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**

A Contraction

We Kit ..



### **Collections that change how we think**



### **Collections that inspire the next generation**



#### **Collections at scale**



# NHM Collection – 2015/16

		Collection	% specimens	# sci. visitor	# loaned
Dept. / Division		size	databased	days	specimens
Ea	arth Sciences				
	Mineral & Planetary Sciences	212,600	95%	165	52
	Envt & Economic Earth Sciences	212,000	95%	105	51
	Invertebrate & Plant Palaeobiol.	7,000,000	1 0%	705	183
	Vert Palaeobiol & Anthropology	1,064,000	4.970	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	1 70/	2,932	716
	Invertebrates	24,403,000	4.270	828	559
T	OTAL	76,101,600	4.50%	7,303	32,784
		/			
		New opportuni	ties	Limited potent	tial 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.



# **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

	artell a sal car			
in the second second		<b>«</b> 1	- 100 >> 274	11851 records
	GBIF QI Catalogue nu	Scientific name	Author	Type status . Locality Country Recorded
A CONTRACTOR OF THE REAL OF	91974.655	Euaugaptilus atlantic	Roe	I state of the
	9 1985.364	Cytheridella tepida		
	BMNH(E)141465	Mycalesis anaxias ae	Fruhstorfer	
在了,并不过这些的 <b>是</b> 外的问题。	9 1982.1.5	Oncholaimus qingda		
新生产的,在14.4%的影响。在14.4%的影响。 14.4%的影响。	9 1989.727	Pseudotanais falcicula		
· · · · · · · · · · · · · · · · · · ·	998.552	Longithorax megalops		
· · · · · · · · · · · · · · · · · · ·	<ul> <li>BMNH(E)29053</li> </ul>	Lachnocnema angola		
A STATE OF A STATE	9 1985.45	Subulella scotti Holdi	Holdich and Bird, 1986	
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	BMNH(E)275274	Nephodia perimede	Druce	
	9 1989.201	Asterocheres bulbos	Malt	
	BMNH(E)405057	Ditomoderus lacroixi	Bomans, 1973	
	BMNH(E)149622	Dercas naganum Tytler	Tytler	and the second se
	BMNH(E)310042	Charaxes acuminatus	van Someren, 1963	
	0 1898.1.28.10	Chalcides ocellatus s		

### Services supporting data quality indicators



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





Outreach



Policy



Government



Training



**Events** 



Volunteers



Monitoring & KPIs



Team communication



Team organization



Innovation



Citizen science



Biq data



Peers



# **Creating a sustained digital culture**

#### Normalising 'digital' within NHM London







Policy



Government



Training



**Events** 



Monitoring & KPIs



Team communication



Team organization



Volunteers



Innovation









Biq data

Peers

# Innovation in image processing

Computer vision to detect, sort & process specimens



- 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



# Innovation in image processing

Computer vision to extract label data







1. Label identification & perspective warp



Cheshire:	Cheshire:
Redesmere,	Red smere,
Chelford,	Che ford.
viii-/, 1935.	viii-ix. 1935.
A.D. Grensted,	A.D.Grensted.
Cheshire.	Cheshire:
Redesmare,	Redesmere,
Chelford.	Chelford,
viil-ix 1935.	viii-ix 1927
A.D.Grensted.	A.D.Gree 1.



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







#### Trait detection



Sex brands in Hesperia comma

### **Specimen data refinery**

Linking processes, creating enhanced records & triaging tasks



# **Citizen science - Transcription**



transcriptions

#### A partnership with Notes from Nature



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



#### Data on transcribers



# **3D Scanning of Bird Beaks**

#### Citizen science to uncover the secrets of beak evolution



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



Crowdsourced beak landmarks

#### LETTER

#### Mega-evolutionary dynamics of the adaptive radiation of birds

Christopher R. Cooney<sup>1</sup>e, Jen A. Bright<sup>13,3</sup>e, Elliot I. R. Capp<sup>1</sup>, Angela M. Chira<sup>2</sup>, Emma C. Hughes<sup>1</sup>, Christopher J. A. Moody<sup>1</sup> Lura O. Navel<sup>1</sup> Zoh K. Varlee<sup>1</sup> & Gazin H. Thornac<sup>1,4</sup>

The origin and expansion of biological diversity is regulated by both developmental trajectories <sup>1,2</sup> and limits on available evolutical	to specifically assess the macroe relevant traits. Here we study the
and a recognition in a potential and infinite on available ecological	the value of any field we study the
nicnes . As timeages diversity, an early and often rapid phase of	tran (pui scape) across an entire-
species and trait proliferation gives way to evolutionary slow-	the processes shaping the accum
nowith as new species pack into ever more densely occupied regions	a goost-scale adaptive radiation.
of ecological niche space". Small clades such as Darwin's finches	Our approach is based around
demonstrate that natural selection is the driving force of adaptive	shape. The avian but is closely
radiations, but now microevocutionary processes scale up to snape	toraging ruches and represe
the expansion of phenotypic diversity over much longer evolutionary	known to have a key role in cla
timescases is uncear. There we address this problem on a global scale	we took 3D scans or museum st
by analysing a crowoscourced dataset of infree-dimensional scanned	(397% of estant general) repre-
bill morphology from more than 2,000 species. We find that bill	diversity. We placed morphologic
diversity expanded early in extant avian evolutionary history, before	on bills (Extended Data Fig. 1) us
transitioning to a phase dominated by packing of morphological	(http://www.markmybird.org) at
space. However, this early phenotypic diversification is decoupted	logical space (morphospace) of e
from temporal variation in evolutionary rate: rates of bill evolution	position and principal compone
vary among lineages but are comparatively stable through time. We	first eight principal component
tind that rare, but major, discontinuities in phenotype emerge from	variation in tell shape (Fig. 1). I
rapid increases in rate along single branches, sometimes leading	describes the volumetric aspect
to depauperate clades with unusual bill morphologies. Despite	sword-billed hummingbird, Ensil
these jumps between groups, the major axes of within-group	large ground linch, Geospiza #
bill-shape evolution are remarkably consistent across birds. We	of shape variation encompasse
reveal that macroevolutionary processes underlying global-scale	(length, width and depth). Var
adaptive radiations support Darwinian" and Simpsonian' ideas of	relate to fine-scale division of t
microevolution within adaptive zones and accelerated evolution	closely related species, but can
between distinct adaptive peaks.	observed among extant birds. N
The role of adaptive radiations as the source of much of the world's	of total variation) are explained
biological diversity has been widely emphasized	retain high phylogenetic signal (
clades have provided insights into the role of natural selection as a	although these higher shape an
diversifying force, but cannot illuminate the processes that shape the	variance, they capture large differ
diversity and discontinuities of radiations over longer evolutionary	of bill shape. The narrow (long ta
time frames. Indeed, at large taxonomic scales, the diversification	compared to the broad distribu-
of clades".12 and traits" shows no evidence of the predicted slow-	Extended Data Table 1), sugge
downs in evolutionary rates, despite there being numerous examples in small clades <sup>3,24-16</sup> . This arearent paradox is potentially resolved by	relatively simple bill shapes and of bill morphospace.
G. G. Simpson's model, in which major jumps to new adaptive zones	We tested an important pred
('quantum evolution') can occur unpredictably throughout clade	ating how niche expansion and n
history. These jumps give rise to rapid lineage expansion into previously	accumulation of bill-shape disp
unoccupied niche space as sub-clades continue to radiate within dis-	history. We estimated multiva
tinct adaptive zones and subzones". Simpson's models introduced the	ancestral state estimates derived
concept of 'mega-evolution'-diversification over large temporal and	trait evolution13 (see Methods)
eputial scales-unifying microrvolution with other factors such as	we calculated disparity as the s
ecological opportunity and evolutionary constraints that shape the	eight shape axes. We compare
macroevolutionary trajectories of radiating lineages. However, although	with two null models-constant
phylogenetic studies involving thousands of species have demonstrated	heterogeneous trait evolution-
heterogeneity in rates of phenotypic evolution 13.37, it is unclear whether	niche-filling processes (see Met
the processes outlined by Simpson have an important role in large-scale	tations, we find that the filling
adaptive radiations. This is because previous studies have been unable	time shows a notable dominant
Department of Annual and Part Sciences Conversion of Reeffacts Diselbers 532 27% DX Technol	of of Gaussiannes, University of Bouth Platitie 3
and a second sec	C. 185. March Concer Department of the Engine



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



# **Halting biodiversity loss**

Digital collections to develop an index of planetary health



# **One Global Digital Museum**

We need a knowledge platform for a 21<sup>st</sup> Century Museum



### **Distributed System of Scientific Collections**

Transforming European NH institutes into an integrated research infrastructure





