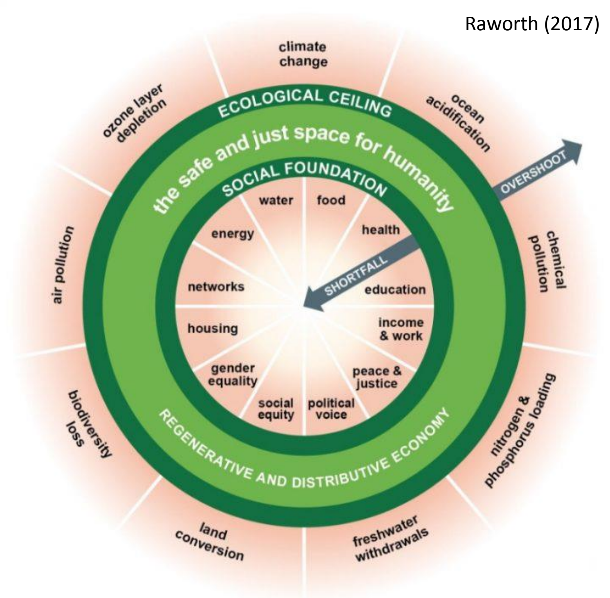


doi:10.1038/nature21074

- Morphology mapped to phylogeny
- Bursts of evolution as new groups emerge
- Fine scale changes thereafter
- Beak diversity expanded early in birds
- Unusual beaks evolved rapidly

Big collections data

Sustainable development: our challenge of the 21st century



e.g. climate change, ocean acidification, pollution, freshwater use, land conservation, ozone depletion

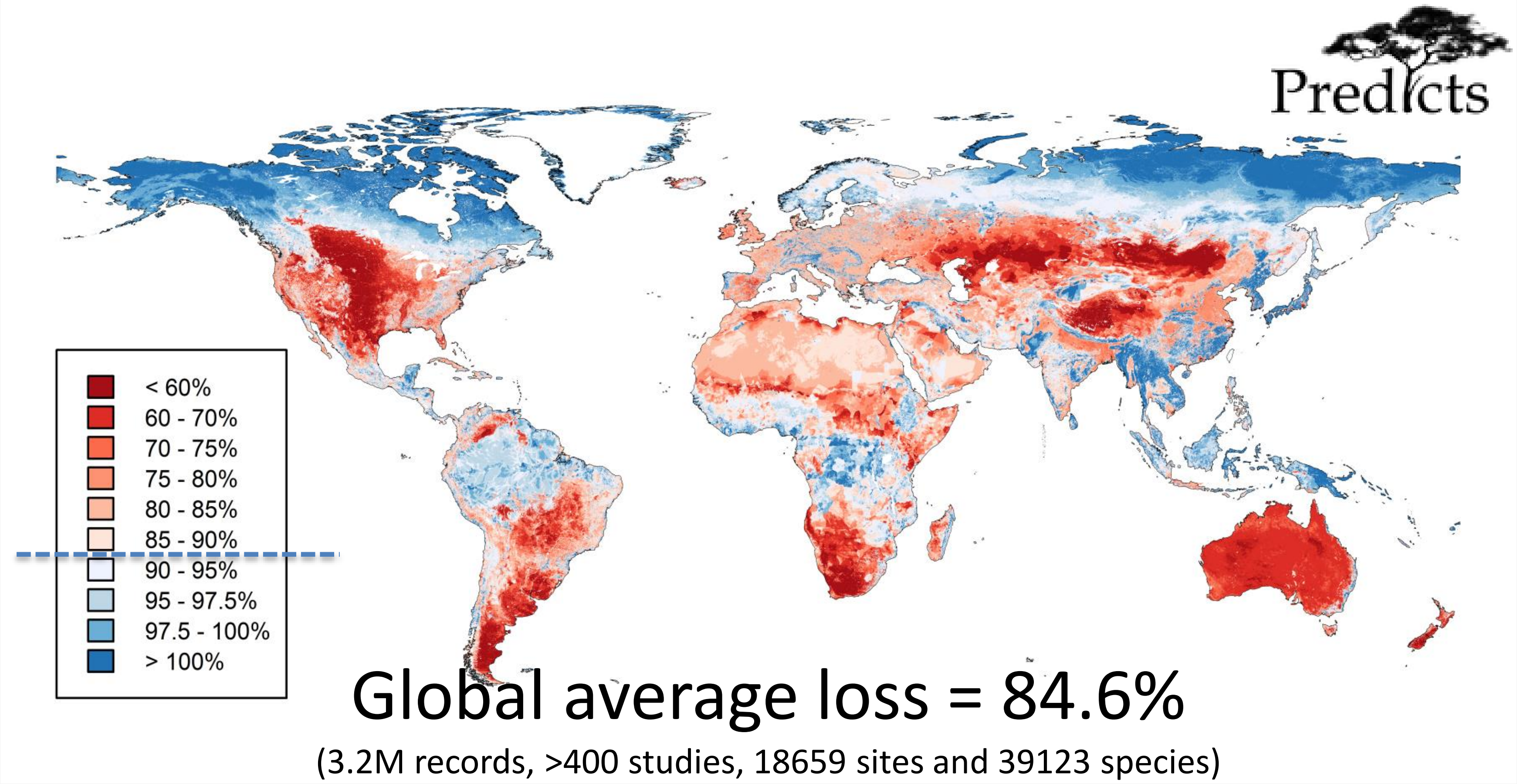
Can collections data tell us…

What will happen to biodiversity in the future?

Have we already exceeded the 'ecological ceiling'?

Measuring biodiversity loss

Aggregating biodiversity data before & after land use change to measure human impact

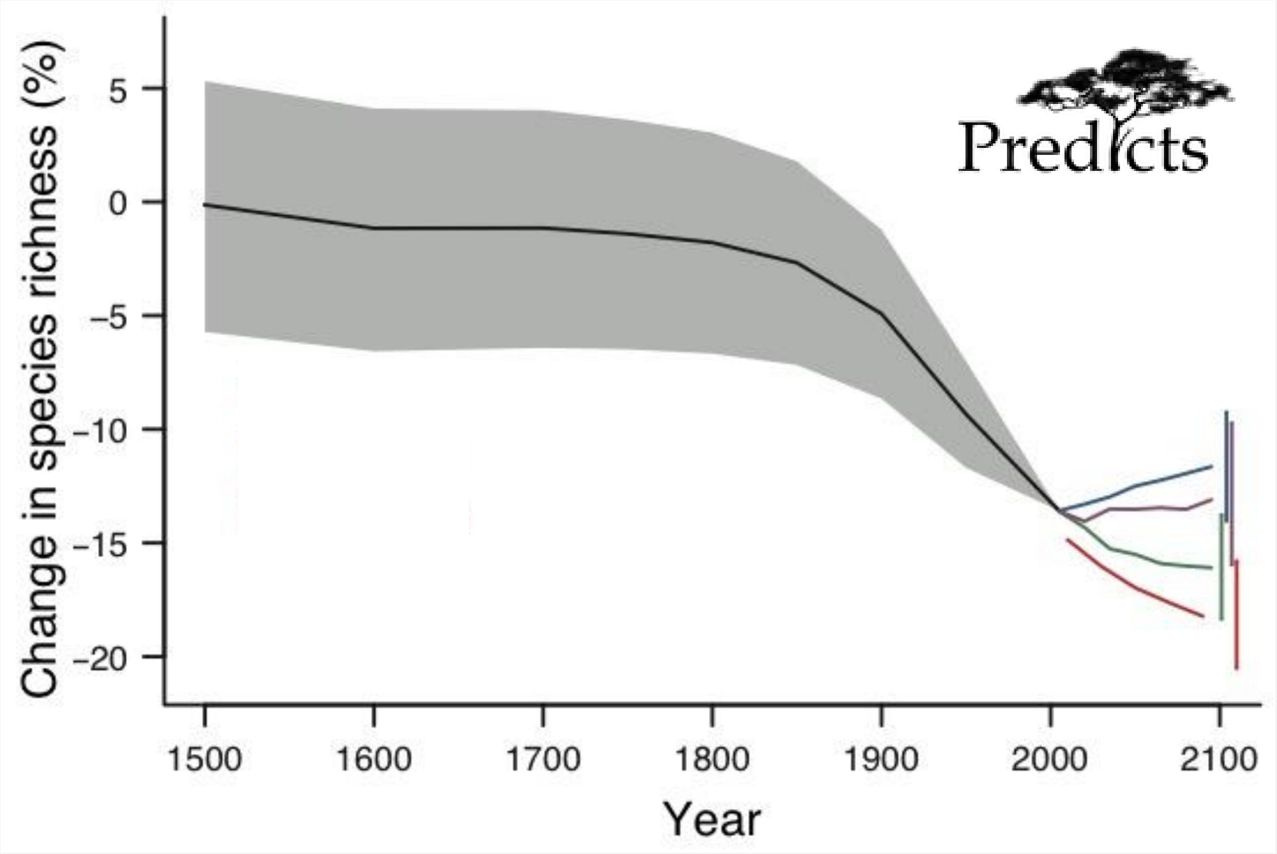


Newbold et al. 2016, *Science* 353:288-291

NHM Portal (http://dx.doi.org/10.5519/0066354)

Halting biodiversity loss

Digital collections to develop an index of planetary health



Newbold et al. 2016, *Science* 353:288-291

Climate Models

MINICAM 4.5:
8.5bn people, afforestation

IMAGE 2.6:
400ppm CO2, 9bn people

AIM 6.0:
860ppm CO2, 10bn people

MESSAGE 8.5:
12.5bn people

Biodiversity loss can be halted under some climate models

One Global Digital Museum

We need a knowledge platform for a 21st Century Museum

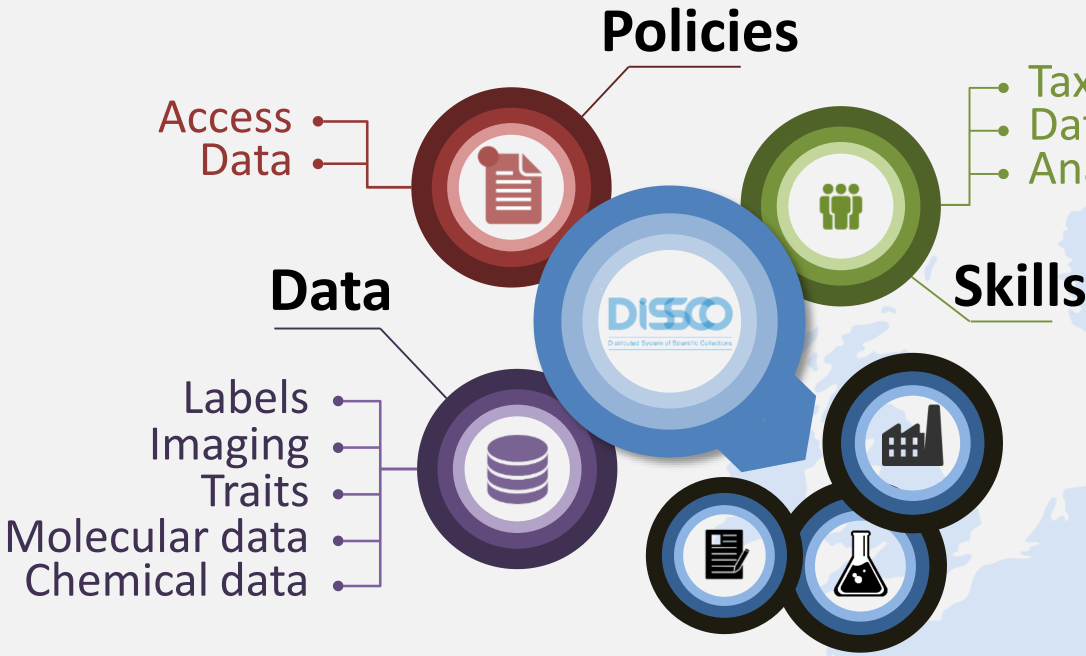
7,015 Collection holding agencies / institutions

181 Museums, 3622 Herbaria

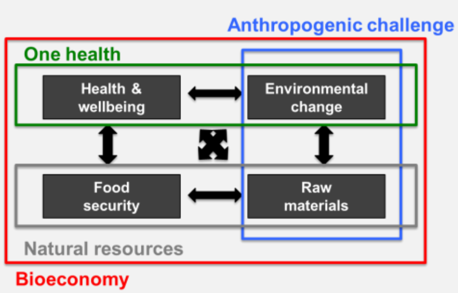
Data from http://grbio.org/

Distributed System of Scientific Collections

Transforming European NH institutes into an integrated research infrastructure



- DiSSCo
- An ESFRI proposal
- 110 inst. 20 countries
  (11 financially, 6 politically)
- Builds on current initiatives
  (eg. GBIF, CETAF, SYNTHESYS)
- 1.5B specimens
- 80% of planetary biodiversity
- Unifying:
  - Collections
  - Data
  - Services
  - Access policy
  - Strategy development
  - Governance



NATURAL HISTORY MUSEUM