

Deciphering natural history and mapping the world through collection labels

SPNHC 2022: Back to Basics: Museum Techniques Skillshare
Wed, 8th June 2022

Krisztina Lohonya

Natural History Museum, London

Presentation DOI: [10.6084/m9.figshare.19762876](https://doi.org/10.6084/m9.figshare.19762876)





Creating discoverable digital collections

- Digitisation efforts
- NHM holds an estimated 80 million objects
- All objects need to be digitised and many transcribed & georeferenced
- What is transcription, georeferencing and why are they important?

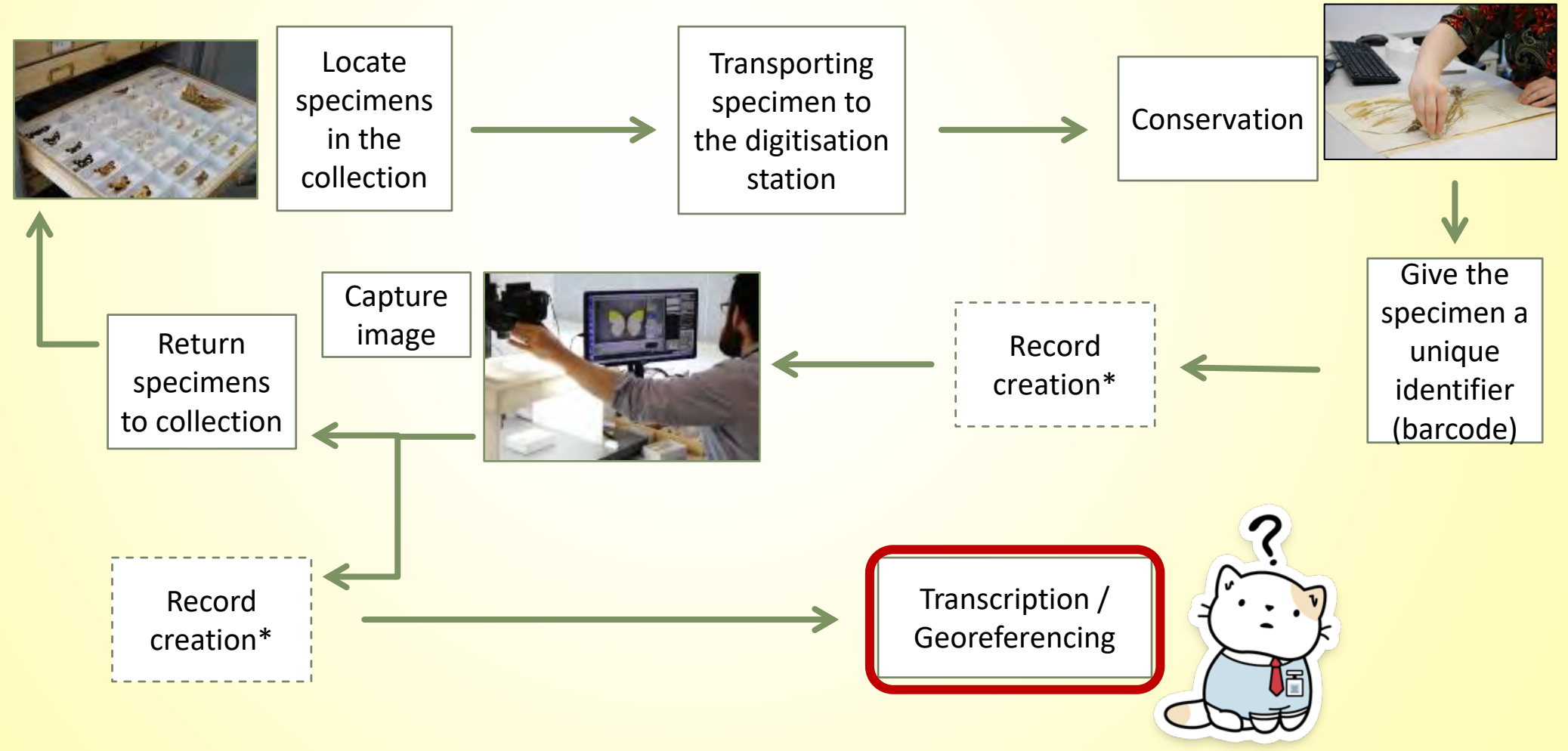
- Challenges along the way
- Solution(s)?

Digitisation of an object from a natural history collection (simplified and idealistic)

1. Collection

2. Imaging Specimens

3. Data Processing



Transcription – My “love-hate” relationship with handwritten labels

Plants of Qinghai, China Convolvulaceae

Cuscuta europaea L.

Chindu Xian, Xiwu Xiang: just S of Chumda, along E side of the Tongtian He (upper Chang Jiang).
Elev. 3440 m 33° 0' N 97° 15' E

Steep sided valley, bottom filled by alluvial fan, mostly colonised by *Elscholtzia frutescens*, slopes of metamorphic rock with *Caragana*, *Cotoneaster*, *Caryopteris* etc., *Artemisia* species common throughout.

Flowers white flushed pink.
Growing as a parasite on *Elscholtzia frutescens*

T. N. Ho (HNWP), B. Bartholomew (CAS), M. Watson (E) and M. Gilbert (MO at BM) 1747, 14 August 1996

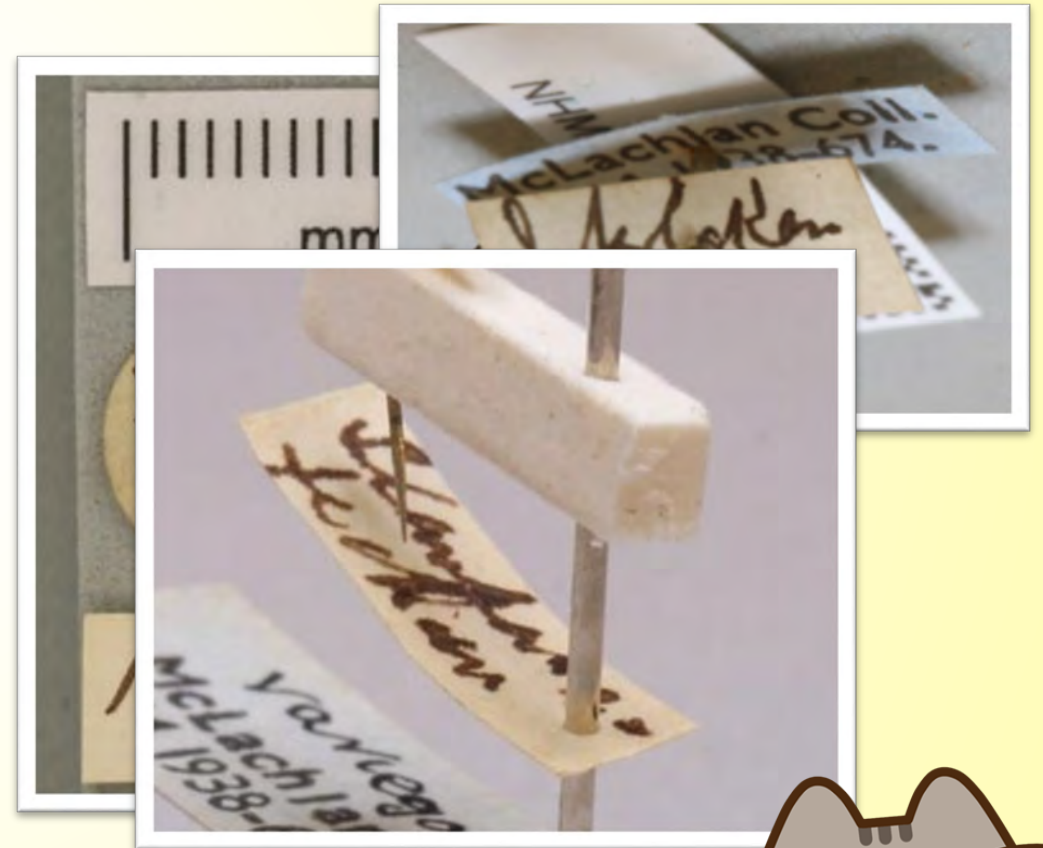
Rhynchosia, Lour.
volubilis, Lour.
O. N. Lathyraceae.
Whampoa, Dana's Island.
Legi m. Octob: 1860.

HERB. H. F. HANCE. Recd. 1887.

MANICA E SOFALA: *Vila Machado, Lamego, margens do rio Munda*

Legi, 29 / II / 1948

Transcription – “The good, the bad and the ugly”



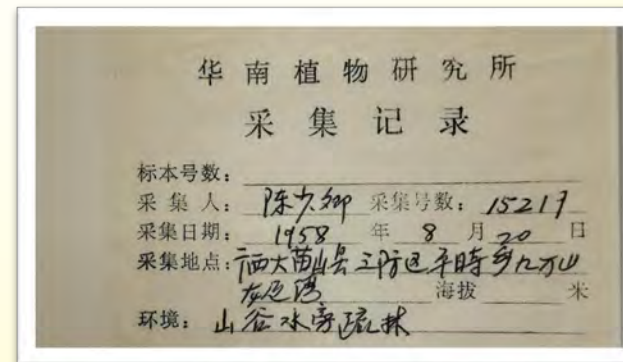
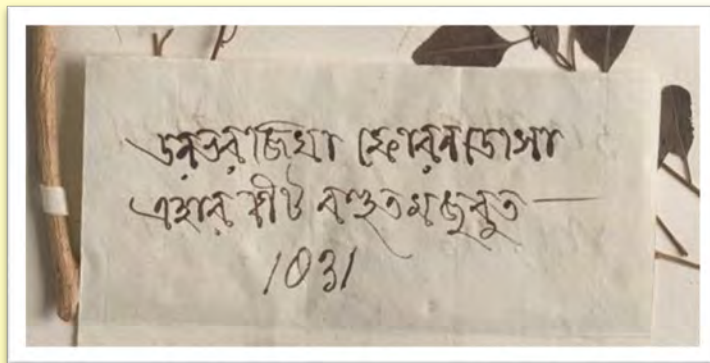


“Just add another language” – When transcription becomes transliteration

- Transcription vs. transliteration vs. romanization

Original word	Transliteration	Transcription
Ελευθερία	<i>Eleutheria</i>	<i>Eleftheria</i>

- ‘Foreign’ languages on labels – foreign for the host country
- Easier with common languages/most widely spoken – e.g. English, Spanish, French
- Harder with languages that are less commonly spoken (e.g. dead languages or endangered languages) and/or languages with non-Latin writing systems (e.g. Chinese, Arabic, Russian)
 - Orthographic transcription – different for each language / could change

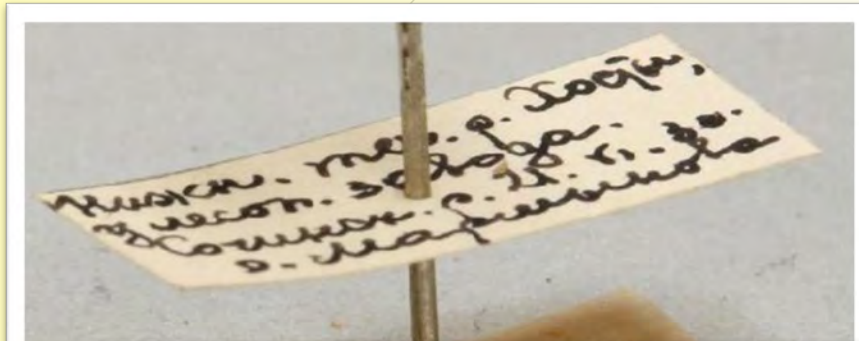


What is the solution?

- Methods developed for certain type of labels and languages
- Some languages that are using a different writing systems have an internationally accepted transliteration system
 - But some do not, those are the harder ones to deal with
- Examples:
 - Chinese – currently pinyin is the internationally accepted system
 - Russian – number of systems exist, for Anglophones BGN/PCGN romanization of Russian is most used
 - Japanese – several systems exist, mostly Hepburn romanization is used
 - Arabic – technically it is not even transliteration, but transcription, due to the unique nature of the language

Romanization system	Romanised name
Modern (pinyin)	Qinghai
Wade-Giles	Ch'ing-hai
Portuguese	Chinghai
German	Tschinghai
French	Ts'ing-hai

Russian example

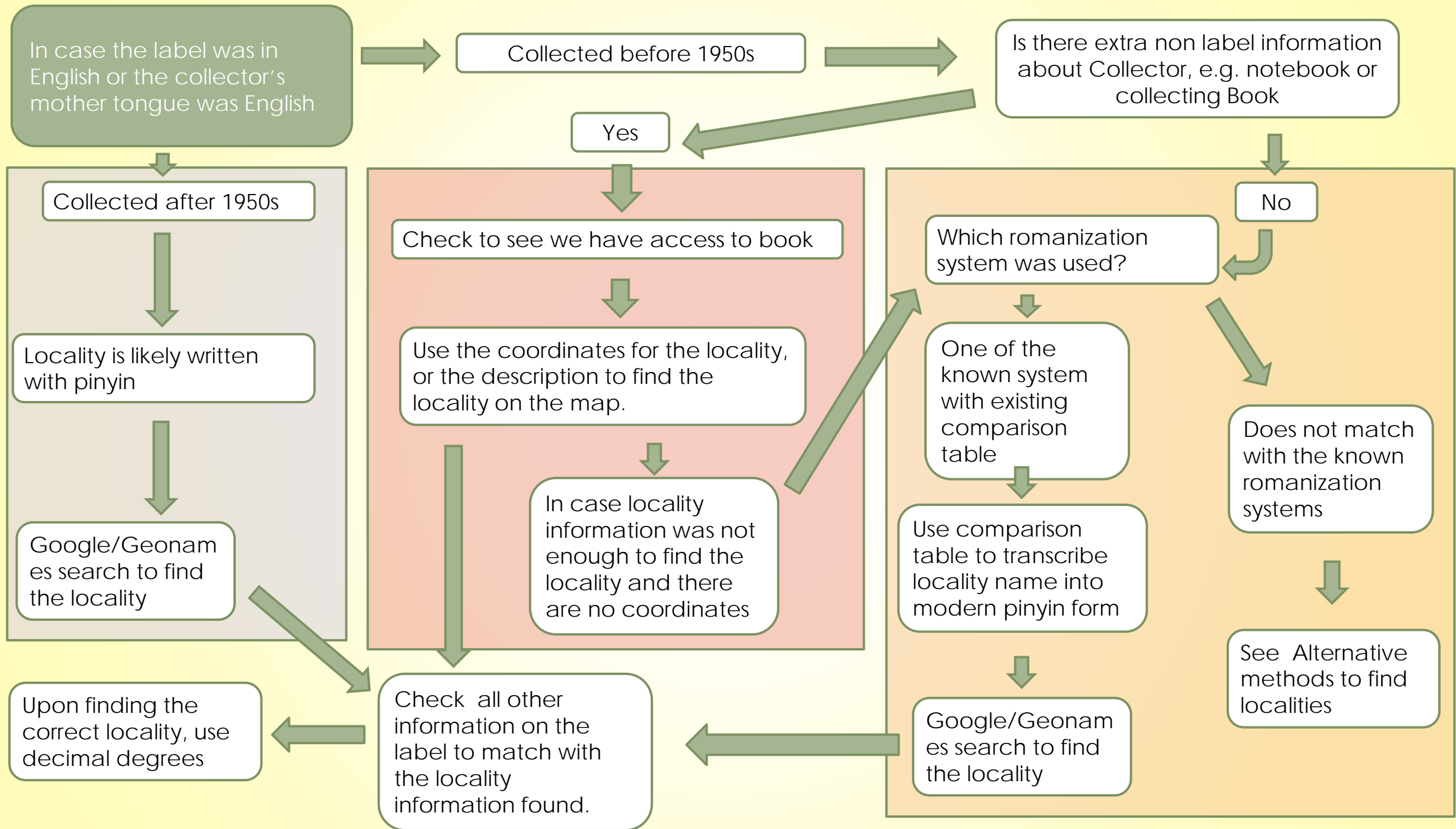


Нижн. Теч. р. Хосты у лесопильного.
завода Сочинск р. 21. VI. 30. О. Мартинова

А	а	А	а
Б	б	Б	б
В	в	В	в
Г	г	Г	г
Д	д	Д	д
Е	е	Е	е
Ё	ё	Ё	ё
Ж	ж	Ж	ж
З	з	З	з
И	и	И	и
Й	й	Й	й
К	к	К	к
Л	л	Л	л
М	м	М	м
Н	н	Н	н
О	о	О	о
Р	р	Р	р
С	с	С	с
Т	т	Т	т
У	у	У	у
Ф	ф	Ф	ф
Х	х	Х	х
Ц	ц	Ц	ц
Ч	ч	Ч	ч
Ш	ш	Ш	ш
Щ	щ	Щ	щ
Ъ	ъ	Ъ	ъ
Ы	ы	Ы	ы
Ь	ь	Ь	ь
Э	э	Э	э
Ю	ю	Ю	ю
Я	я	Я	я

Lower. Tech. R. Khosta at the
sawmill plant Sochi r. 21. VI.
30. O. Martynova





Chinese example



HERB: H. F. HANCE.
Ad coenobium buddhisticum Filoitsz secus fl. North River, prov. Cantoniensis, in murorum fissuris. Ipse legi



FROM THE HERBARIUM OF H. F. HANCE.
At the Buddhist monastery Filoitsz on the North River, prov. of Canton, growing in the clefts of the walls. Collected by self



I was unable to believe that the two could be conspecific. In September last, however, in an excursion I made with Mr. Sampson up the North River to the Tsing-yune Pass, about 120 miles above Canton, in the magnificent dense woods encompassing the renowned Buddhist monastery of Filoi-tsz,—which, let me note, *en passant*, is



Solutions

- More available methodologies
- Visual aids for transcription
- Documentation
- Having a diverse team working on transcription and georeferencing
- Make effective use of polyglots in the institution
- Technological approaches
 - Annotation platforms

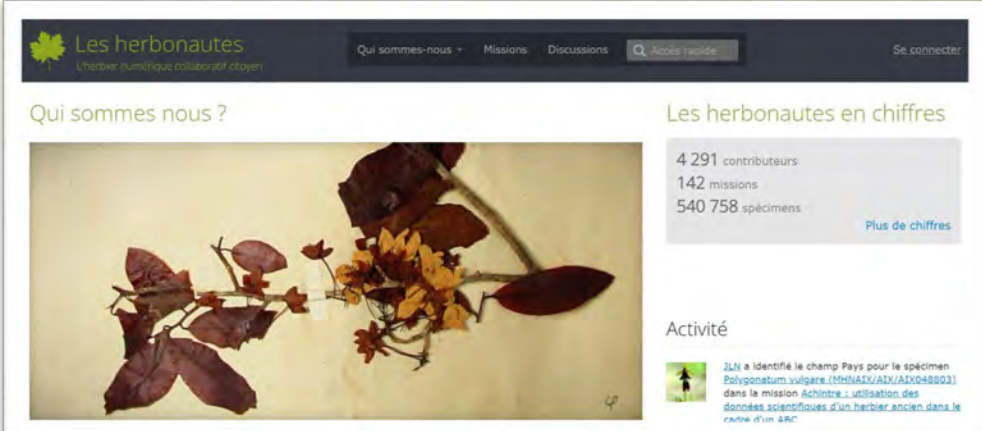
An alphabet of around 1800.

А а А А А А А а а а а а
В в В в В в В в В в В в

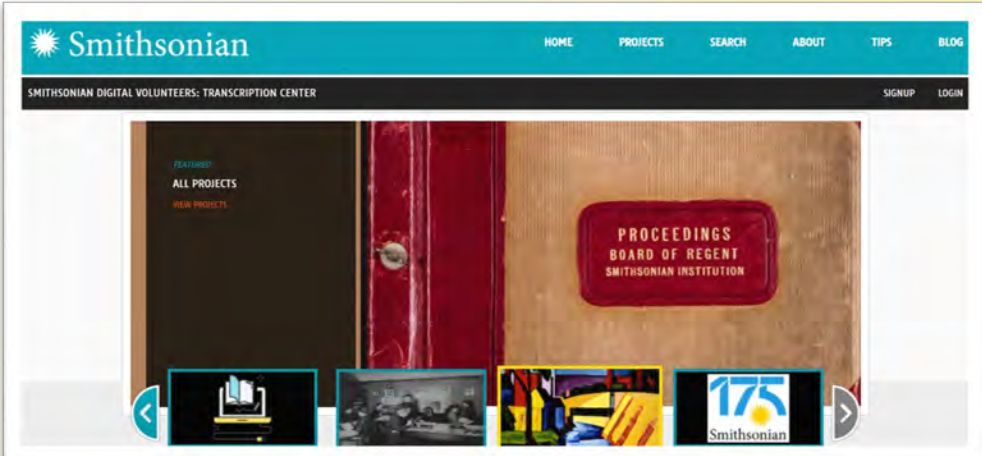
А	а	А	а	Р	р	Р	р
Б	б	Б	б	С	с	С	с
В	в	В	в	Т	т	Т	т
Г	г	Г	г	У	у	У	у
Д	д	Д	д	Ф	ф	Ф	ф
Е	е	Е	е	Х	х	Х	х
Ё	ё	Ё	ё	Ц	ц	Ц	ц
Ж	ж	Ж	ж	Ч	ч	Ч	ч
З	з	З	з	Ш	ш	Ш	ш
И	и	И	и	Щ	щ	Щ	щ
Й	й	Й	й	Ъ	ъ	Ъ	ъ
К	к	К	к	Ы	ы	Ы	ы
Л	л	Л	л	Ь	ь	Ь	ь
М	м	М	м	Э	э	Э	э
Н	н	Н	н	Ю	ю	Ю	ю
О	о	О	о	Я	я	Я	я

Solutions

- Developing digital volunteering programs/platforms
- E.g.:
 - Especially viable for local collections; e.g. British & Irish collections' transcription and georeferencing outsourced to British and Irish audience
 - Herbonauts - <http://lesherbonautes.mnhn.fr/>
 - Smithsonian - <https://transcription.si.edu/>
 - Zooniverse
 - <https://www.zooniverse.org/organizations/md68135/notes-from-nature>



The screenshot shows the homepage of 'Les herbonautes', a digital volunteering platform. The header includes the logo and navigation links: 'Qui sommes-nous', 'Missions', 'Discussions', 'Accès rapide', and 'Se connecter'. The main content area features a large image of a pressed plant specimen. To the right, a statistics box titled 'Les herbonautes en chiffres' displays: 4 291 contributeurs, 142 missions, and 540 758 spécimens, with a 'Plus de chiffres' link. Below this, an 'Activité' section shows a user profile and a recent activity: 'JLN a identifié le champ Pays pour le spécimen Polygonatum vulgare (MHN:AIX/AIX048803) dans la mission Achivre : utilisation des données scientifiques d'un herbier ancien dans le cadre d'un ZNC'.



The screenshot shows the 'Smithsonian Digital Volunteers: Transcription Center' website. The header features the Smithsonian logo and navigation links: 'HOME', 'PROJECTS', 'SEARCH', 'ABOUT', 'TIPS', and 'BLOG'. Below the header, the text 'SMITHSONIAN DIGITAL VOLUNTEERS: TRANSCRIPTION CENTER' is visible, along with 'SIGNUP' and 'LOGIN' links. The main content area displays a large image of a book cover titled 'PROCEEDINGS BOARD OF REGENT SMITHSONIAN INSTITUTION'. At the bottom, there are several small thumbnail images, including one with the text 'ALL PROJECTS' and another with the '175 Smithsonian' logo.

Tips & Tricks

- ▶ Always have the original label on hand/photograph of it
- ▶ Clustering records by data
 - ▶ First transcribing collector and date, then cluster them together to ease transcription and geo referencing.
- ▶ Relax and enjoy the detective experience of natural history through the collections!





Thank you for your attention!