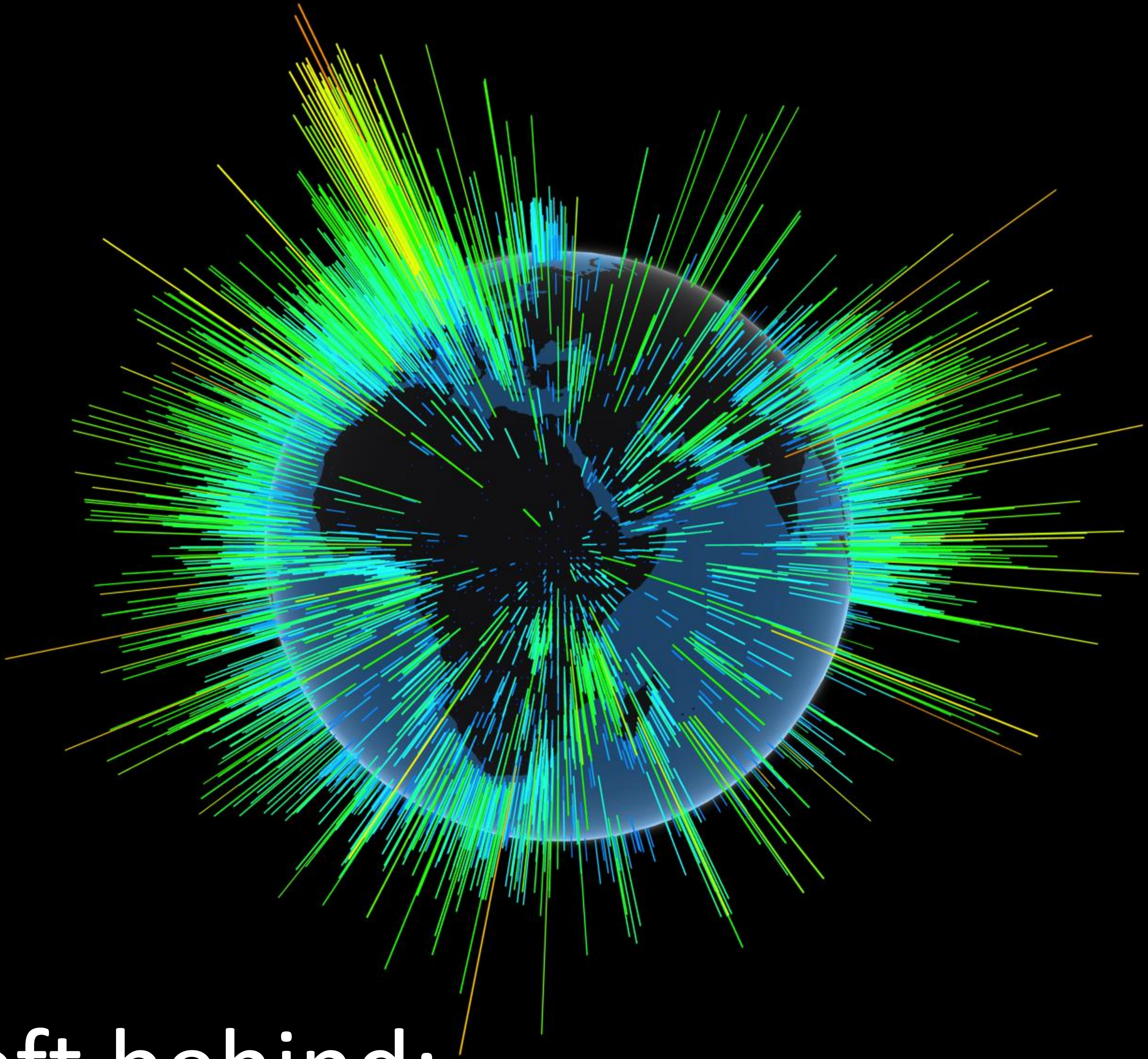


No specimen left behind:

The NHM Digital Collection Programme

Vince Smith, Natural History Museum, London
Digitisation of Biological Collections, 13 April 2014

1.5-3 BILLION SPECIMENS

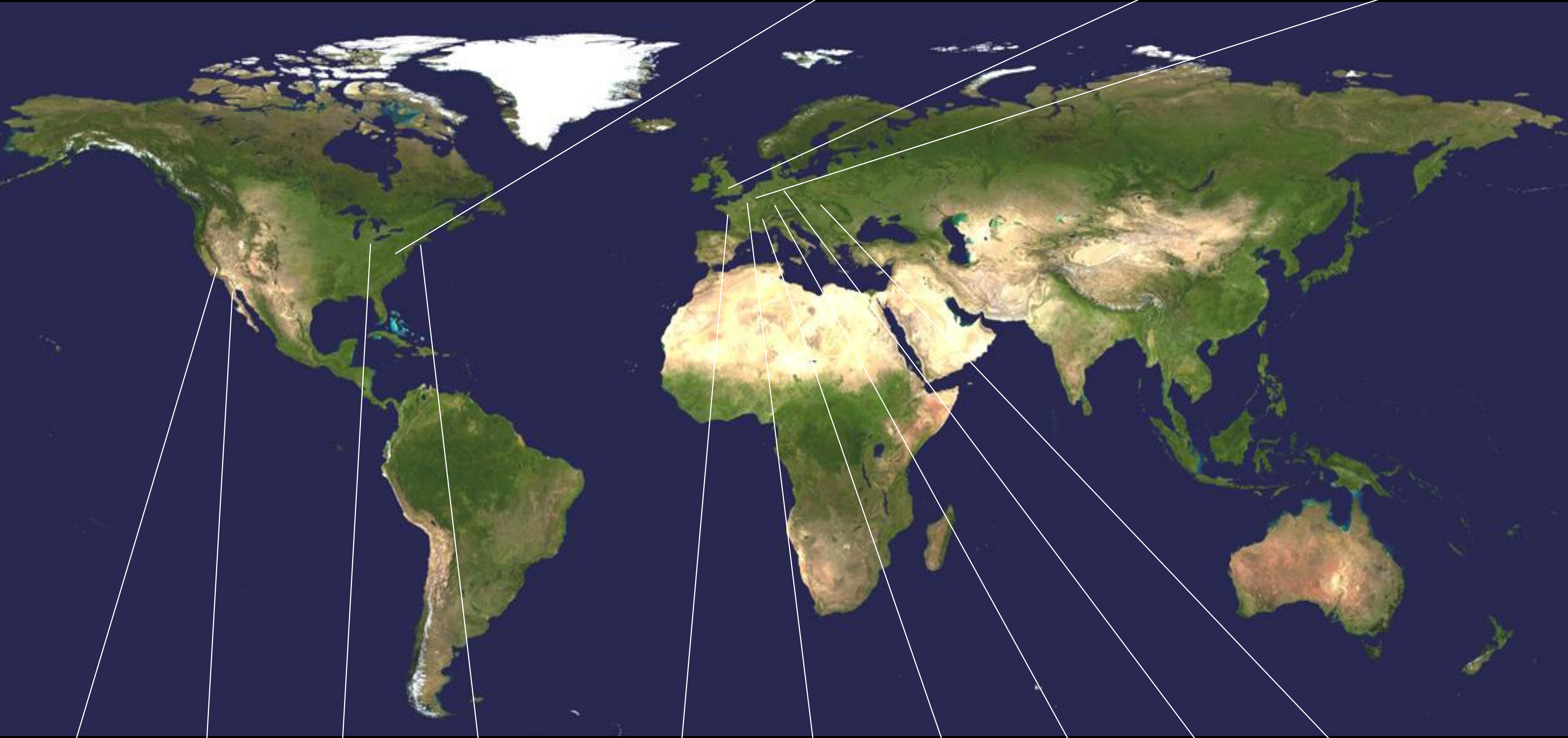
1.9 million species

300 years of collection

Washington
125M

London
80M

Paris
60M



San Francisco
28M

Los Angeles
35M

Chicago
25M

New York
30M

Brussels
37M

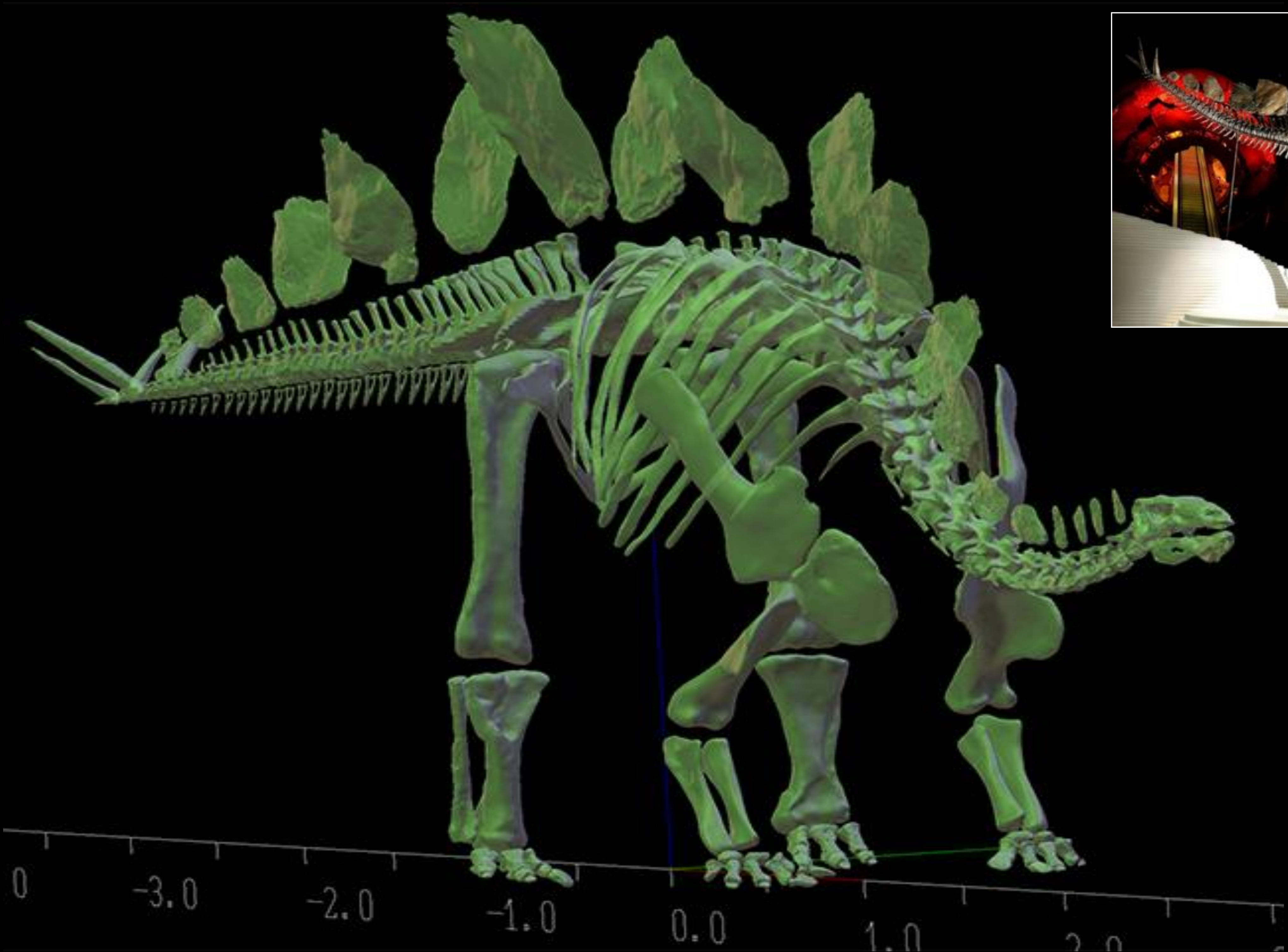
Leiden
37M

Vienna
35M

Frankfurt
40M

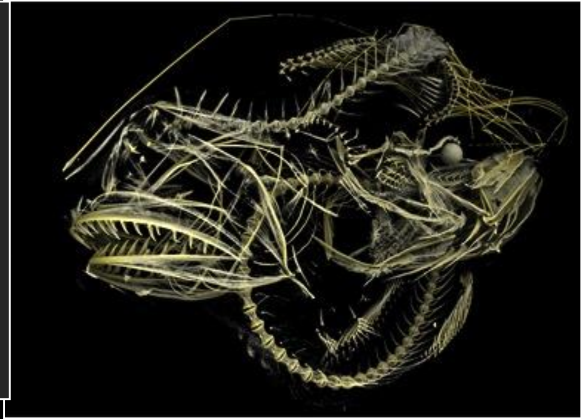
Berlin
30M

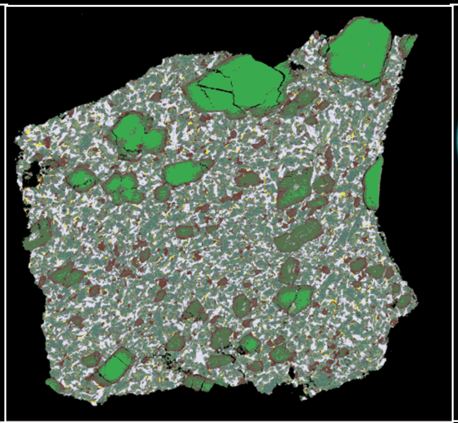
St P’berg
32M

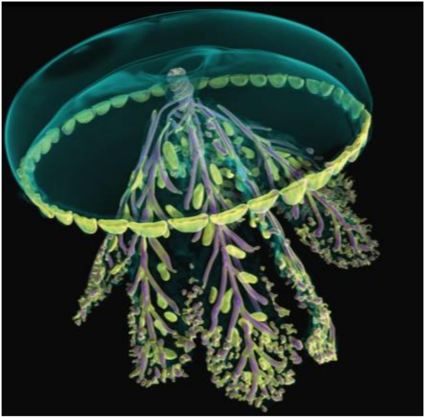


Digital surrogates of specimens:

- 3D models and printing
- CT imaging of internal structures
- Atomic, physical and chemical analysis







© CASSON MANN 201

Using digitisation to enhance public engagement with natural history specimens:

- **Augmented reality in gallery**
- **Animated content for pre- and post-visit apps**

An inordinate fondness for beetles:
400,000 species described so far, 100,000 types among the NHM's 10M beetle specimens

NHM collection

Collection area	No of objects	No of type specimens	Physical register	Digital data
Palaeontology	6,919,207	43,146	2,364,232	340,636
Mineralogy	423,563	615	425,000	402,727
Botany	5,863,000	172,750	127,200	645,222
Entomology	33,753,257	612,796	57,197	255,000
Zoology	27,501,350	325,000	1,986,000	1,160,216
Library & archives	5,460,000	-	-	-
TOTAL	**79,920,377**	**1,154,307**	**4,959,629**	**2,803,801**

**<3% of NHM specimens are digitised,
& even fewer are 'computable'**

Digital NHM: a strategic priority

Museum Strategy 2015-2020

Thee big narratives

①Origins and evolution
②Living diversity
③Sustainable futures

Four priority areas

①Digital
②National
③International
④London

Priority Area #1: Digital

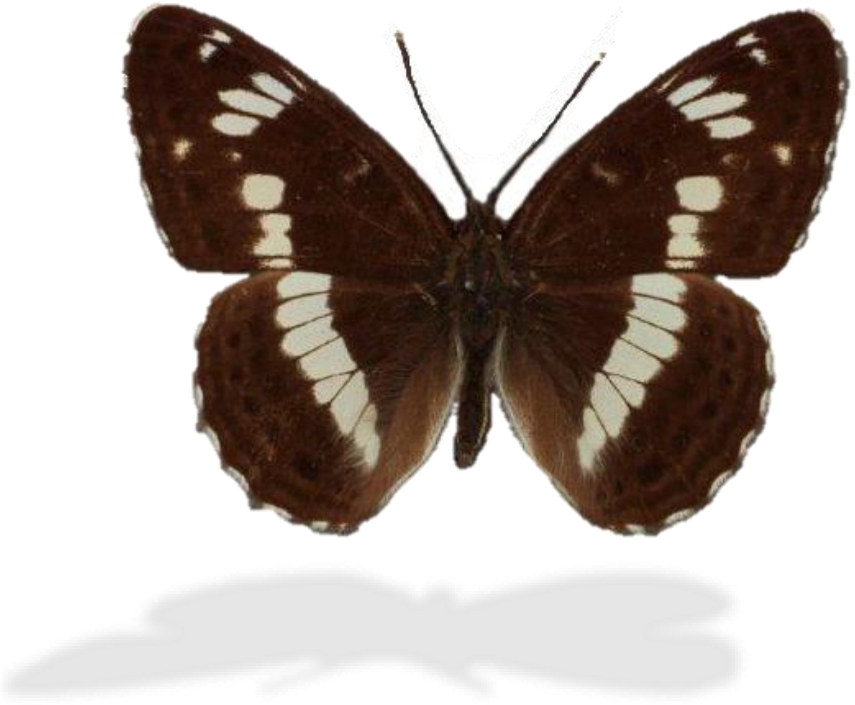
Objectives	Key projects
Big, open data	Mass collection digitisation* NHM Data portal* Crowd-sourcing of data*
Global communities	New website* Social media channels Membership
Innovative platforms	Location-aware Wi-Fi* Museum App* In-gallery technology

NATURAL HISTORY MUSEUM

iCollections: 2013-2015

- A pilot for the Digital Collections Programme
- iCollections digitisation criteria:
  - Entire collection
  - No existing digitisation pipeline
    (pinned, slide & herbarium specimens)
  - High research potential
    (phenology, morphometrics, migration patterns, pest associations, automated species recognition)
  - Curation opportunity

- Data outputs
  - Image(s)
  - Metadata (what, when & where)
  - Georeference point localities



NHM iCollections project:

- UK butterflies & moths
- 500k specimens
- 2 mins per specimen
- £1 per specimen

Large-scale digitisation:

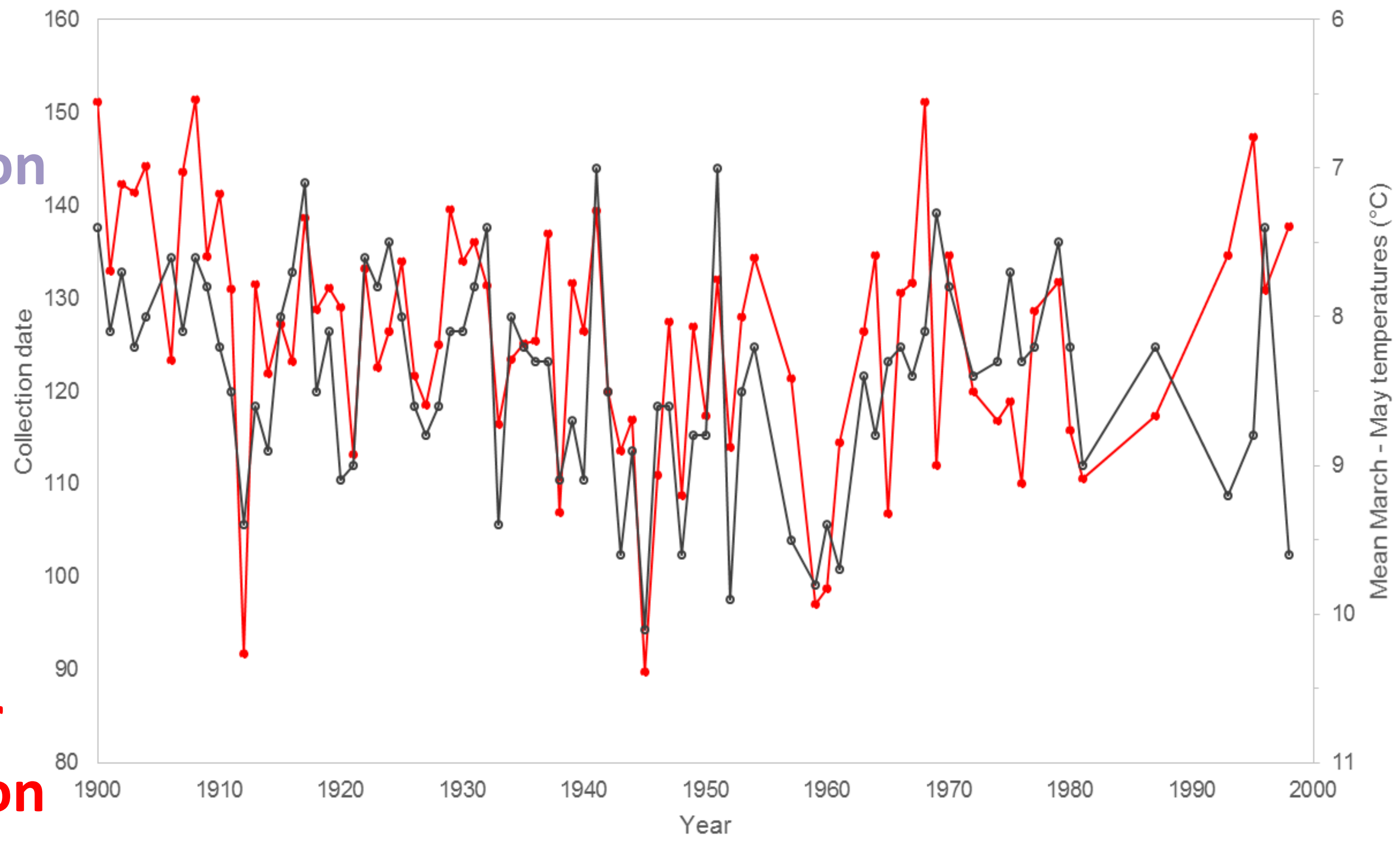
- High-throughput digitisation workflows
- Informatic pipelines
- Computer-assisted object recognition

iCollections digitisation process

iCollections research: long-term trends in phenology

Later collection

Earlier collection



Cooler springs

Warmer springs

Orange tip butterfly

(*Anthocharis cardamines*)

- **Many species emerging earlier and earlier each year**
- **Initial collection date & temperature highly correlated**
- **A unique marker on phenological response before recent climate change**
- **Longer time perspective than most observational records (BMS post-1976)**
- **Museum data available for rare or hard to record species**

Brooks, Self, Toloni & Sparks, 2014, *Int. J. Biometeorol.*
DOI 10.1007/s00484-013-0780-6

NHM Digital Collections Programme

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

- an online specimen- / lot-level data base for all holding

- core meta-data and / or images for key parts of the collection, and

- flexible informatics and visualisation tools”

Target = 20 million specimens digitised in 5 years

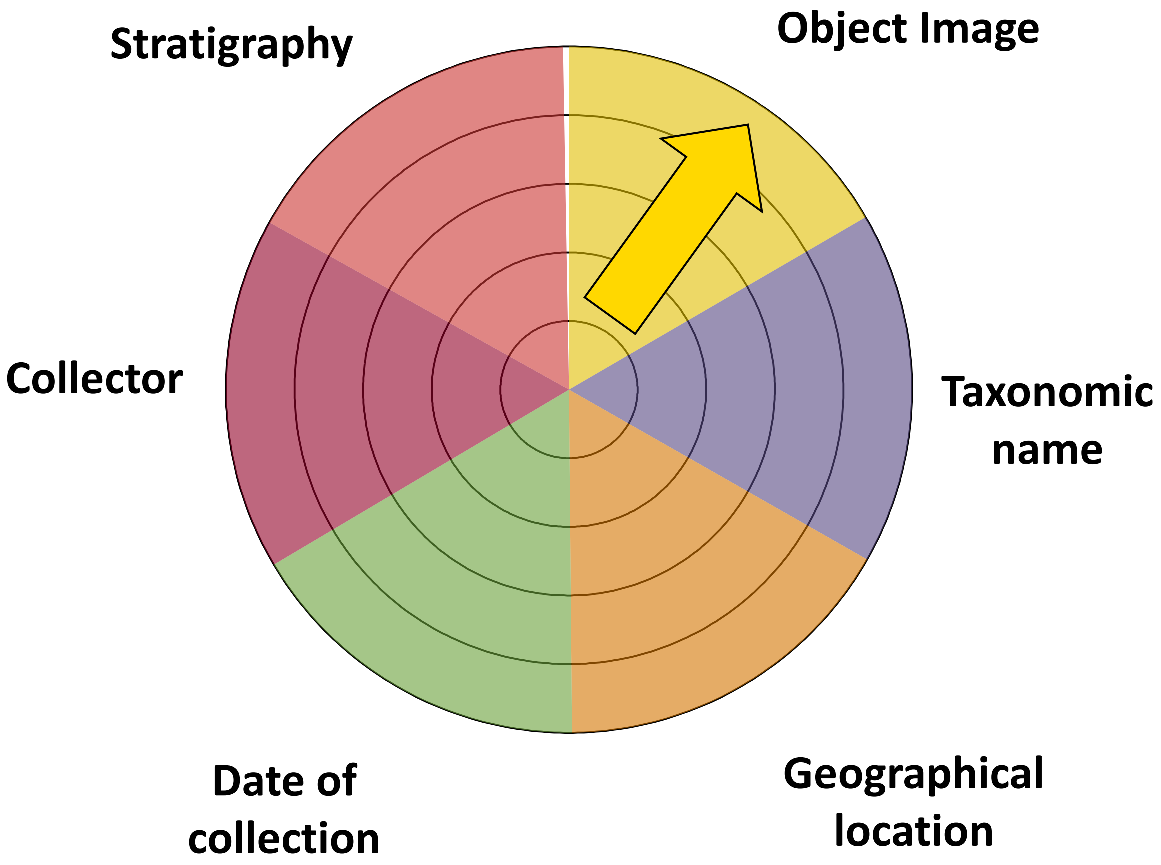
	2 year	5 year	10 year
POLICY & PROTOCOL	Defined data policy and standards	Policies embedded in NHM operating practises	Leaders of process in the digital curatorial world
DATA CAPTURE	Prioritised digitisation Workflows piloted	Portfolio of mass digitisation output projects	Some major collections digitised
PEOPLE & SKILLS	Task force formed and operating	Best-practice processes integrated into training	Digital curation as a core part of our 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 & initiatives	Collaborative research material published	Major contributions to grand challenges
ACCESS	Live NHM Data Portal	Tools , visualisations & analytics	Integrated global network of users

Digital Collections Task Force

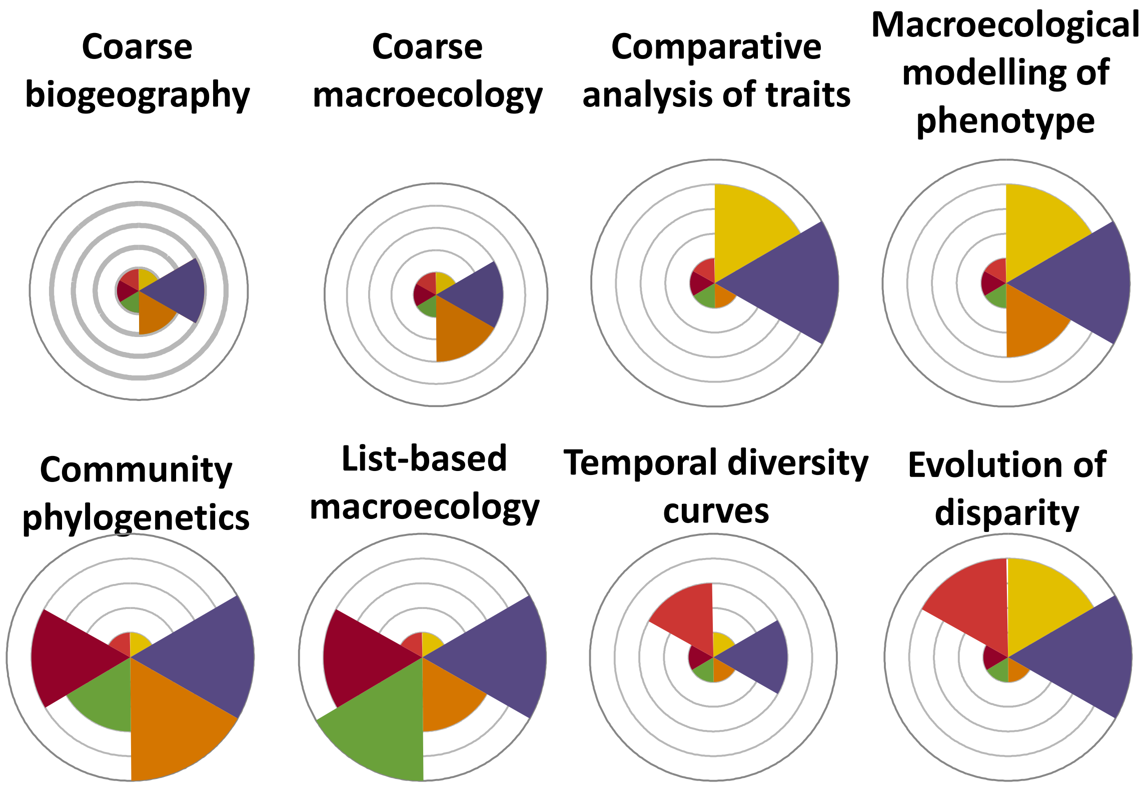
- Core pool of expertise, ideas and capacity for the Museum's digitisation activity
- Key individuals across departments
- 50% or more of their time on projects under DCP
- Individuals seconded to relevant projects
- Coordinated by a steering group made up of units lead individuals and chaired by the Head of Collections Digitisation
- Membership expected to change as the portfolio of projects evolves
- **Change from within**

Question-oriented prioritisation

Core meta-data categories:



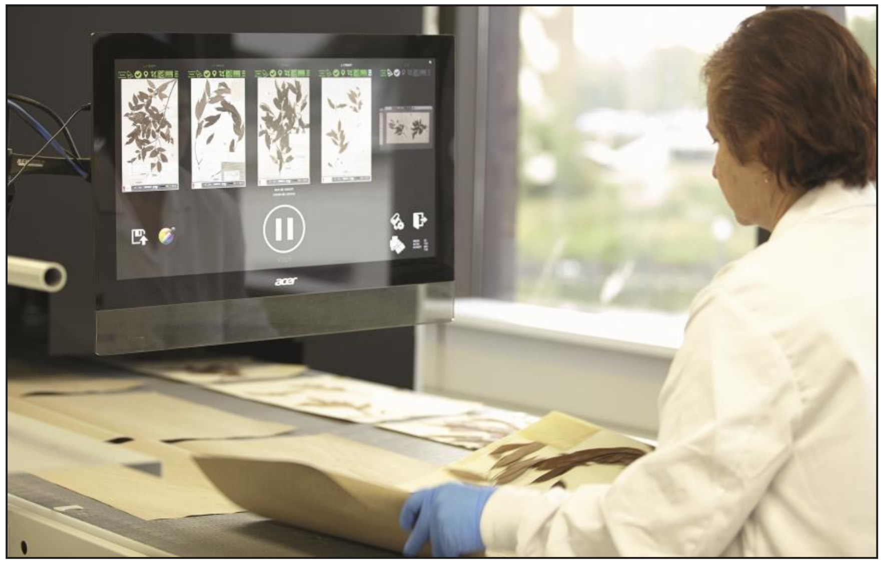
What meta-data are required for different types of study?



Digital Collections Programme: pilot projects

Pilot 1 – Herbarium sheets

- Trial partnership with Kew & Picturae (cf Naturalis)
- Conveyor-based imaging equipment
- Outsourcing of label transcription
- C.70,000 sheets
- January-April 2015

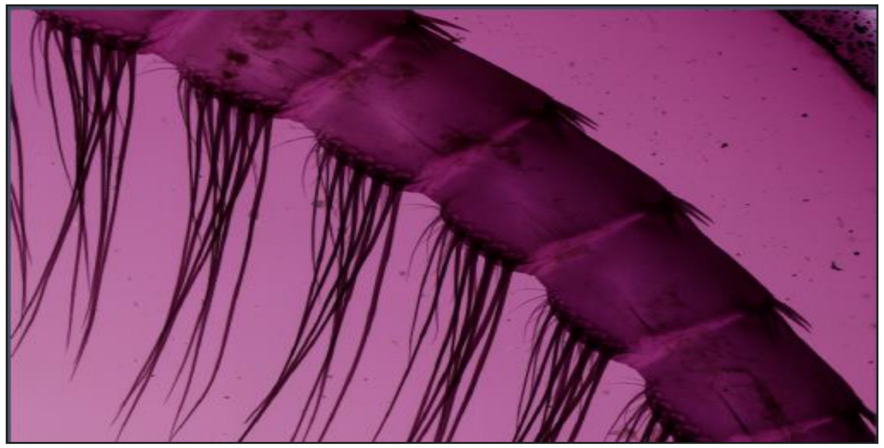


Pilot 2 – Palaeontology

- British Fossils (Mesozoic vertebrates)
- Standardised photography
- Capturing taxonomic & stratigraphic metadata
- February 2015 – March 2016

Pilot 3 – Microscopic slides

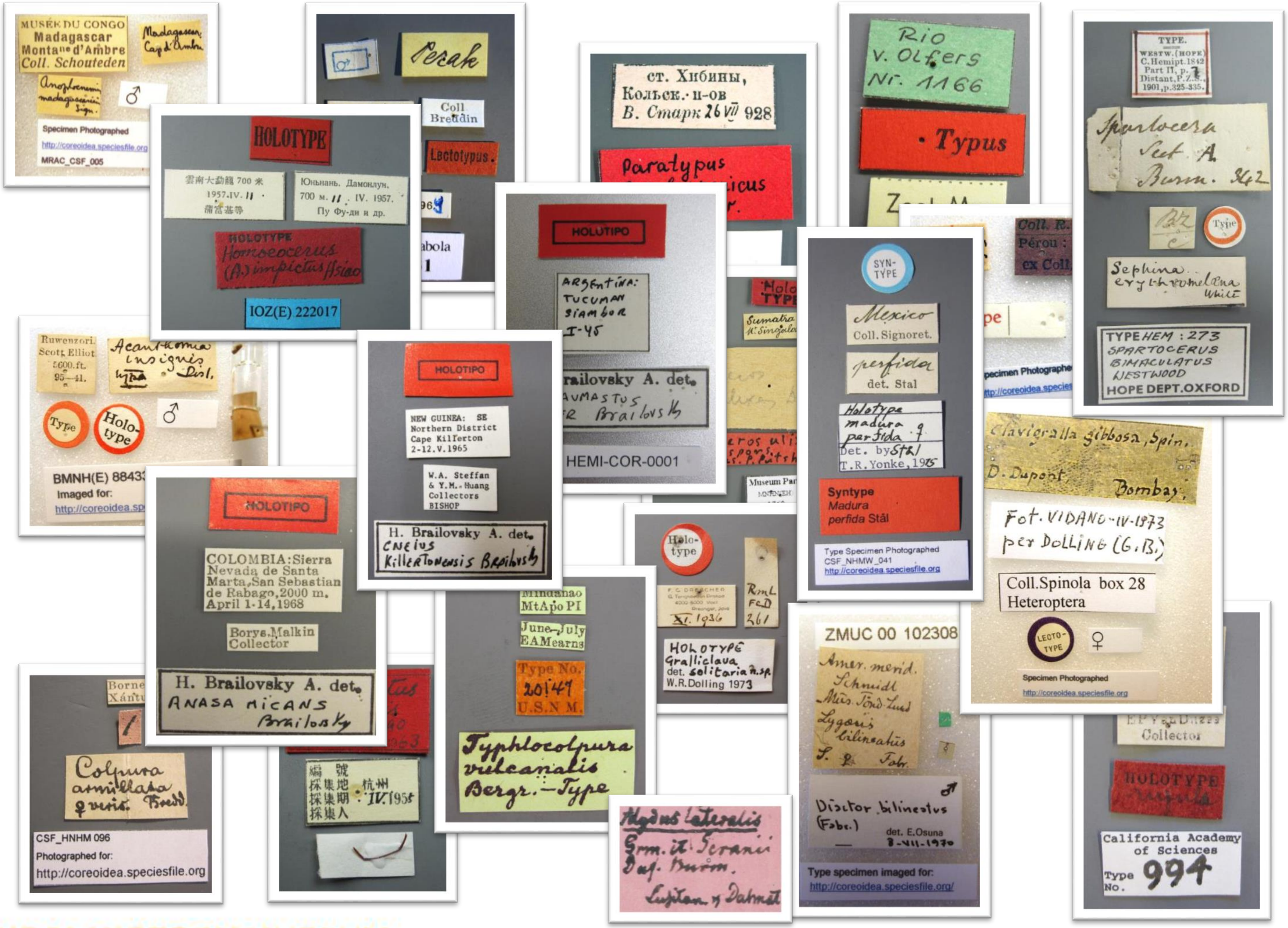
- High-throughput slide scanner
- Satscan to capture label information
- Call for pilot project ideas
- April 2015 onwards



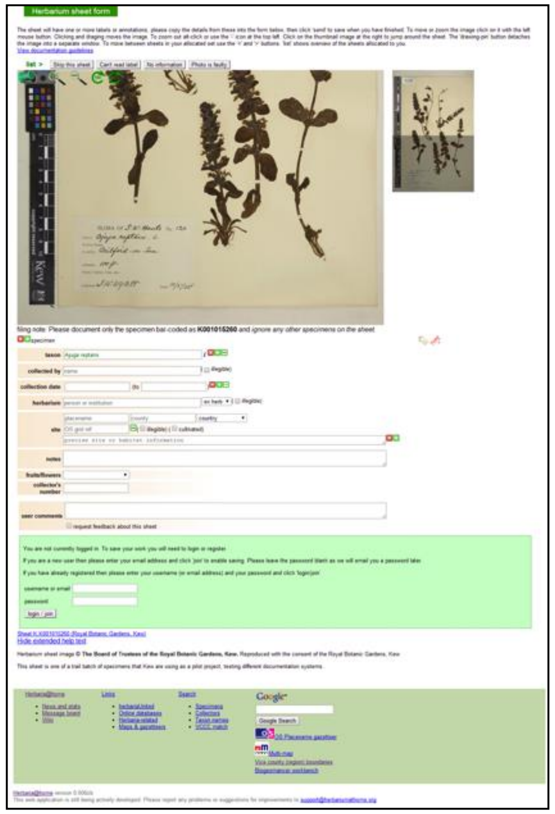
NHM & Kew Herbarium sheet digitisation

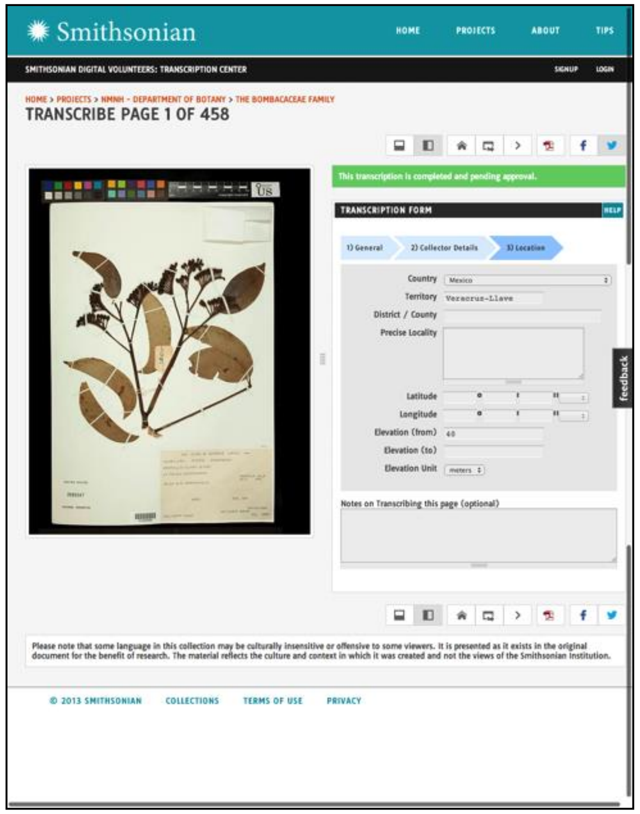
33k Specimens per day, 3 shifts (6am-10pm), Netherlands collection complete in 1.5 years
€1.29 Euros per specimen image (if outsourced), transcription at similar cost

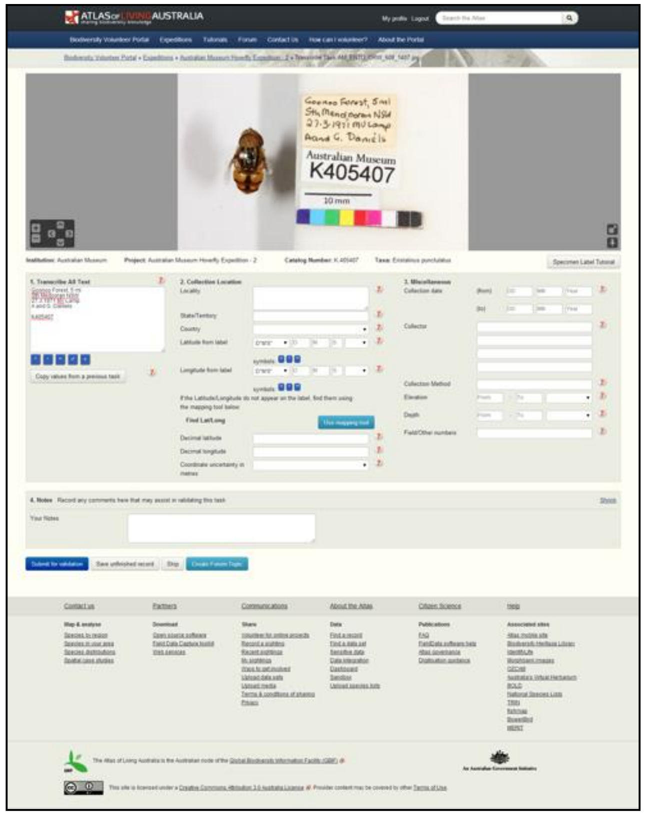
Crowdsourcing

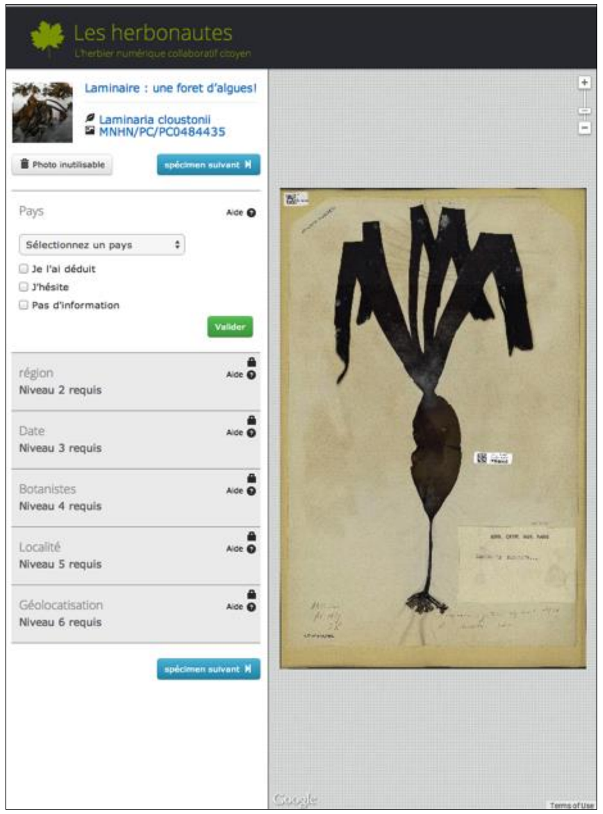


Crowdsourcing platforms









Herbarium @Home
http://herbariaunited.org/atHome/

Smithsonian Transcription Center
https://transcription.si.edu/

Atlas of Living Australia
http://volunteer.ala.org.au/

Les Herbonautes
http://lesherbonautes.mnhn.fr/

Research by the people, for the people.
We have hundreds of projects waiting for your help to achieve amazing goals.

Notes from Nature
http://www.notesfromnature

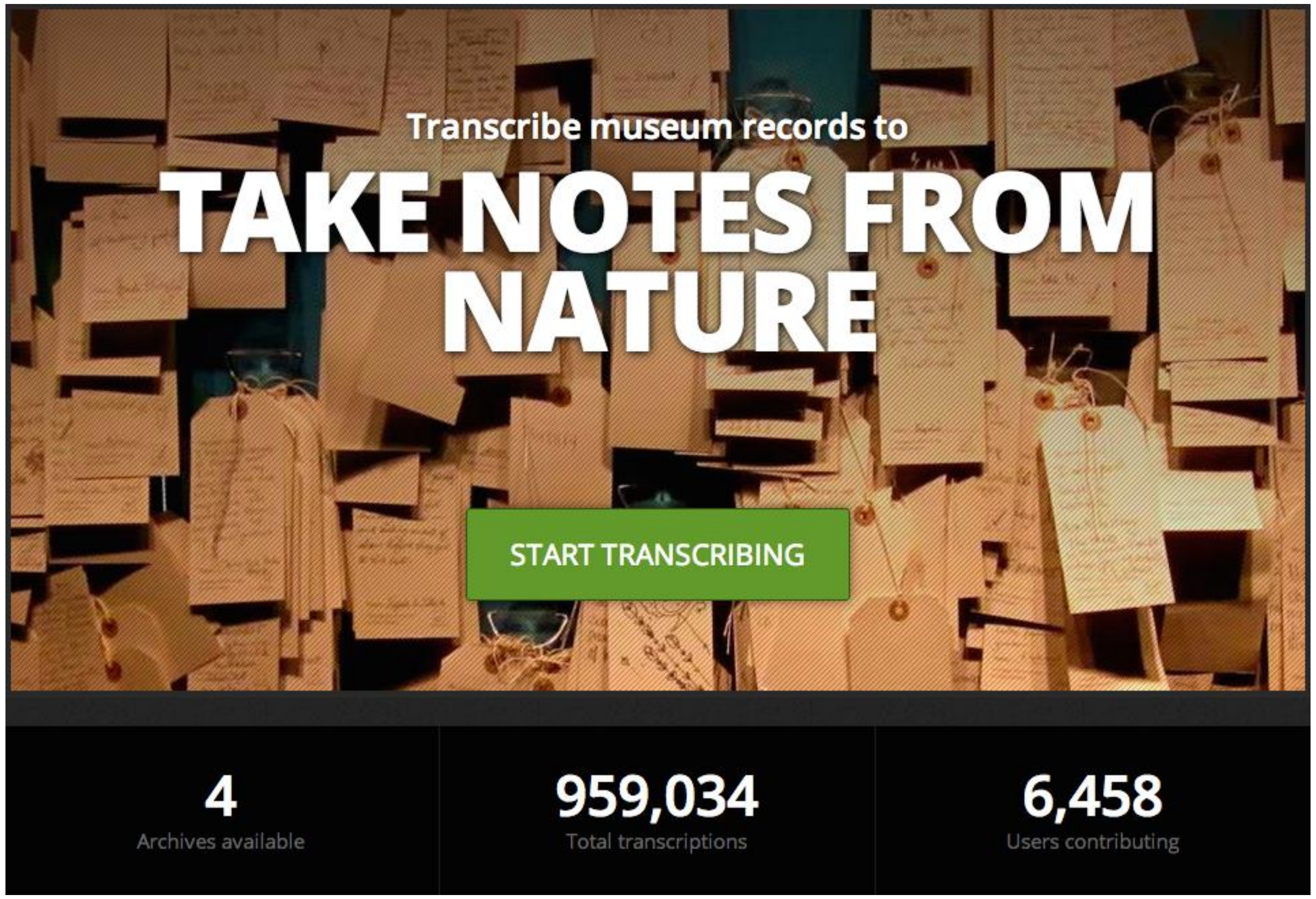
Crowdcrafting
http://crowdcrafting.org

Notes from Nature (Zooniverse)

NHM Ornithology Registers (1837 – 1990)

Progress

- Total Images: 2,950
- Active Images: 126
- Complete Images: 2,824
- 307,305 transcriptions
- Circa 1/3rd by one person

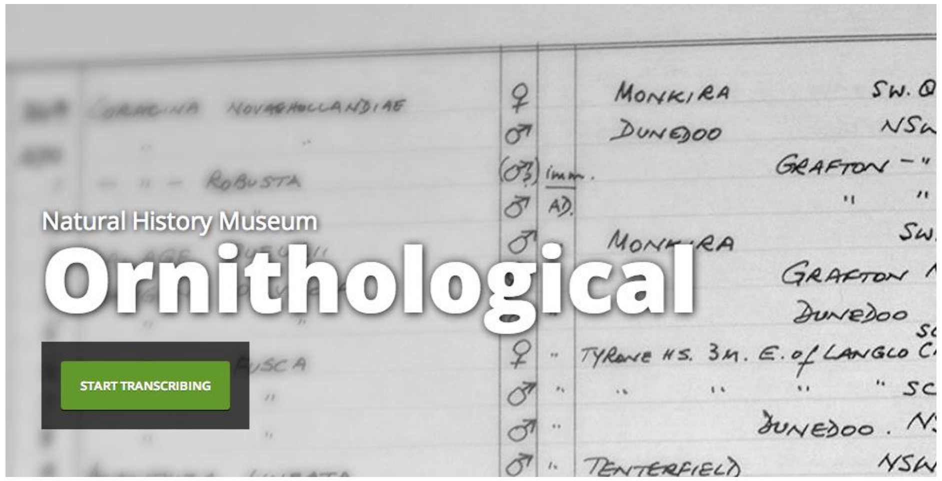


Transcribe museum records to

TAKE NOTES FROM NATURE

START TRANSCRIBING

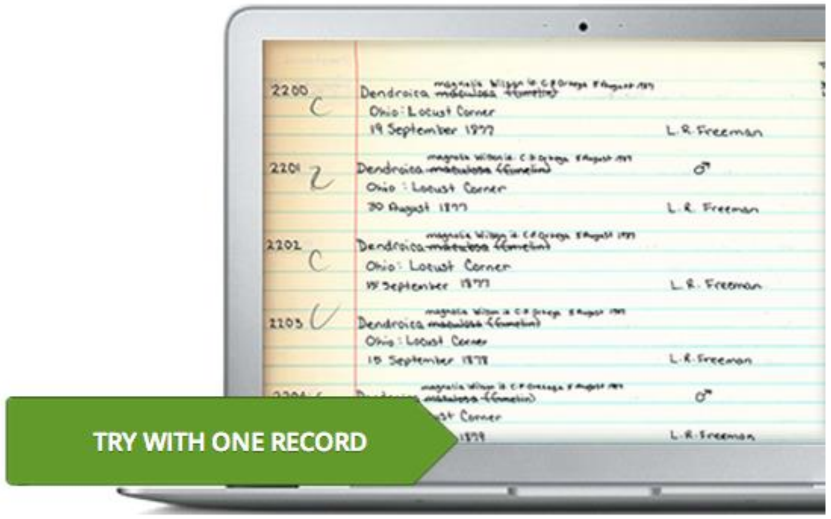
4	959,034	6,458
Archives available	Total transcriptions	Users contributing



Natural History Museum

Ornithological

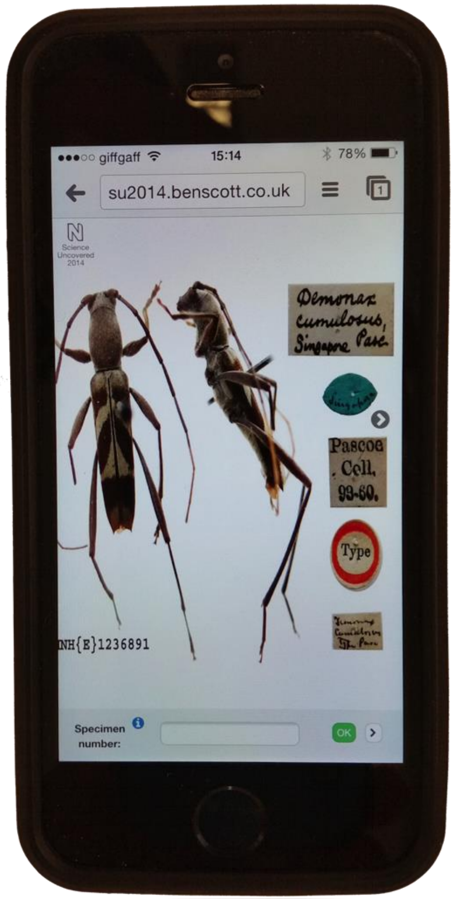
START TRANSCRIBING

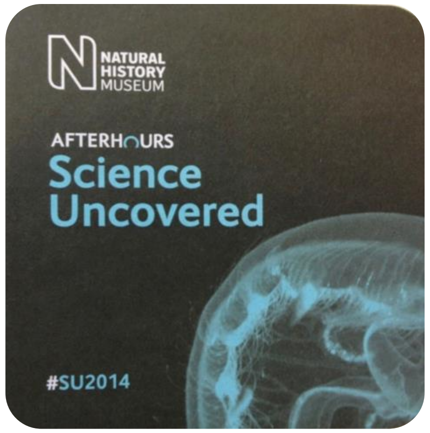


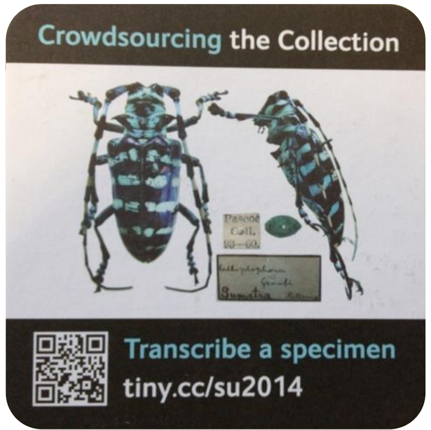
TRY WITH ONE RECORD

Science uncovered 2014: “Crowdsourcing the Collection”

- Demonstrate digitisation process
- Engage the public in transcription
  - Digitise; web publish images; public transcription; data publication
- Dedicated mobile website



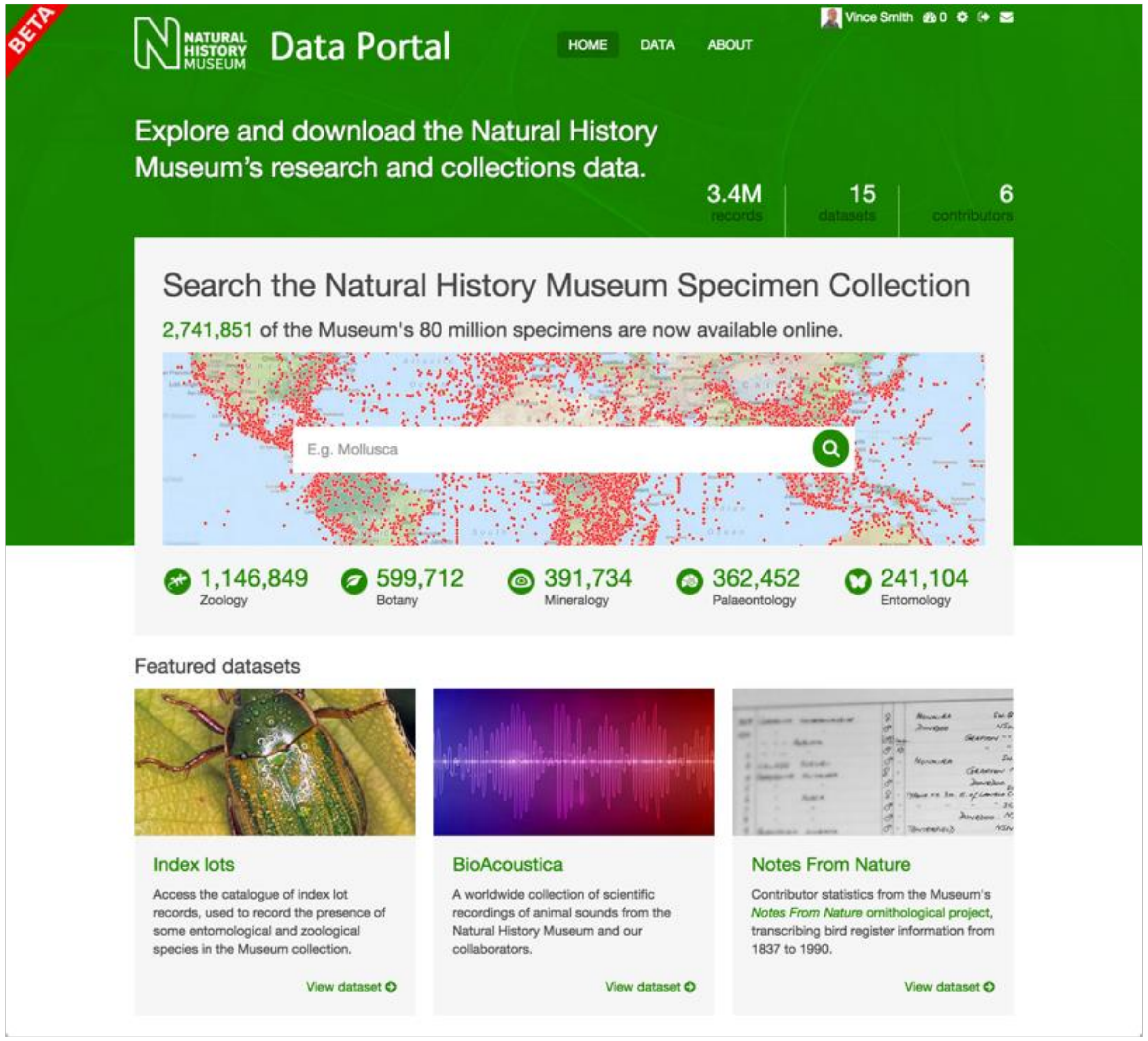




DEMO

NHM Data Portal

- A platform for deposition and discovery of NHM research & collections data
- Weekly updates from NHM collection database
- Stable, citable (DataCite) identifiers on datasets & records
- Transparent data quality indicators
- Export, import, & web-services available (Linked Open Data coming soon)
- Built using CKAN, with enhanced mapping, statistical and filtering functions



Explore and download the Natural History Museum's research and collections data.

3.4M records | 15 datasets | 6 contributors

Search the Natural History Museum Specimen Collection

2,741,851 of the Museum's 80 million specimens are now available online.

1,146,849 Zoology | 599,712 Botany | 391,734 Mineralogy | 362,452 Palaeontology | 241,104 Entomology

Featured datasets

Index lots

Access the catalogue of index lot records, used to record the presence of some entomological and zoological species in the Museum collection.

View dataset

BioAcoustica

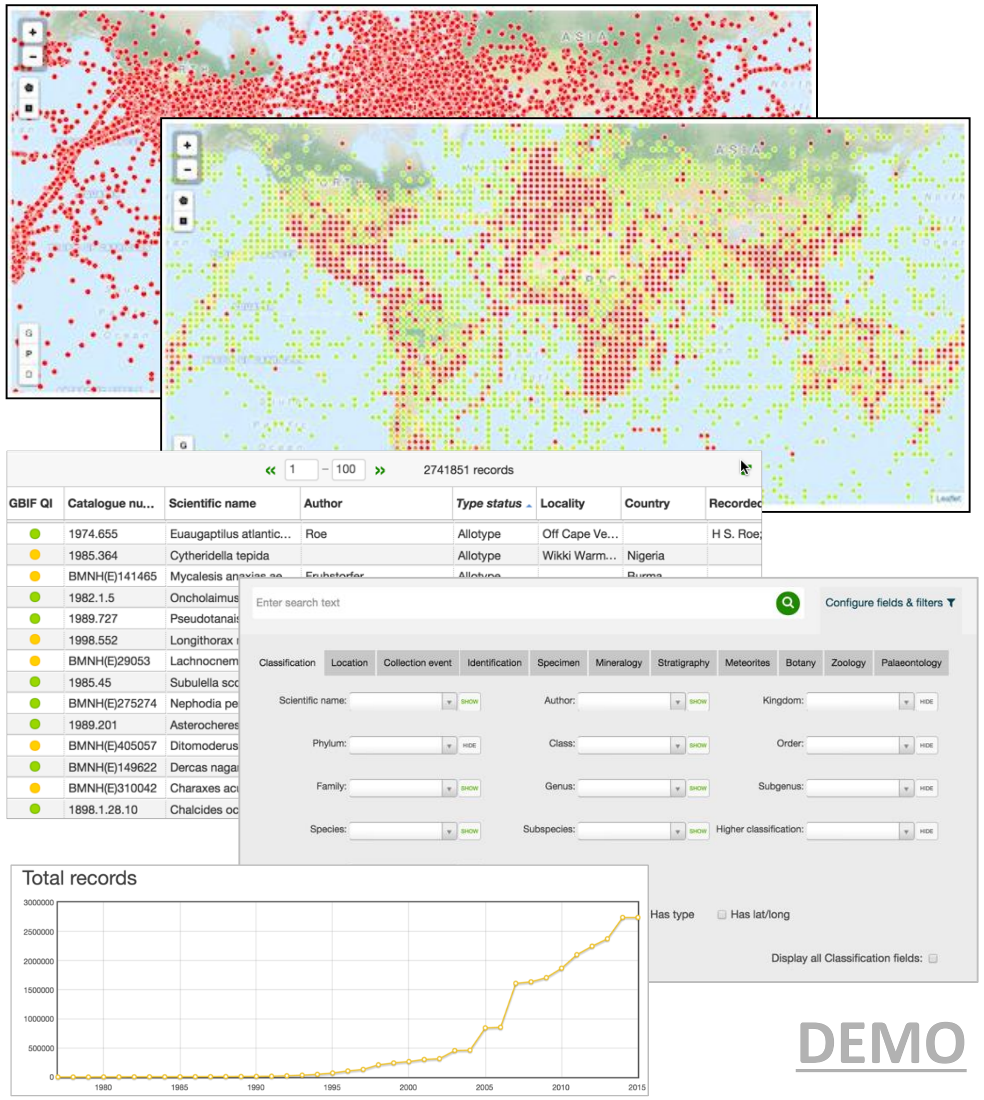
A worldwide collection of scientific recordings of animal sounds from the Natural History Museum and our collaborators.

View dataset

Notes From Nature

Contributor statistics from the Museum's *Notes From Nature* ornithological project, transcribing bird register information from 1837 to 1990.

View dataset



2741851 records

GBIF QI	Catalogue nu...	Scientific name	Author	*Type status*	Locality	Country	Recordec
🟢	1974.655	Euaugaptilus atlantic...	Roe	Allotype	Off Cape Ve...		H S. Roe;
🟡	1985.364	Cytheridella tepida		Allotype	Wikki Warm...	Nigeria	
🟡	BMNH(E)141465	Mycalesis anaxias ae	Fruhstorfer	Allotype		Burma	
🟢	1982.1.5	Oncholaimus					
🟢	1989.727	Pseudotanais					
🟡	1998.552	Longithorax					
🟡	BMNH(E)29053	Lachnocnem					
🟢	1985.45	Subulella scc					
🟢	BMNH(E)275274	Nephodia pe					
🟢	1989.201	Asterocheres					
🟡	BMNH(E)405057	Ditomoderus					
🟢	BMNH(E)149622	Dercas nagai					
🟡	BMNH(E)310042	Charaxes acu					
🟢	1898.1.28.10	Chalcides oc					

Scientific name: SHOW | Author: SHOW | Kingdom: HIDE

Phylum: HIDE | Class: SHOW | Order: HIDE

Family: SHOW | Genus: SHOW | Subgenus: HIDE

Species: SHOW | Subspecies: SHOW | Higher classification: HIDE

Has type □ Has lat/long

Display all Classification fields: □

DEMO

NHM Open Data Policy

Open-by-default policy on collections data & research data sets

- Creative Commons (CC) Zero waiver on data
- CC BY 4.0 licence on images

Exceptions to default

- Exceptions require justification, which include:
  - *Commercial value*
  - *Research competitiveness*
  - *Third party rights*
  - *Sensitive collection information*
  - *Donor or funder conditions*
  - *Confidential documents*

PSI Directive

- EU legislation on reuse of public sector information, extended to cover museums, libraries and public archives. To be implemented in July 2015
- General principal that information should be available for re-use for free or at marginal cost, and if reused, made available to all under equal terms.

NATURAL HISTORY MUSEUM