

Making molehills out of mountains

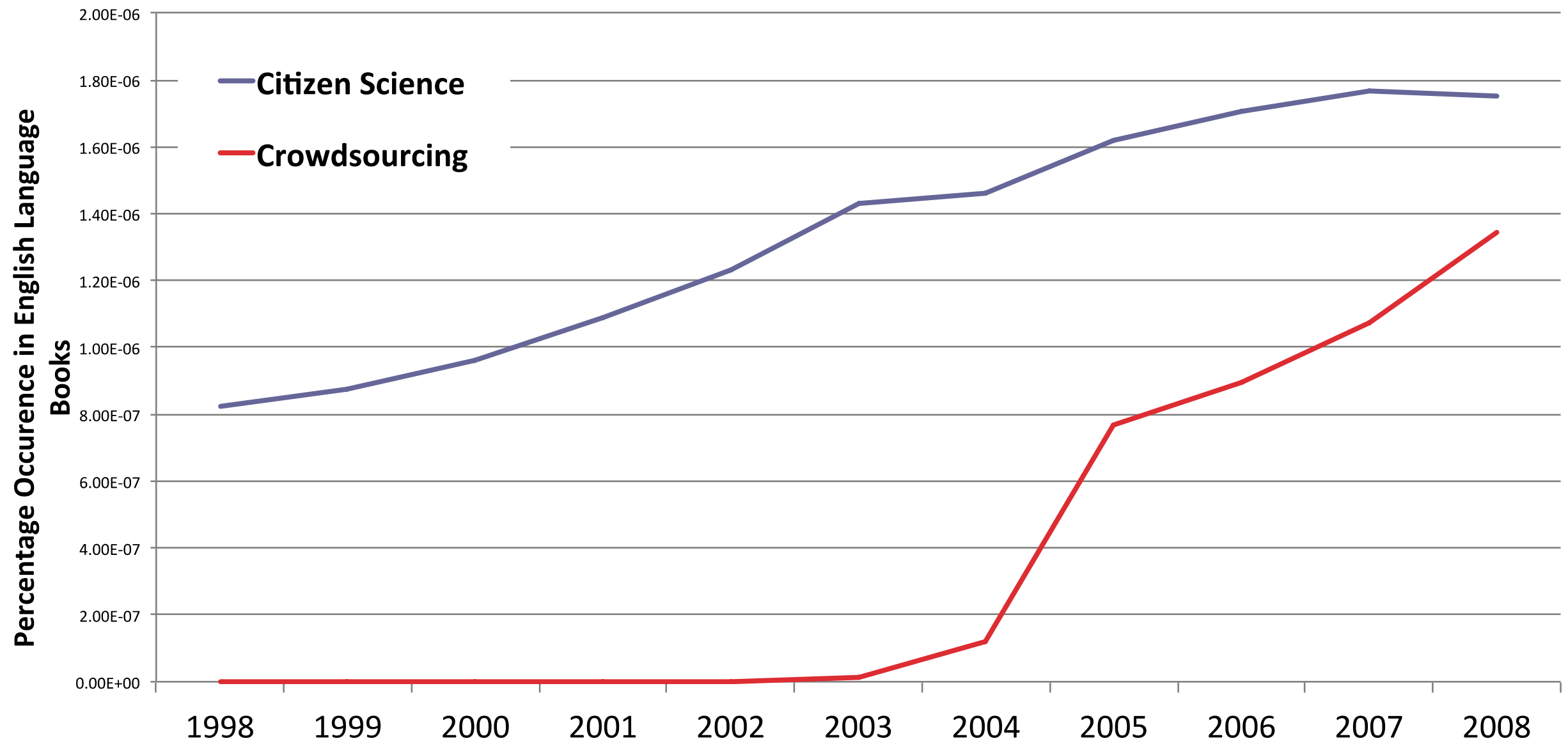
Crowdsourcing digital access to natural history collections

Laurence Livermore

Vincent S. Smith and John Tweddle

29th SPNHC Annual Meeting – June 2014

Google Ngram of "Citizen Science" and "Crowdsourcing"



Data from Google Ngram Viewer, <https://books.google.com/ngrams> Search settings: phrases "crowdsourcing" and "citizen science"; between 1998-2008; from the corpus English;

What is crowdsourcing?

- Lack of consistent definition but generally¹:
 - Involves a **crowd**
 - Has a **task** and **goal**
 - **Rewards** the crowd
 - Has a distinct **crowdsourcer**
 - **Benefits** the crowdsourcer
 - Is an **online** and **open participatory** process

1. Estelles-Arolas, E., and F. Gonzalez-Ladron-De-Guevara. "Towards an Integrated Crowdsourcing Definition." *Journal of Information Science* 38.2 (2012): 189-200

Why is it relevant to Natural History Collections?





Mantle et al (2012) **Whole-drawer imaging for digital management and curation of a large entomological collection.** *ZooKeys* 209: 147–163, doi: 10.3897/zookeys.209.3169

Explore, visualize and download
the Natural History Museum's collection
and research data.

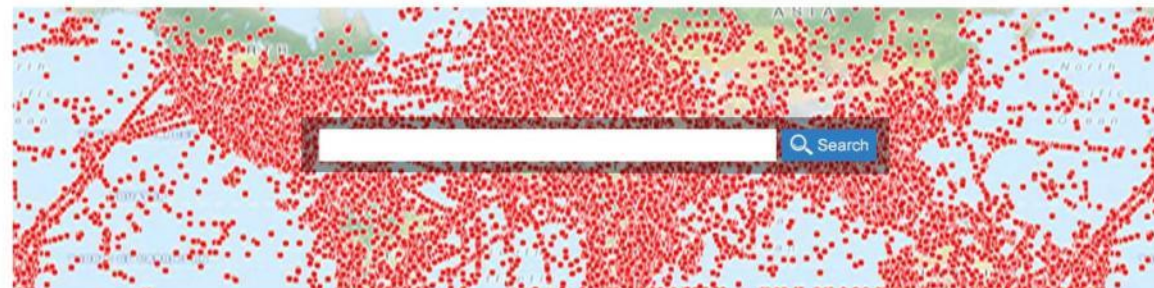
1000000
Records

100
Datasets

1000
Users

Search the Natural History Museum Specimen Collection

The Natural History Museums 3,547,342 specimens are now available online.



648,367 Botany 334,241 Paleontology 402,777 Mineralogy 1,169,583 Zoology 992,324 Entomology

Featured datasets

Featured datasets



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent sit amet nibh rutrum, iaculis dolor id, cursus quam. Nam eu odio nibh. Aenean nibh tortor, lobortis nec fermentum eu, ornare ut quam. Etiam lobortis, enim ac dictum sodales, sapien leo imperdiet lacus, ac congue odio massa ut tellus.

[Read more](#)

Featured datasets



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent sit amet nibh rutrum, iaculis dolor id, cursus quam. Nam eu odio nibh. Aenean nibh tortor, lobortis nec fermentum eu, ornare ut quam. Etiam lobortis, enim ac dictum sodales, sapien leo imperdiet lacus, ac congue odio massa ut tellus.

[Read more](#)

Featured datasets



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent sit amet nibh rutrum, iaculis dolor id, cursus quam. Nam eu odio nibh. Aenean nibh tortor, lobortis nec fermentum eu, ornare ut quam. Etiam lobortis, enim ac dictum sodales, sapien leo imperdiet lacus, ac congue odio massa ut tellus.

[Read more](#)

Use the data



Explore

Explore the Museum's [x] collection
and research data.

[Start exploring](#)



Download

All datasets are available to
download directly or via an API.

[Find out more](#)



Cite

Cite Museum's data in your research
all datasets have a DOI.

[Find out more](#)

MUSEE DU CONGO
Madagascar
Montagne d'Ambre
Coll. Schouteden

Madagascar
Cap d'Ambre

Anopheles
madagascariensis
Sign.

♂

Specimen Photographed

<http://coreoidea.speciesfile.org>

MRAC-CSF_005

HOLOTYPE

雲南大勐龍 700 米
1957.IV.11
蕭富基等

Юньнань. Дамонлун,
700 м. 11. IV. 1957.
Пу Фу-ди и др.

HOLOTYPE
Homoeocerus
(A.) *impictus* Hsiao

IOZ(E) 222017

Coll.
Bretfordin

Lectotypus.

96

abola

1

ст. Хибины,
Кольск. п-ов
В. Старк 26 VII 928

Paratypus
...icus

HOLUTIPO

ARGENTINA:
TUCUMAN
SIAMBOUR
I-45

Brailovsky A. det.
AUMASTUS
BR Brailovsky

HEMI-COR-0001

HOLOTIPO

NEW GUINEA: SE
Northern District
Cape Killerton
2-12.V.1965

W.A. Steffan
& Y.M. Huang
Collectors
BISHOP

H. Brailovsky A. det.
CNEIUS
KILLERTONENSIS Brailovsky

HOLOTIPO

COLOMBIA: Sierra
Nevada de Santa
Marta, San Sebastian
de Rabago, 2000 m.
April 1-14, 1968

Borys Malkin
Collector

H. Brailovsky A. det.
ANASA MICANS
Brailovsky

Ruwenzori,
Scott Elliot
5600 ft.
95-41.

Acanthomia
insignis
type Dist.

♂

Type

Holo-
type

BMNH(E) 88433

Imaged for:

<http://coreoidea.speciesfile.org>

HOLOTIPO

COLOMBIA: Sierra
Nevada de Santa
Marta, San Sebastian
de Rabago, 2000 m.
April 1-14, 1968

Borys Malkin
Collector

H. Brailovsky A. det.
ANASA MICANS
Brailovsky

Borneo
Xantus

Colpura
amillata
F. v. Br. det.

CSF_HNHM 096

Photographed for:

<http://coreoidea.speciesfile.org>

dissimilis
Hsiao
正模標本編號 1963

編號
採集地 杭州
採集期 IV. 1958
採集人

Typhlocolpura
vulcanalis
Bergr. - Type

Mindanao
Mt Apo PI
June-July
E.A. Mearns

Type No.
20147
U.S.N.M.

Holo-
type

F.G. DRECHER
G. Tengchong District
4000-5000 Meters
Greenland, Java

HOLOTYPE
Grallidava
det. *solitaria* n.sp.
W.R. Dolling 1973

Mydas lateralis
Grm. it. Scranii
Det. Burm.
Lupitan. y Dahmet

Rio
v. Olfers
Nr. 1166

• Typus

SYN-
TYPE

Mexico
Coll. Signoret.

perfidus
det. Stal

Holotype
madura
perfidus f.
Det. by Stal
T.R. Yonke, 1985

Syntype
Madura
perfidus Stal

Type Specimen Photographed
CSF_NHMH_041
<http://coreoidea.speciesfile.org>

ZMUC 00 102308

Amer. merid.
Schmidt
Miss. Ford-Lund
Lygus
bilineatus
L. & Fabr.

Disctor bilineatus
(Fabr.) det. E. Osuna
8-11-1970

Type specimen imaged for:
<http://coreoidea.speciesfile.org/>

TYPE.
WESTW. (HOPE)
C. Hemipt. 1842
Part II, p. 7
Distant, P.Z.S.,
1901, p. 325-335.

Spartocera
Set. A
Burm. 342

132
C

Type

Sephina
erythromelana
White

TYPE HEM: 273
SPARTOCERUS
BIMACULATUS
WESTWOOD
HOPE DEPT. OXFORD

Coll. R.
Pérou:
ex Coll.

Specimen Photographed
<http://coreoidea.speciesfile.org>

Clavigralla gibbosa, Spin.
D. Dupont
Bombay.

Fot. VIDANO-IV-1973
per DOLLING (G.B.)

Coll. Spinola box 28
Heteroptera

LEOTO-
TYPE

♀

Specimen Photographed
<http://coreoidea.speciesfile.org>

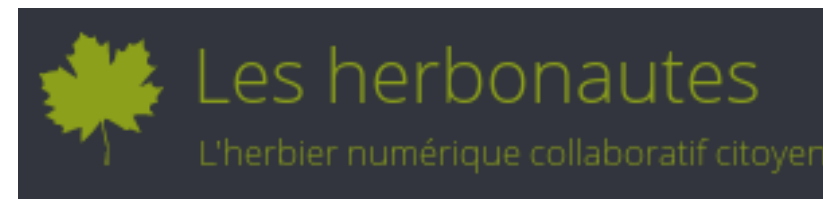
E.P. van Duzee
Collector

HOLOTYPE
...us

California Academy
of Sciences
Type No. 994

What platforms have been developed?

herbaria@home



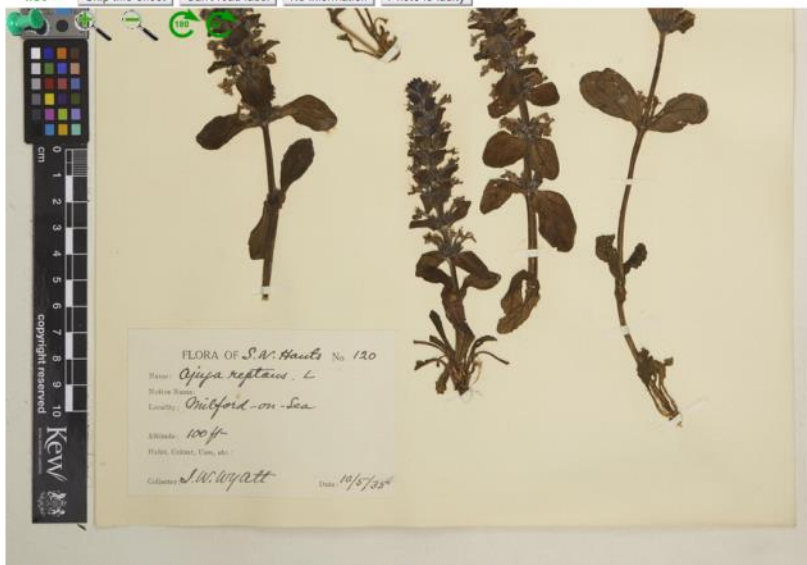
Herbarium at Home

- Herbarium sheet transcription platform
- Pioneering NH crowdsourcing tool – launched in 2007
- Established and experienced community
- 135k transcriptions, community of ~ 400
- Two stage entire field documentation/review

Herbarium sheet form

The sheet will have one or more labels or annotations, please copy the details from these into the form below, then click 'send' to save when you have finished. To move or zoom the image click on it with the left mouse button. Clicking and dragging moves the image. To zoom out alt-click or use the '-' icon at the top left. Click on the thumbnail image at the right to jump around the sheet. The 'drawing-pin' button detaches the image into a separate window. To move between sheets in your allocated set use the '<' and '>' buttons. 'list' shows overview of the sheets allocated to you.
[View documentation guidelines](#)


list > Skip this sheet Can't read label No information Photo is faulty





FLORA OF S.W. Hants No. 120
 Name: *Ajuga reptans* L.
 Native Name:
 Locality: *Orford-on-Sea*
 Altitude: *100 ft*
 Habitat, Culture, Use, etc.:
 Collector: *S. W. Hants* Date: *19/5/35*

filing note: Please document only the specimen bar-coded as **K001015260** and ignore any other specimens on the sheet.

specimen


taxon 

collected by  (illegible)

collection date (to) 

herbarium ex herb ☐ (illegible)

placename county country

site  (illegible) ☐ cultivated

notes

fruits/flowers

collector's number

user comments

☐ request feedback about this sheet

You are not currently logged in. To save your work you will need to login or register.
 If you are a new user then please enter your email address and click 'join' to enable saving. Please leave the password blank as we will email you a password later.
 If you have already registered then please enter your username (or email address) and your password and click 'login/join'.

username or email
 password



[Sheet K001015260 \(Royal Botanic Gardens, Kew\)](#)
[Hide extended help text](#)

Herbarium sheet image © The Board of Trustees of the Royal Botanic Gardens, Kew. Reproduced with the consent of the Royal Botanic Gardens, Kew
 This sheet is one of a trial batch of specimens that Kew are using as a pilot project, testing different documentation systems.

Herbaria@home Links Search

- [News and stats](#)
- [Online databases](#)
- [Herbaria-related](#)
- [Maps & gazetteers](#)
- [Specimens](#)
- [Collectors](#)
- [Taxon names](#)
- [VCCC match](#)

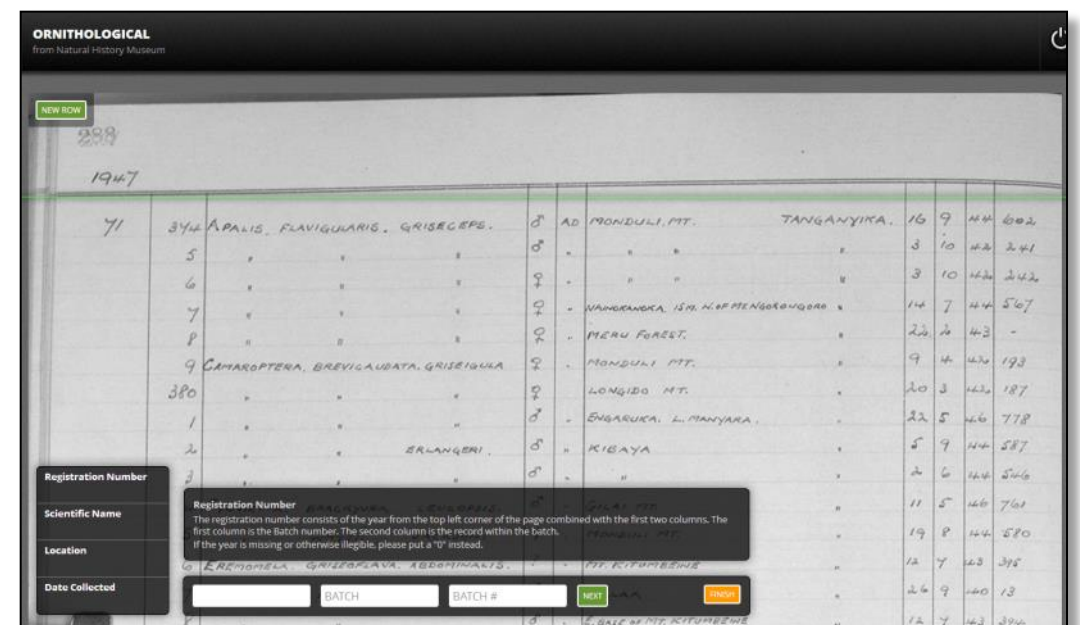
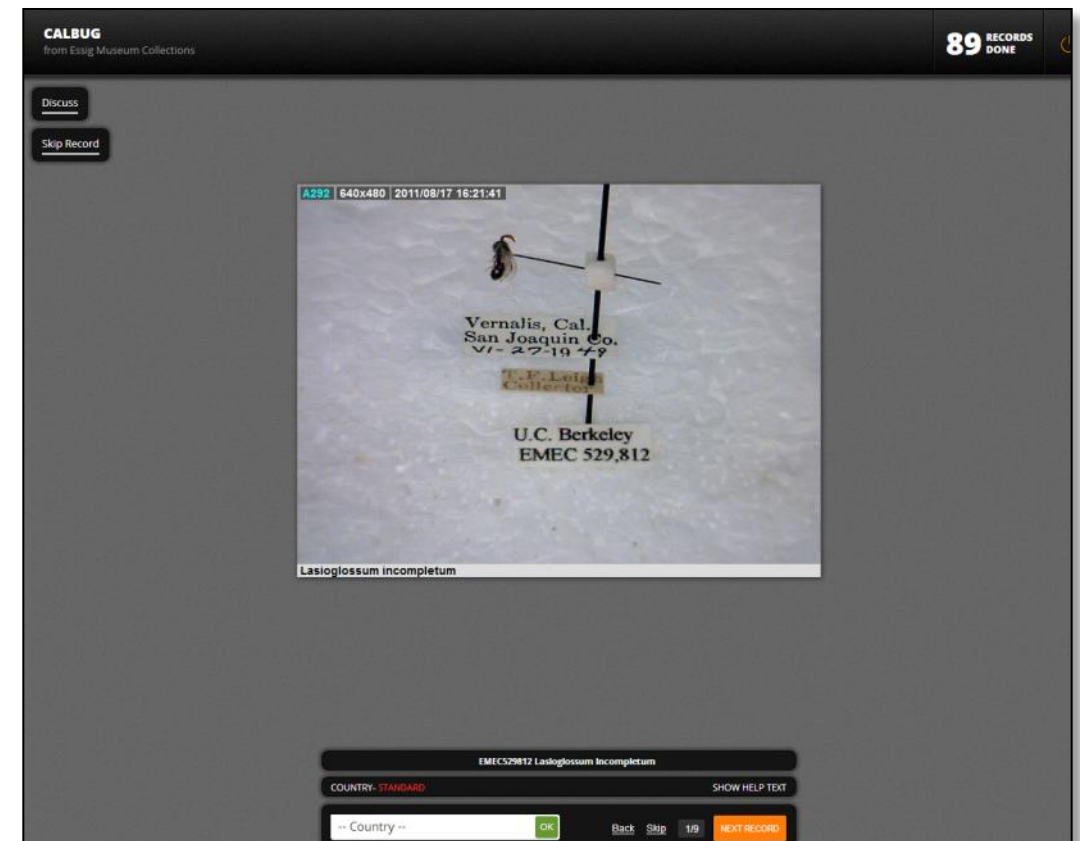
Google

 [OS Placename gazetteer](#)
 [Multi-map](#)
[Vice county \(region\) boundaries](#)
[Biogeomancer workbench](#)

Herbaria@home version 0.006zb
 This web application is still being actively developed. Please report any problems or suggestions for improvements to support@herbariumathome.org

Notes from Nature

- General natural history collections transcription platform
- Launched in 2013
- Part of large, highly active, Zooniverse platform (5.7k/1.1M)
- Four projects (specimen label or register transcription) from large institutes
- Staged transcription interface but no review



<http://www.notesfromnature.org/>

Atlas of Living Australia - Volunteer Portal

- General natural history collections transcription platform
- Launched in 2011
- Not limited to Australian collections or Australian specimens
- Supports alternative transcription approaches
- Admin-based validation/reviewing
- Supports mobile and widescreen resolutions

The screenshot displays the 'Atlas of Living Australia' Volunteer Portal interface. At the top, there's a navigation bar with links like 'Biodiversity Volunteer Portal', 'Expeditions', 'Tutorials', 'Forum', 'Contact Us', 'How can I volunteer?', and 'About the Portal'. A search bar is also present. The main content area shows a specimen image of a hoverfly with a label that reads: 'Goonoo Forest, 5 mi Sth Mendooran NSW 27.3.1971 M. Lamp A and G. Daniels Australian Museum K405407 10 mm'. Below the image, there's a form for transcription with four sections: 1. Transcribe All Text (with a text area and a 'Copy values from a previous task' button), 2. Collection Location (with fields for Locality, State/Territory, Country, Latitude from label, and Longitude from label, plus a 'Find Lat/Long' button), 3. Miscellaneous (with fields for Collection date, Collector, Collection Method, Elevation, Depth, and Field/Other numbers), and 4. Notes (with a text area for 'Your Notes'). At the bottom, there's a footer with links for 'Contact us', 'Partners', 'Communications', 'About the Atlas', 'Citizen Science', and 'Help'. The footer also includes the GBIF logo and a Creative Commons Attribution 3.0 Australia License notice.

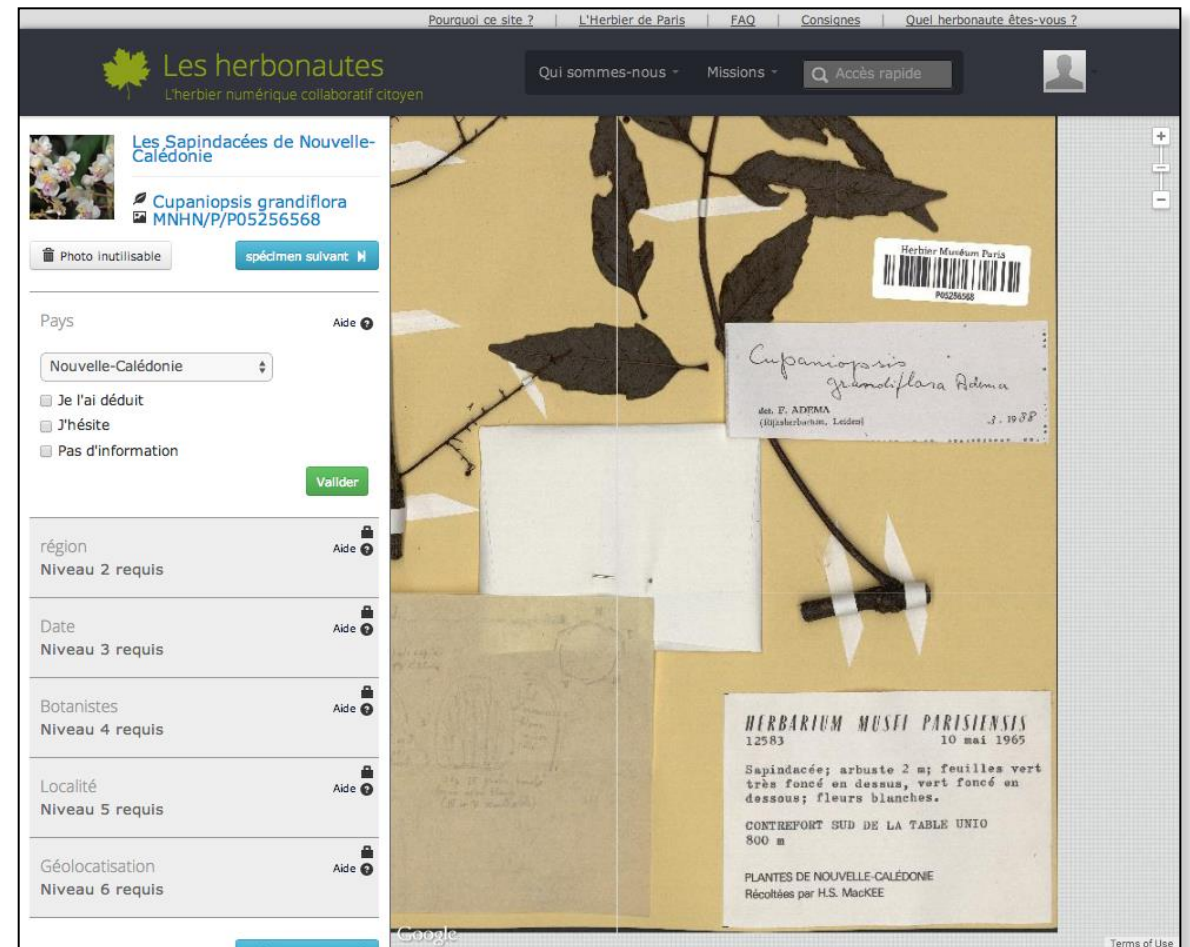
Smithsonian Digital Volunteers: Transcription Center

- General transcription platform – not just for NH but all Smithsonian museums
- Young platform – launched July 2013
- Many small projects (~40) with small communities
- Two-stage transcribing/reviewing

The screenshot displays the Smithsonian Digital Volunteers: Transcription Center interface. At the top, the Smithsonian logo is on the left, and navigation links (HOME, PROJECTS, ABOUT, TIPS) are on the right. Below this, a dark banner reads "SMITHSONIAN DIGITAL VOLUNTEERS: TRANSCRIPTION CENTER" with "SIGNUP" and "LOGIN" links. The main content area shows a breadcrumb trail: "HOME > PROJECTS > NMNH - DEPARTMENT OF BOTANY > THE BOMBACACEAE FAMILY". Below this, it says "TRANSCRIBE PAGE 1 OF 458". A color calibration chart and a photograph of a botanical specimen (a branch with leaves and flowers) are on the left. To the right of the specimen is a "TRANSCRIPTION FORM" with a "HELP" link. The form has three tabs: "1) General", "2) Collector Details", and "3) Location". The "General" tab is active, showing fields for "Country" (Mexico), "Territory" (Veracruz-Llave), "District / County", "Precise Locality", "Latitude", "Longitude", "Elevation (from)" (40), "Elevation (to)", and "Elevation Unit" (meters). A "Notes on Transcribing this page (optional)" text area is at the bottom of the form. A "feedback" button is on the right side of the form. At the bottom of the page, there is a disclaimer: "Please note that some language in this collection may be culturally insensitive or offensive to some viewers. It is presented as it exists in the original document for the benefit of research. The material reflects the culture and context in which it was created and not the views of the Smithsonian Institution." and a footer with "© 2013 SMITHSONIAN", "COLLECTIONS", "TERMS OF USE", and "PRIVACY".

Les Herbonautes

- Herbarium sheet transcription platform
- Launched in 2012
- 21 projects and over 1,400 participants
- Unique amongst platforms with user levels based on experience/skill
- Redundant entries



<http://lesherbonautes.mnhn.fr/>

PyBossa / Crowdcrafting.org

- PyBossa - free, open-source crowdsourcing architecture/platform
- Crowdcrafting.org – free project hosting site
- Micropasts project allows sign-in using social network accounts



British Museum Bronze Age Index Drawer B8: Contribute

Please transcribe this index card

You can use your mouse wheel to zoom in/out the picture, as well as drag & drop to navigate in the document.

Object Type

Top of card

Date of Discovery

Year Month Day

Right Corner

Right Corner

Site

Left of the card

Georeference the site

Geolocate the site in a map

Search

Map

I believe the location I have given on the map is accurate to within:

☒ 100 metres

☐ 1 kilometre

☐ 10 kilometres

Grid Reference

Top left of card

Collection

Right of card

Publications

Centre of card

Length

In millimetres

Width

In millimetres

Edge

In millimetres

Thickness

In millimetres

Weight

In grammes

Composition

Below measurements

Description

Below measurements

Surface

Left of card

Patina

Towards bottom left of card

Associations

Bottom centre of card


Remarks

Bottom centre of card

Transcriber's Comments

Add here your comments, if any

Submit the transcription

the British Museum UCL  PYBOSSA

© Email Twitter Facebook Flickr API

What about the crowd?

Participant motivation - why does it matter?

CS isn't free and participation isn't a given

1. Projects need to appeal

Understanding interests, motivations & expectations of contributors is central to effective project design, uptake & enjoyment

2. And yield useful science data

Link between level of engagement and motivation of participants and resultant quality and volume of data & longevity of participation



Why do people participate?



***Social
community***

Science

Volunteering

Photography

Gamer

***Biological
recording***

Teacher

**Museum
volunteer**

Student

**Professional
scientist**

**Community
leader**

Journalist

What motivates the crowd?

Amateur naturalists	Local communities	Educators
Interest in project and its scientific goals	Interest in project and its scientific goals	Interest in project and its scientific goals
Enjoyment of finding, identifying, recording wildlife	Deep affinity with location (historical, social, cultural value)	Participation in 'real' science
Aesthetic appreciation of nature	Projects that support shared, local goals	Fit with curriculum
Personal learning and reputation	Opportunity to share knowledge	Rapidity of feedback
To support conservation activity	New activities in a familiar space	Ease - ready-made, accessible materials
Social aspects	Social aspects	Location and cost

[OPAL, 2013; Roy et al., 2012; Tweddle et al., 2012; Robinson et al., 2013; Ellis & Waterton, 2005; Grove-White et al., 2007 ; Hobbs & White, 2012]

On-going support & reward – what works?

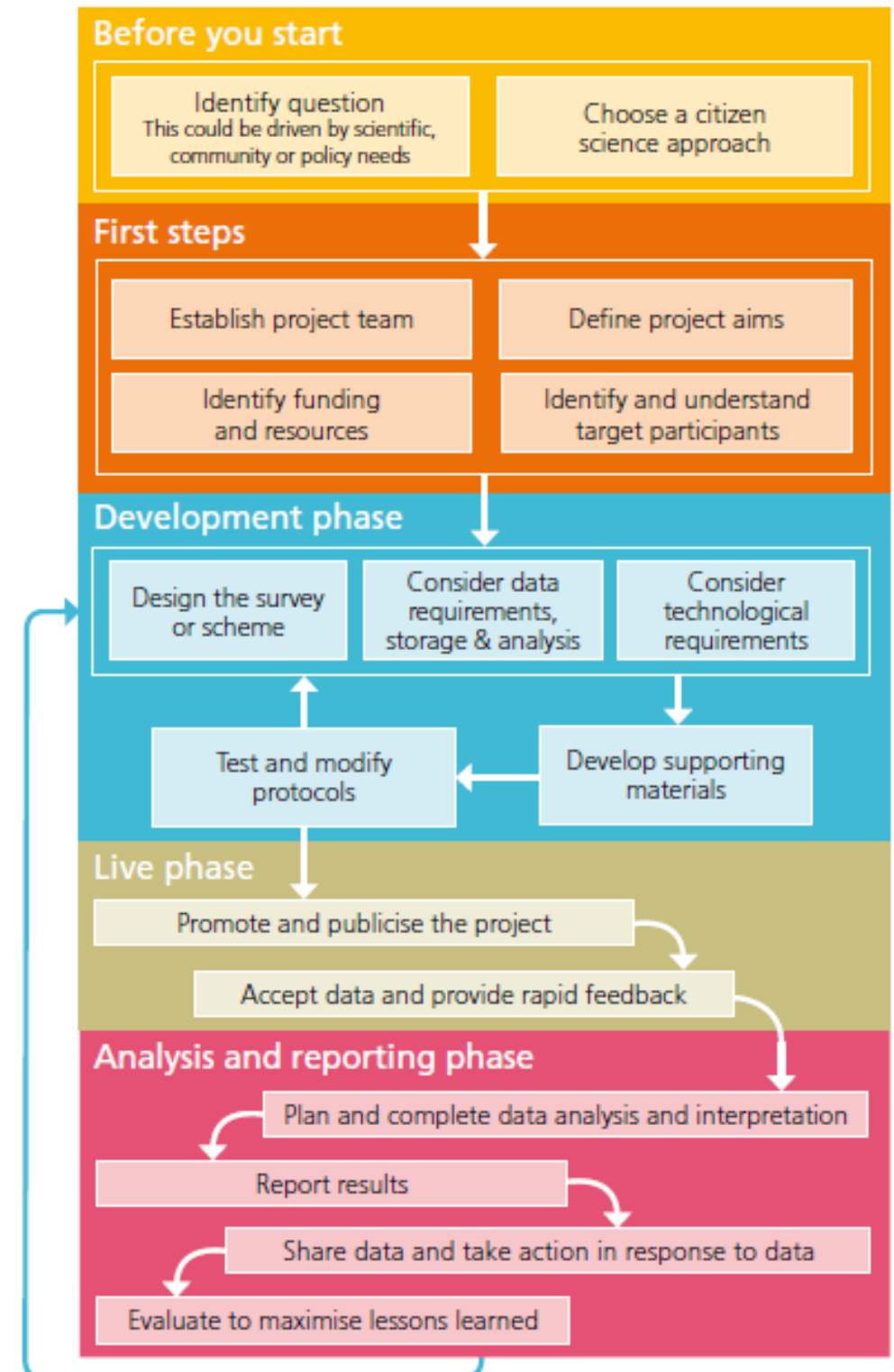
- Ongoing, rapid feedback and thanks
- Evidence that the data are being used
- **Social interaction and community**
- **Personal learning and progression**
- Recognition and reputational gain (incl. super-contributors)
- Awards, games, badges, leaderboard (work for some people, not others)

Top 20 users

	user	specimens
1	oldnick	24047
2	hallucigenia	10550
3	qgroom	9743
4	nacs12	9444
5	johned	7609
6	keith barnett	6108
7	xmhcman	5540
8	shaw30	4657
9	daniel king	4020
10	chrisu	3414
11	siandef	3340
12	tom humphrey	3140
13	kenbalkow	2971
14	CLL	2686
15	lowfatspread	2672
16	mossysal	2563
17	portwild	2485
18	mikedaps	2460
19	johnhawksford	2445
20	rmwalls	2151

So what does this mean as a practitioner?

- Projects need to be personally and socially relevant to succeed
- Motivations of participants often quite different to those of project designer
- One size rarely fits all - danger of making assumptions
- Key to success is working with and understanding target participants – and adapting



End goal: SYNTHEsys Report – JRA: Physical to Digital

- Quantitative comparisons between project methods
- Demographics and information on the crowd
- Call for data – if you have any relevant information please share!

I.livermore@nhm.ac.uk

SYNTHEsys 
Synthesis of systematic resources

Acknowledgements

SYNTHESYS Report Collaborators:

Pieter van den Besselaar¹, Henk Koerten¹, Sarah Phillips², Lucy Robinson³, Vincent S. Smith³ and John Tweddle³

1. VU University Amsterdam, Netherlands; 2. RBG Kew, Surrey, UK; 3. NHM, London, UK

Crowdsourcing Data providers: Tim Conyers, Tom Humphrey, Ruth Benny

Additional slides

De-motivation factors

Amateur naturalists	Local communities	Educators
Data being used for purposes other than intended ('misuse')	Poor communication (top-down)	Confidence (e.g. in identification ability)
Projects that detract from interests	Transient support	Projects that are hard to link to curriculum
Lack of recognition	Lack of recognition	Location and cost

How do we ensure the data are fit for use?

- Two main methods to capture data:
 - Multiple transcription passes (NfN)
 - Single pass crowd reviewing/correction (Smithsonian)
- Autocomplete fields (ALA)
- Flag problems or questions (for community or expert response)
- Quality indicators for data

Generally interesting facts and observations

- Small subset of dedicated users contribute the most e.g. (H@H <1% users did >32%)
- H@H - user behaviour changes over time