Digital Collections – towards a UK & European vision

Vince Smith, NHM London

Bristol Workshop - Towards a strategic approach to mobilising UK museum biodiversity data, M-Shed, UK

8 March 2018

Natural Science Collections

Bio- & geo-diversity

Leading scientific facilities

Engaging the public

Discovering, describing and interpreting life on Earth

Tackling societal challenges

Sustainable future

Molecular repository

NHM Digital Collections Programme (2014-2024)

"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

NHM Digital Programme

	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

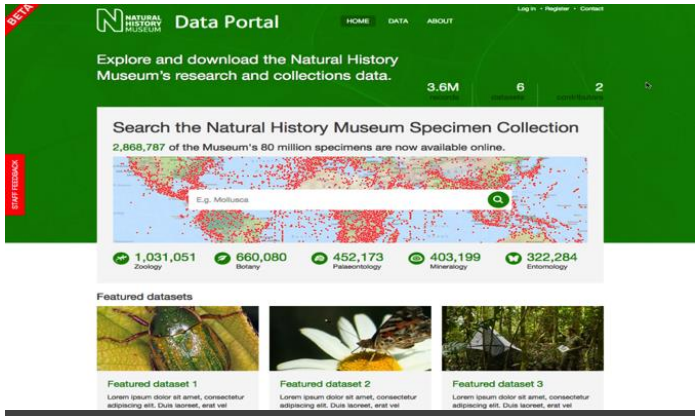
- 2, 5 & 10 yr. phases
- A change programme
- Addresses historic under investment
- Multiple digitisation projects
- Focus on innovation & science

NATURAL HISTORY MUSEUM

Pilots

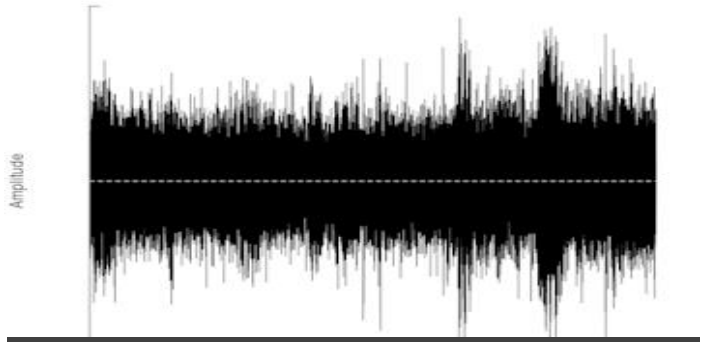
Crowdsourcing

Micro Collections



Informatics

Lepidoptera



Acoustics & 3D

Open Herbarium

Environmental Change

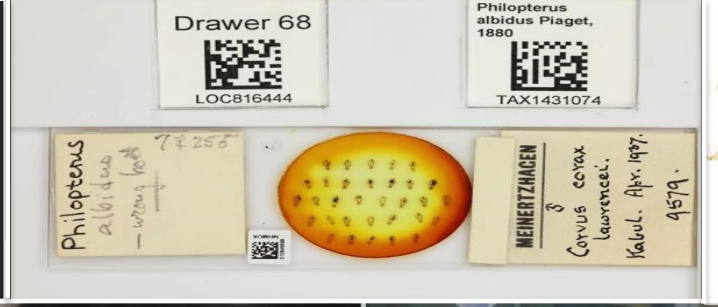
- UK butterflies & moths
- 800k specimens
- 2 mins per specimen
- £1 per specimen

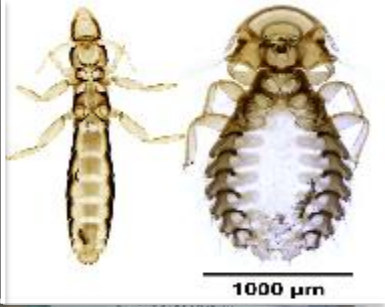
Large-scale digitisation:

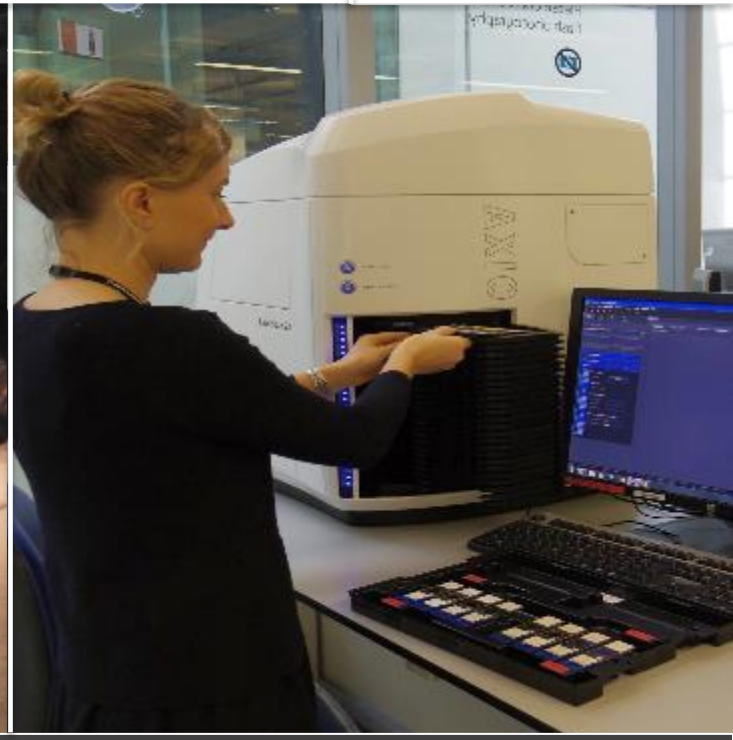
- High-throughput digitisation workflows
- Informatic pipelines
- Computer-assisted object recognition

Parasites & Vectors:

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



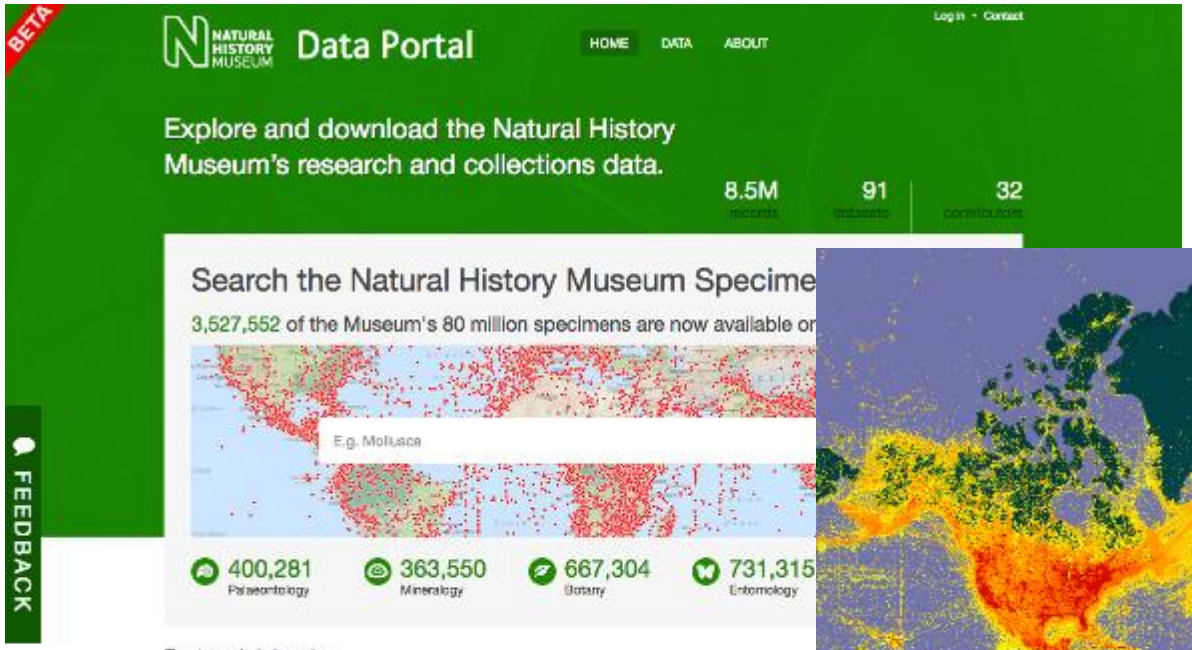


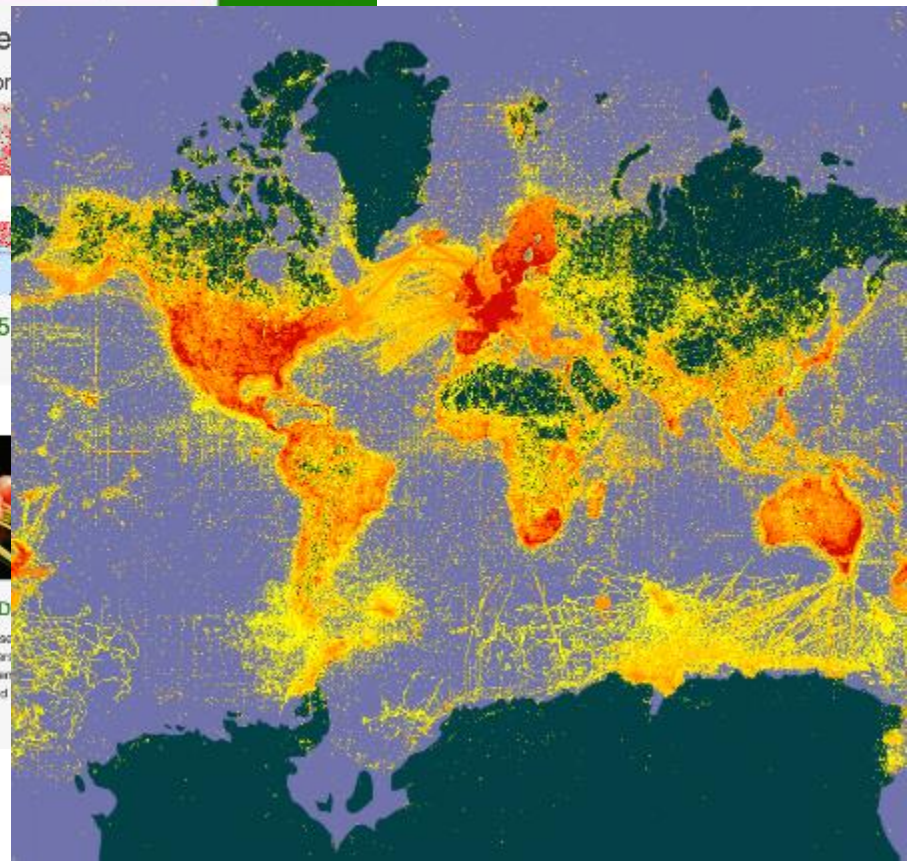


Parasitic lice (new spp. & cospeciation)

- 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 Portal





- NHM research & collections data
- Access, reuse & citation
- 9.1m records, 95 datasets since 2015
- Images, sound, video & 3D
- Default open licensing
- GBIF
- 10B records downloaded in 139k events, 61 papers

Impact

Impact

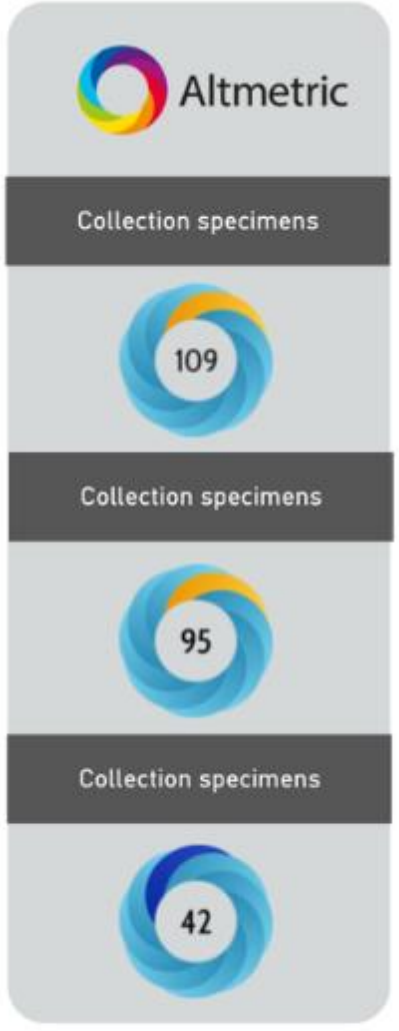
Show publication list

Publications citing NHM data (GBIF only)

GBIF topic tags by frequency

59	50	22
Publications	Peer-reviewed	Open Access

Count of citations by source



* **Altmetrics** are non-traditional metrics used to measure the impact of a scholarly work. The 'donuts' above mainly reflect tweets referencing an NHM data DOI.

NATURAL HISTORY MUSEUM

Challenges

1. Mass Digitisation

Specimen image

2. Data Extraction

Transcription
(e.g. Zooniverse)

OCR
(e.g. reCAPTCHA)

Georeferencing
(Online mapping tools)

Image Recognition
(e.g. Image Search)

3. Using the Data

Linking to Archives and Literature
(e.g. Biodiversity Heritage Library)

Analytical Tools
(e.g. OpenRefine)

Data Visualisation
(e.g. CartoDB)

Search
(e.g. WikiData)

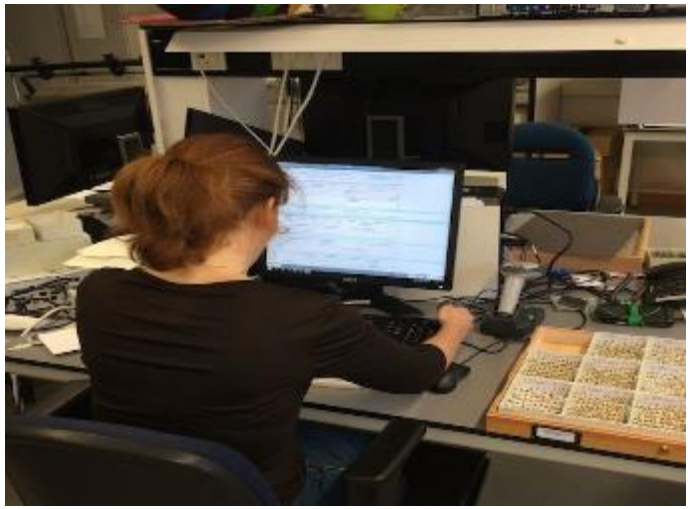
4. End Products

NHM in an App

Digital Exhibitions
(Virtual Tours)

NATURAL HISTORY MUSEUM

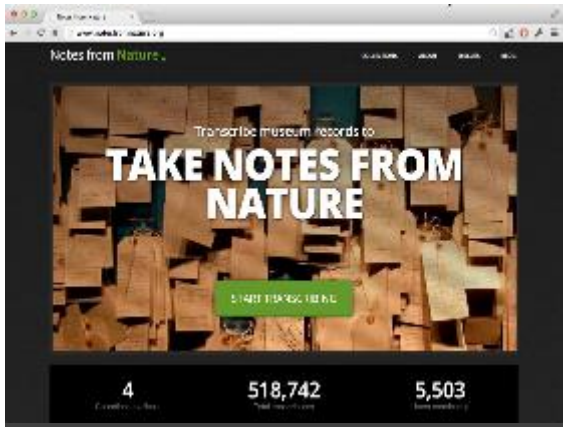
Workflows



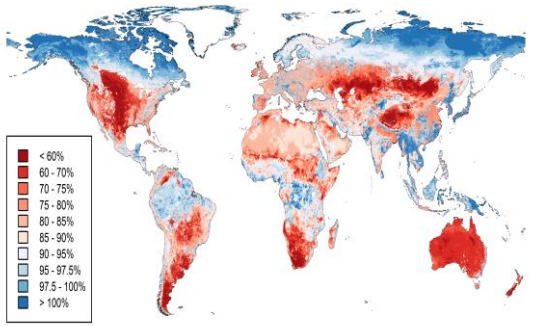
NATURAL HISTORY MUSEUM

Research

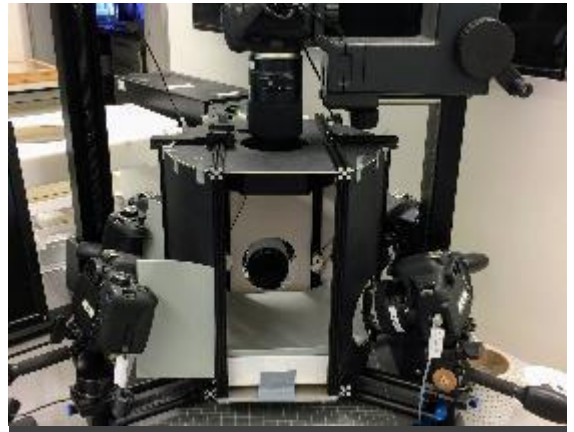
Environmental Change



Citizen science

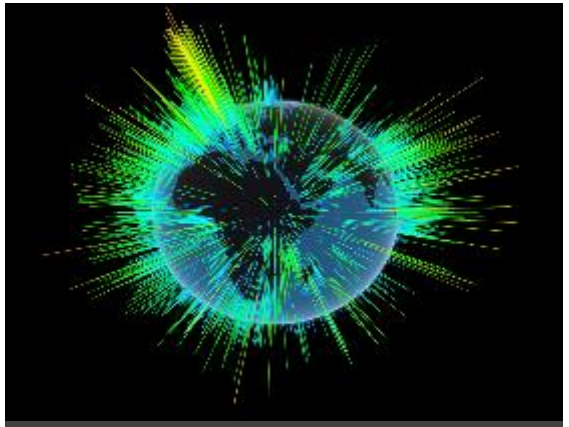


Measuring biodiversity loss

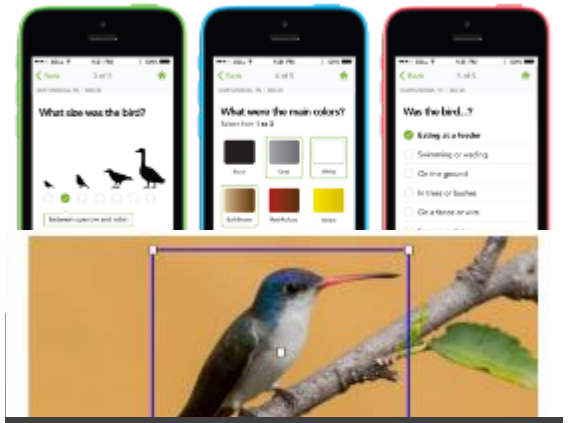


Digitisation workflows

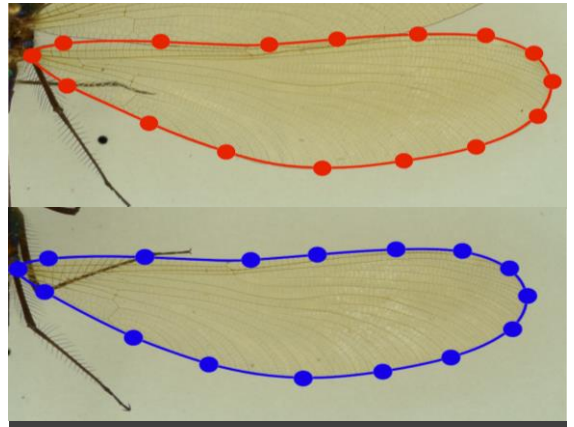
Measuring hyper diversity



Data visualisation



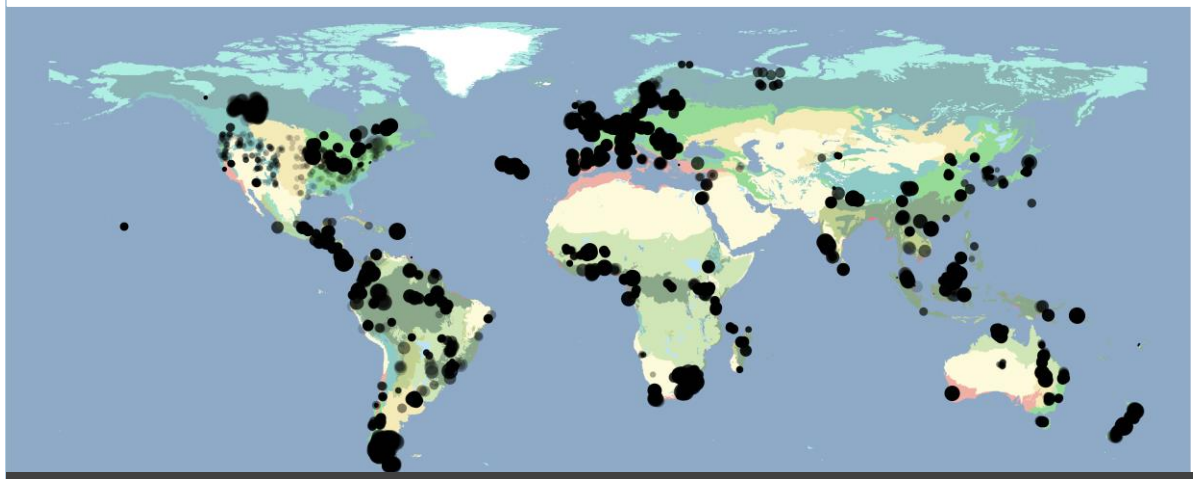
Species recognition



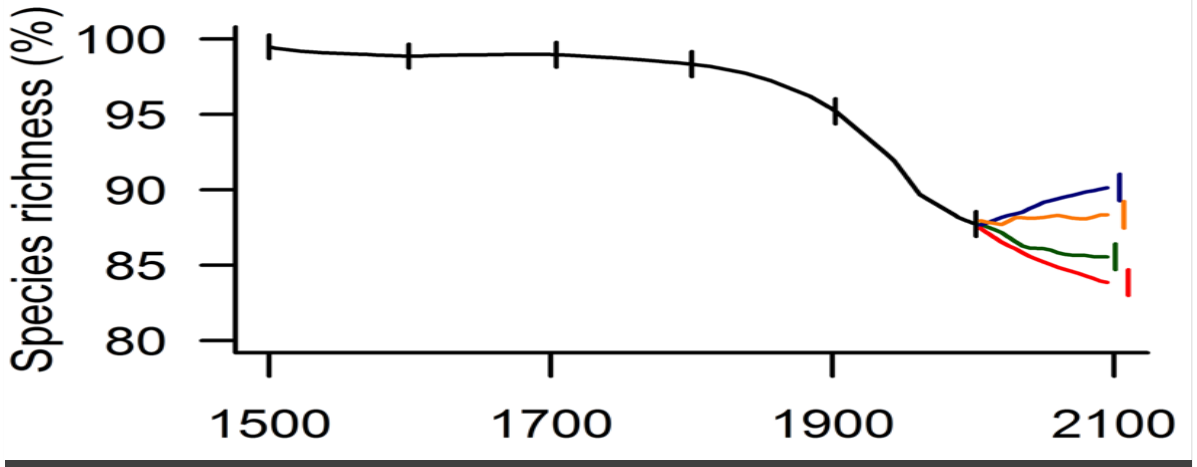
Computer vision

NATURAL HISTORY MUSEUM

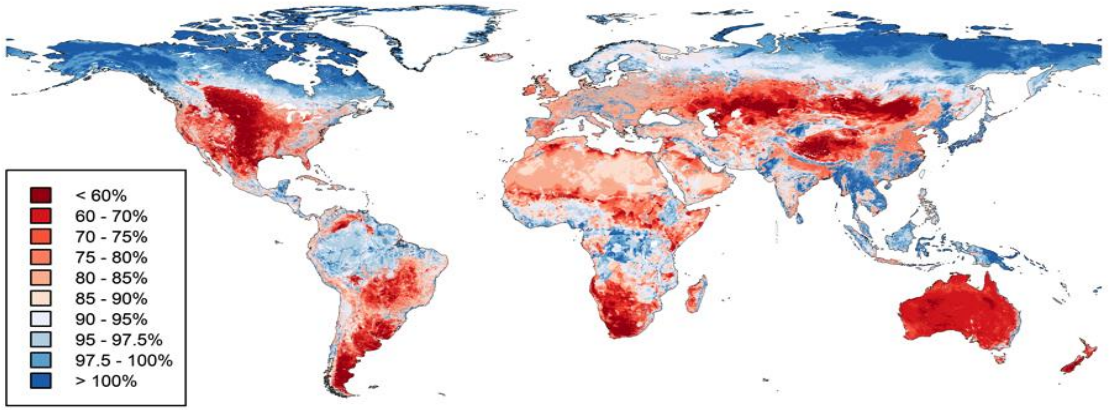
Big Data



Database (biodiversity, land use & human population)



Projection (predicted change to species richness)



Model (Biodiversity intactness Index – BII)

Newbold et al. 2016
Science 353:288-291

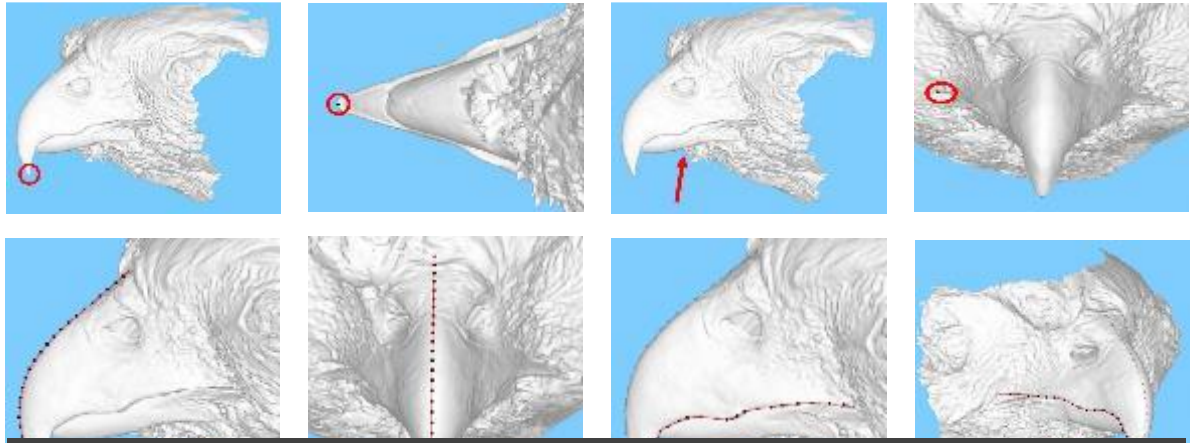
NATURAL HISTORY MUSEUM

Engagement

Wired

Project Plumage

Visiteering



Mark-My-Bird

NATURAL HISTORY MUSEUM

Europe

55% of the world‘s collections with rich historical & global distribution

> **1.5 billion** specimens

> **80%** of world’s species

> **5,000** scientists employed

> **16,000** scientific visitors pa

> **10 million** public visitors pa

> **25 million** web visitors pa

NATURAL HISTORY MUSEUM

Projects

Major digital initiatives

SYNTHESYS
Synthesis of systematic resources

ViBRANT
Virtual Biodiversity

eMonocot

EU BON

OPEN UP!

EDIT

Major funders

European Commission

NERC SCIENCE OF THE ENVIRONMENT

Alignment with organisational agendas

Biodiversity Information Standards
TDWG

CETAF

RDA
RESEARCH DATA ALLIANCE

Major collaborators

Catalogue of Life

GBIF

eol

GGBN

- 13-year programme of physical & digital access
- SYNTHESYS1,2,3 (>50k access days, 3.8k sci., 4.5k pubs.)
- Focus on science, policy and public engage. activities
- A unique role within the infrastructure landscape
- Foundation for deeper integration

NATURAL HISTORY MUSEUM

DiSSCo

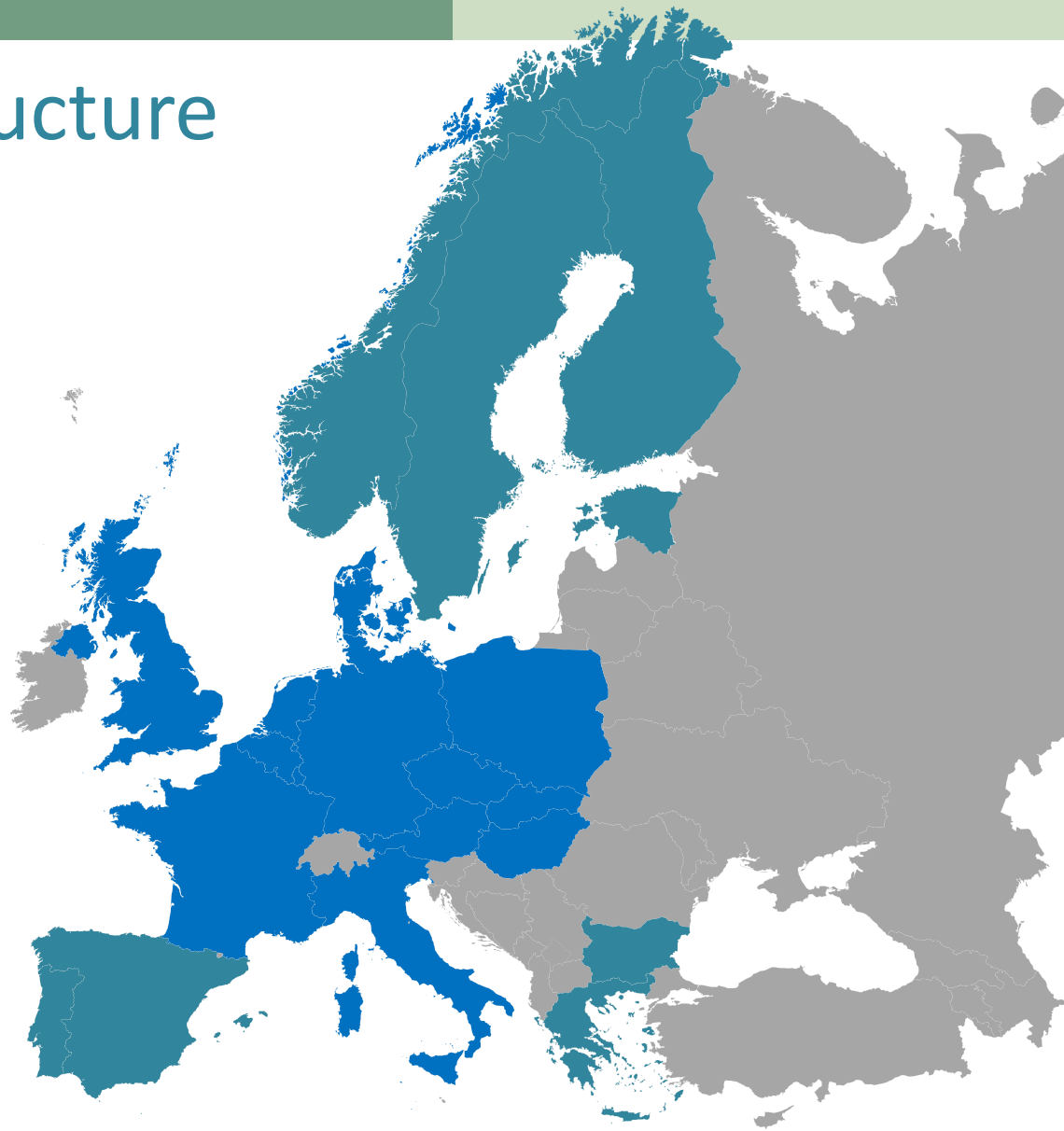
DiSSCo: A new European infrastructure

114 National Facilities

21 Countries

- Largest ever formal agreement between natural science collection facilities
- Centralised governance model already in place
- Supporting network of working groups

With political support by 11 European Governments



NATURAL HISTORY MUSEUM

Services

1 e-Science services

A one-stop shop for services providing unified **discovery, access, interpretation and analysis** of complex linked data

2 Physical and remote access services

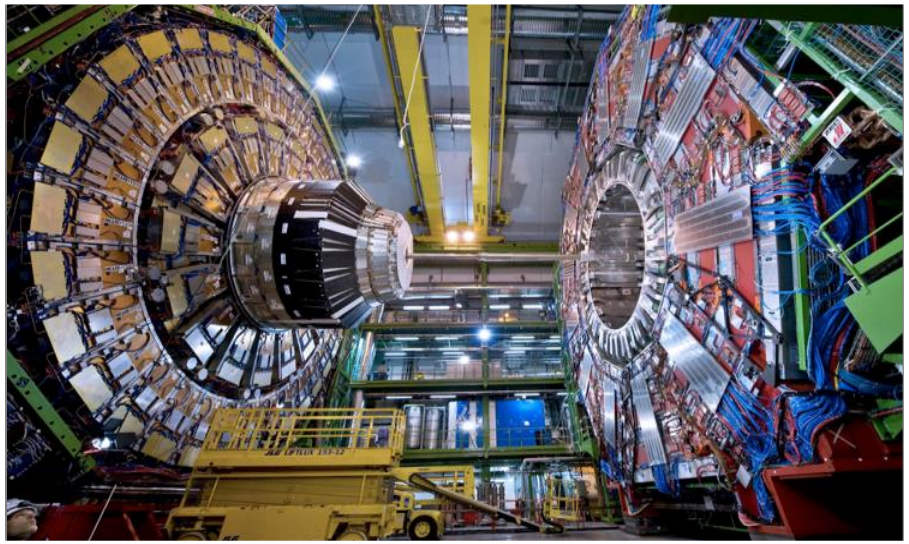
A universal harmonised **physical access service** and **digitisation on demand** service. **Supporting mass-scale, yet synchronised content mobilisation across European Collections**

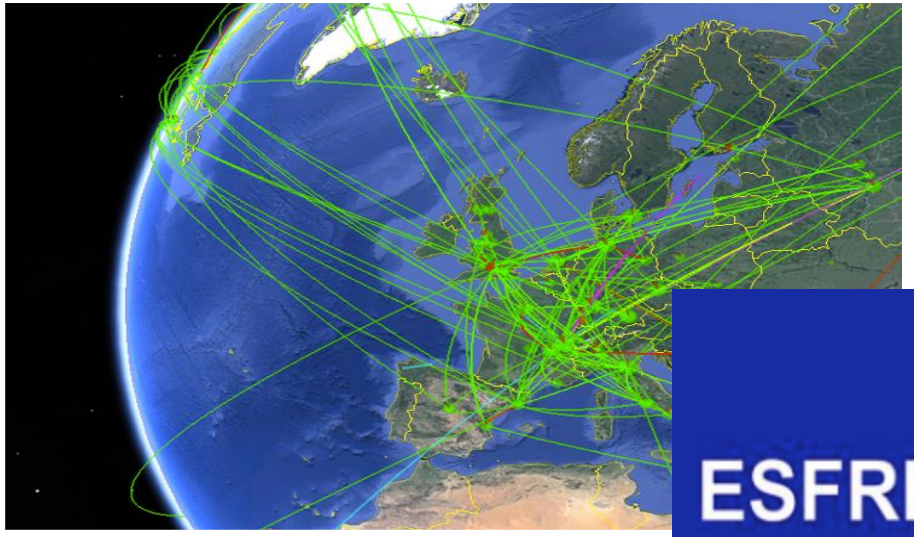
3 Support & Training services

Integrated **user support desk** and implementation of **multi-modal training programmes** to enhance data skills

NATURAL HISTORY MUSEUM

ESFRI

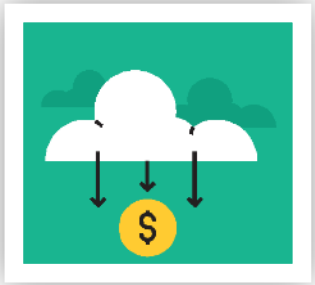




- Support a coherent approach to policy-on European research infrastructures (RI's)
- Facilitate initiatives leading to better use & development of RI's
- Complex, two-year roadmap admission process
- Two delegates per country, nominated by the national research/science agencies.

NATURAL HISTORY MUSEUM

Costs



Design Study

€10.7M

Resourcing

EC funding
Consortium investments

2018

Preparatory & Construction

€89.6M

EFSI
Consortium (in-kind)
Member State contributions

H2020 (CSA)
COST Action
Consortium (in-kind)
Member State contributions

H2020 (RIA)
Public Private Partnerships
Consortium (in-kind)
Member State contributions

2025

Operational

€12.1M / year

(plus additional digitisation investments)

European Investment Bank
EFSI
Consortium (in-kind)
Member State subscriptions

NATURAL HISTORY MUSEUM

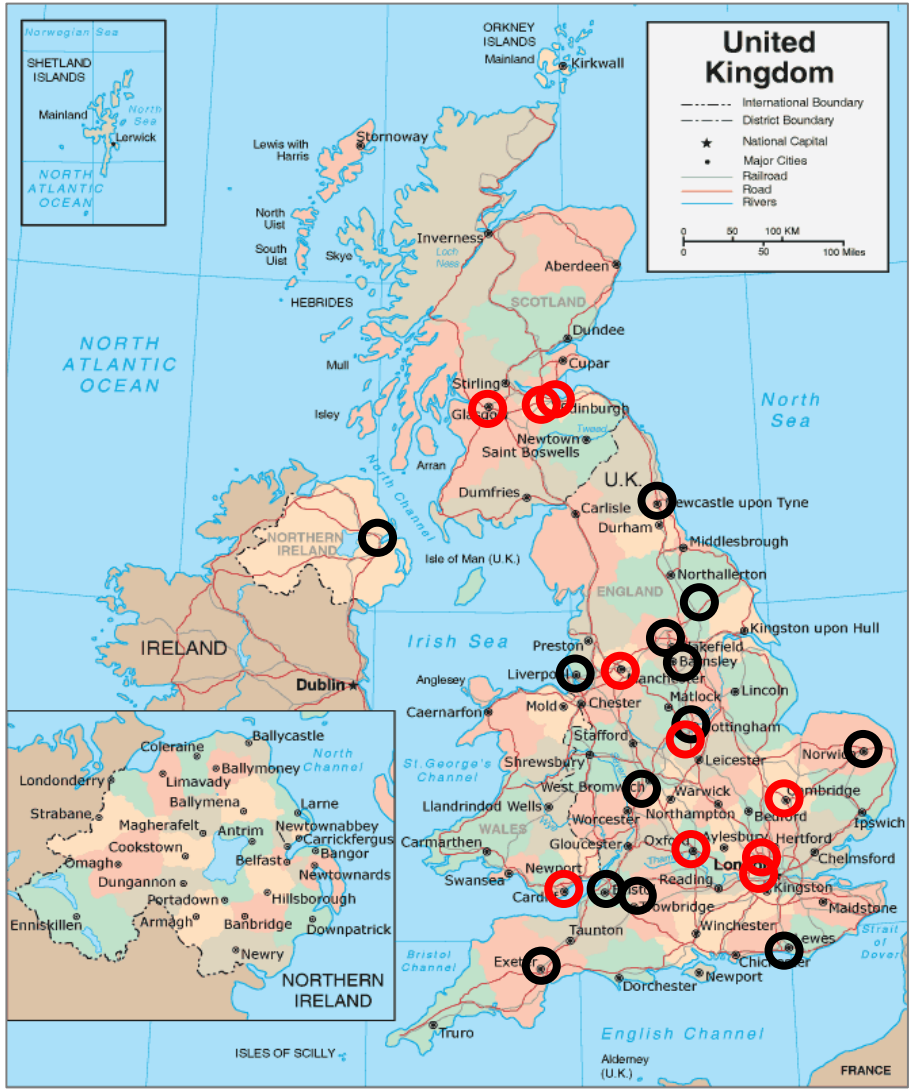
An urgently needed initiative

- Official ESFRI decision due June 2018
- If successful, DiSSCo is open to new organisations
  - Opportunity for leadership at a global level
  - Direct response to identified needs in the European and international RI landscape
  - Lowers the barrier for big, open science practices across tens of thousands of users

Potential to improve the case for support for UK collections?

NATURAL HISTORY MUSEUM

UK



- Potential to increase UK coordination on:
  - Regional collections
  - Observer networks
  - Technical support
  - Citizen science
  - Education
  - Others?
- What might the “One-Collection” vision mean for the UK?
- Can we be more than the sum of our parts?

Digital Communities

- Community-led projects
- Volunteer network
- Edu. prog. & tools
- Citizen science

Digital innovation

- Open data portal
- Mobile apps
- 3D imaging
- Digitisation pipelines

Focal collections

- Plants
- Fossils
- Butterflies and moths