

2023年2月18日

第40回自然史標本の発信に関する研究会

Zooniverseを用いた クラウドソーシングによる 標本データ入力

海老原 淳

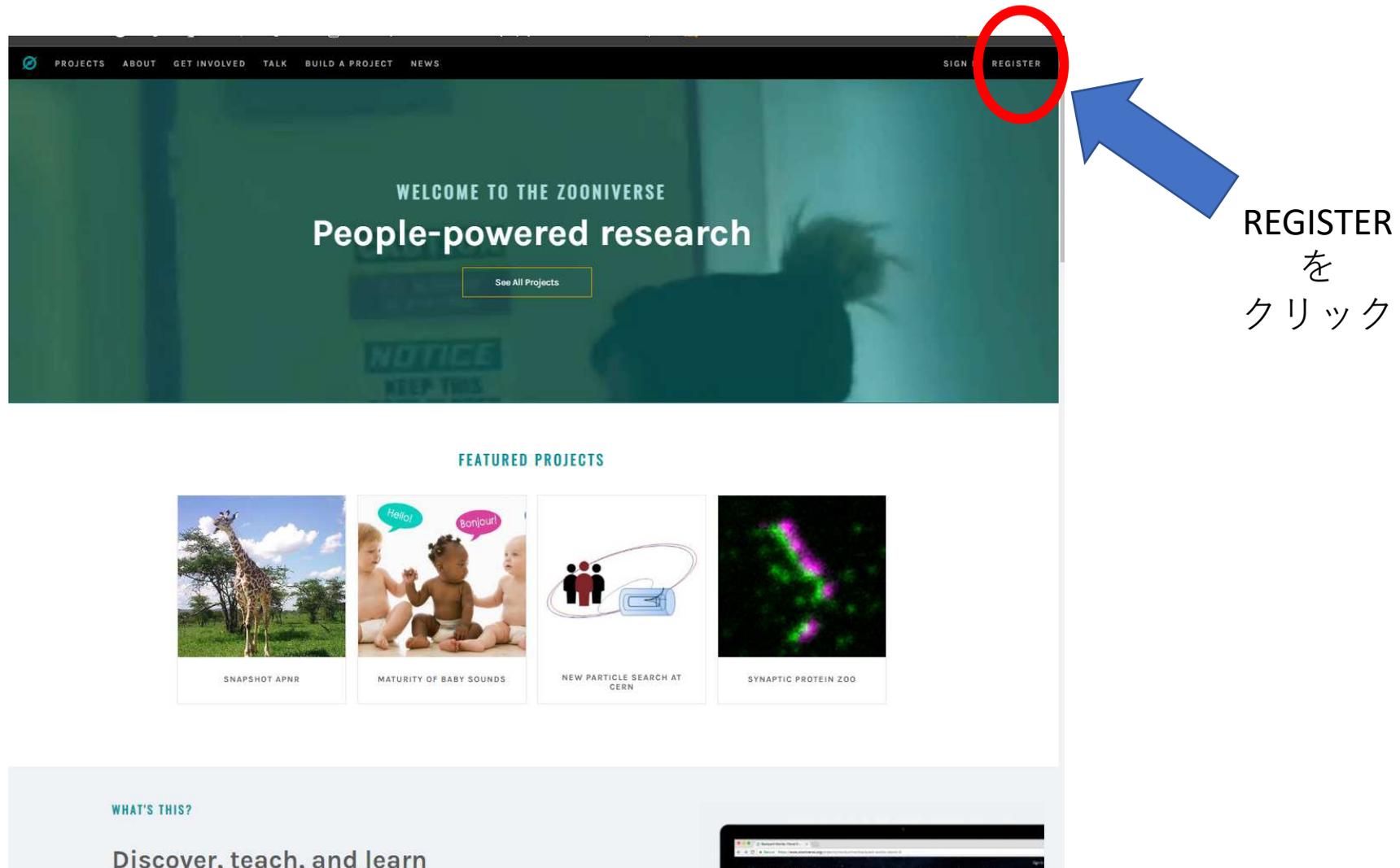
(国立科学博物館植物研究部)

はじめに

- 今回は米国などの市民科学プロジェクトで幅広く使用されている「Zooniverse（ズーニバーズ）」を用いて、クラウドソーシングによる標本データ入力を体験していただきます。
- まずは、アカウント作成をしていただきます。

Zooniverseのアカウント作成方法

1. Zooniverse(<https://www.zooniverse.org/>)にアクセスし、右上の黒い部分にある「REGISTER」をクリックする。



The image shows a screenshot of the Zooniverse website homepage. The top navigation bar is dark with white text for 'PROJECTS', 'ABOUT', 'GET INVOLVED', 'TALK', 'BUILD A PROJECT', 'NEWS', 'SIGN', and 'REGISTER'. The 'REGISTER' button is circled in red, and a blue arrow points to it from the right. Below the navigation bar, the main content area has a dark green background with the text 'WELCOME TO THE ZOOVERSE' and 'People-powered research'. A 'See All Projects' button is visible. Below this, there is a 'FEATURED PROJECTS' section with four project cards: 'SNAPSHOT APNR' (giraffe), 'MATURITY OF BABY SOUNDS' (babies), 'NEW PARTICLE SEARCH AT CERN' (particle detector), and 'SYNAPTIC PROTEIN ZOO' (microscopy image). At the bottom, there is a light blue footer with the text 'WHAT'S THIS?' and 'Discover, teach, and learn'.

REGISTER
を
クリック

2. アカウント作成に必要な情報を入力する。

4か所入力が必要です。

①User name ユーザー名

半角英数字で任意のユーザー名を考えてください。（例：taro-kahaku1）

Looks good と出れば、そのユーザー名はまだ使われていないので使えます。

プロジェクト参加者同士のやりとりは、このユーザー名が使われます。

②Password パスワード

③Confirm password パスワード確認用
任意のパスワードを設定し、2箇所と同じものを入力。忘れないようにしてください。

④Email address メールアドレス

パスワードを忘れた時のために、実際にメールが届くアドレスを登録しておくといいです。

WELCOME TO THE ZOOVERSE

SIGN IN REGISTER

If you are under 16 years old, tick this box and complete the form with your parent/guardian.

User name *Required*

①

You'll use this name to log in. It will be shown publicly.

Password *Required*

②

Confirm password *Required*

③

Email address *Required*

④

Real name *Optional*

We'll use this to give you credit in scientific papers, posters, etc.

You agree to our [privacy policy](#) (required)

It's okay to send me email every once in a while. (optional)

I'd like to help test new projects, and be emailed when they're available. (optional)

2. アカウント作成に必要な情報を入力する。

18歳未満でなければ**チェックは不要**

The image shows a registration form with the following fields and annotations:

- Age Confirmation:** If you are under 16 years old, tick this box and complete the form with your parent/guardian. (An arrow points from the 18-year-old annotation to this checkbox.)
- User name:** *Looks good* *Required*. Below the field: *You'll use this name to log in. It will be shown publicly.*
- Password:** *Required*
- Confirm password:** *Looks good* *Required*
- Email address:** *Looks good*
- Real name:** *Optional*. Below the field: *We'll use this to give you credit in scientific papers, posters, etc*. (An arrow points from the 'Real name' annotation to this field.)
- Agreements:**
 - You agree to our [privacy policy](#) (required) (An arrow points from the 'Privacy policy' annotation to this checkbox.)
 - It's okay to send me email every once in a while. (optional)
 - I'd like to help test new projects, and be emailed when they're available. (optional) (An arrow points from the 'Email notifications' annotation to this checkbox.)
- Register Button:** A blue button labeled "Register".

Real name (本名) 欄は、空欄のままで差し支えなし。

「You agree to our privacy policy」に**チェックを入れる**。(必須)

下2つのチェックボックスは、**チェックを外す**。
(メール通知は不要なため)

3. 記入できたら、水色のRegisterボタンを押して、情報を送信する。

SIGN IN REGISTER

If you are under 16 years old, tick this box and complete the form with your parent/guardian.

User name Looks good Required

You'll use this name to log in. It will be shown publicly.

Password Required

Confirm password Looks good Required

Email address Looks good

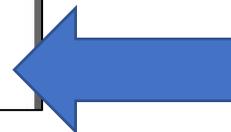
Real name Optional

We'll use this to give you credit in scientific papers, posters, etc

You agree to our [privacy policy](#) (required)

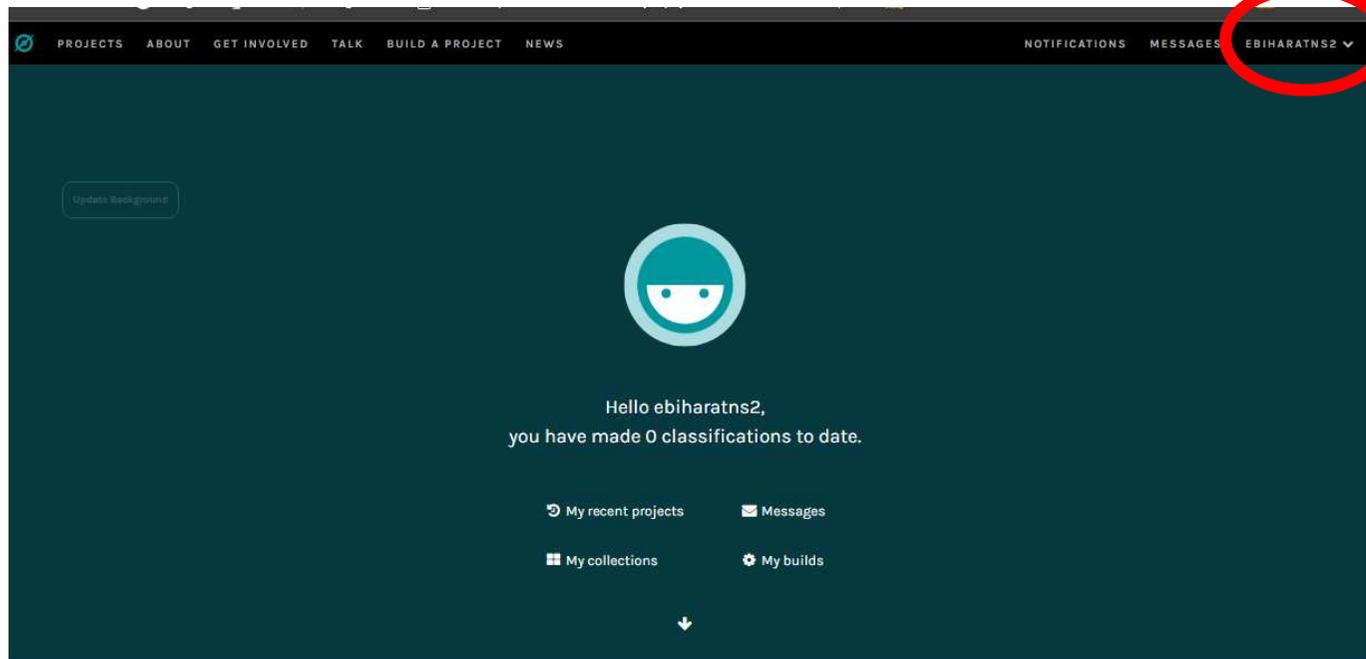
It's okay to send me email every once in a while. (optional)

I'd like to help test new projects, and be emailed when they're available. (optional)



REGISTER
を
クリック

4. この画面に切り替わったら、アカウント作成完了です。

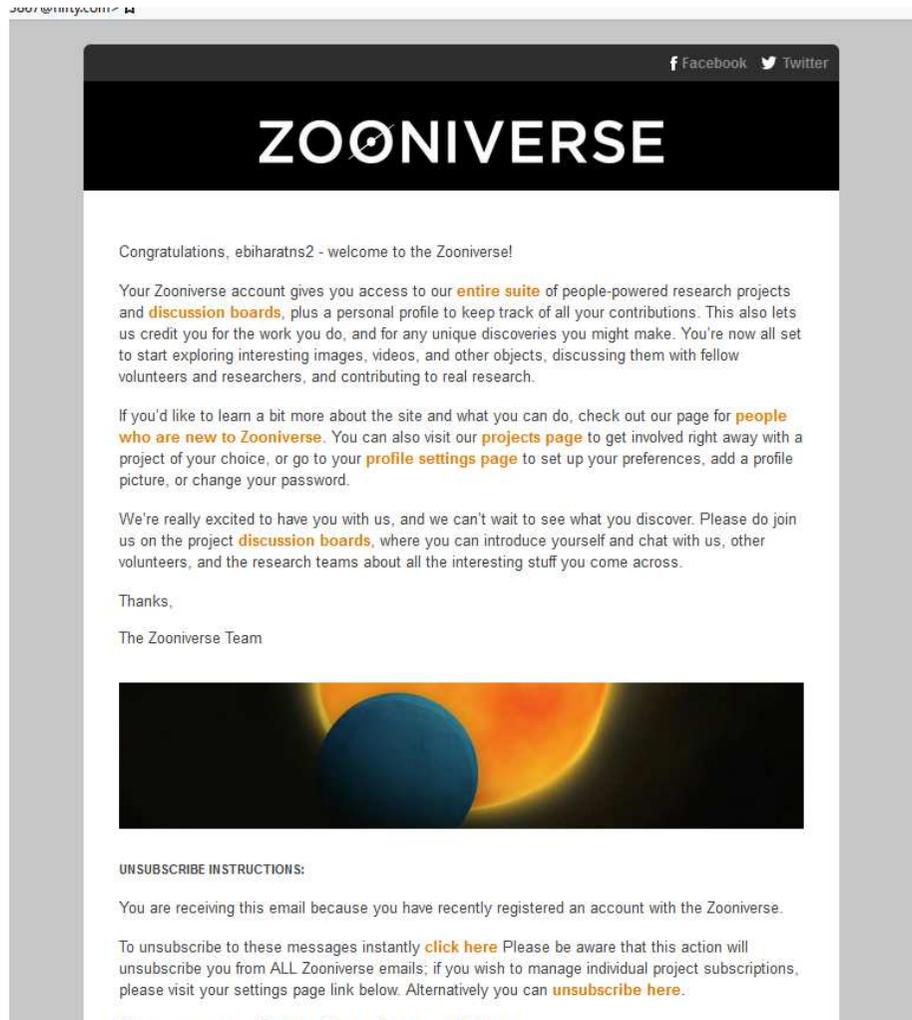


ここに、自分の登録したユーザー名が表示されていれば、成功です。

Zooniverseの準備は、一旦完了です。



5. 登録完了メールがZooniverseから届く。



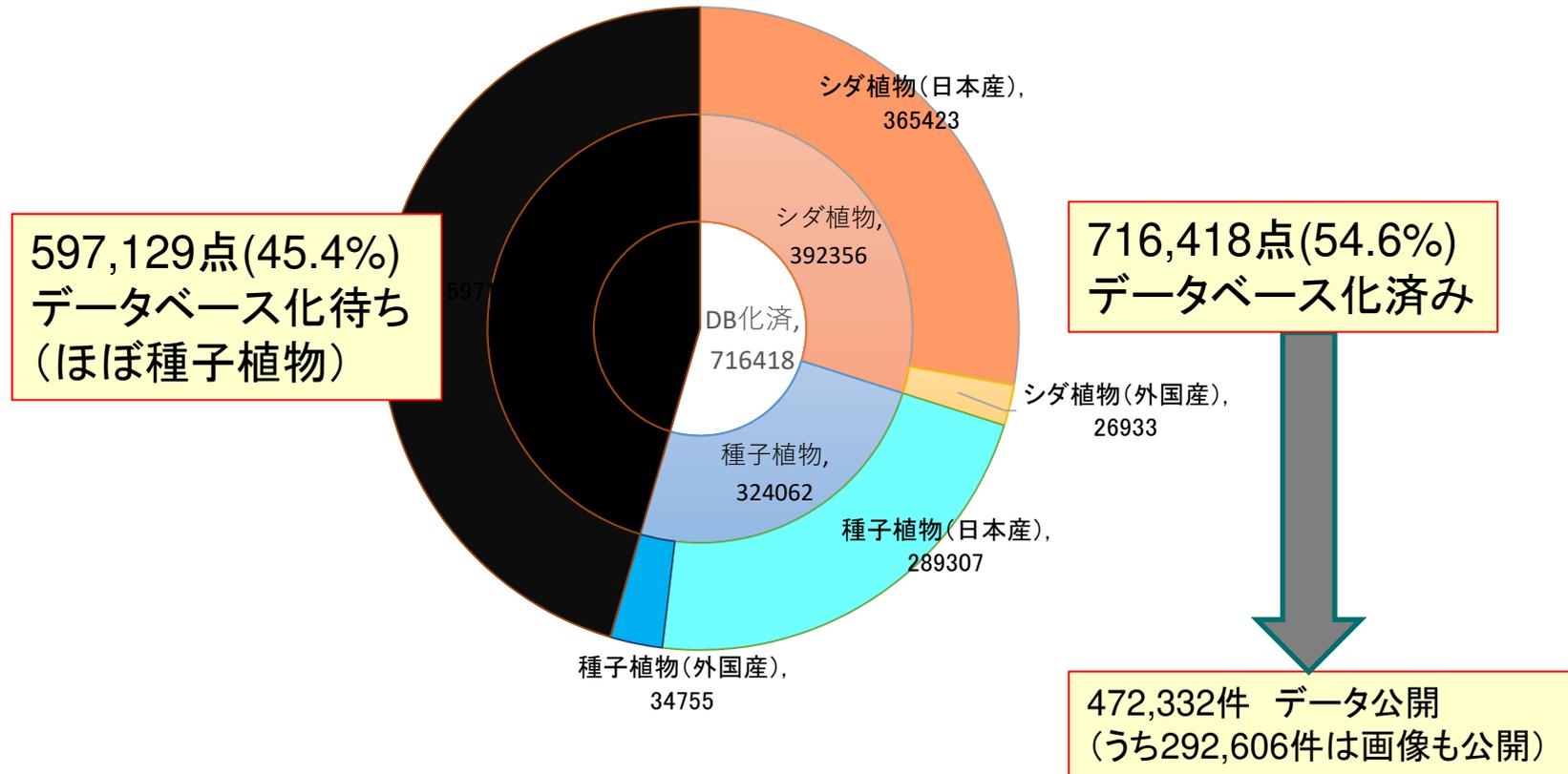
確認メールなので、無視して差し支えありません。

※メールが届かない方は、正しいメールアドレスが入力されていないか、迷惑メールに分類されている可能性があります。

イントロダクション：
自然史標本とクラウドソーシング

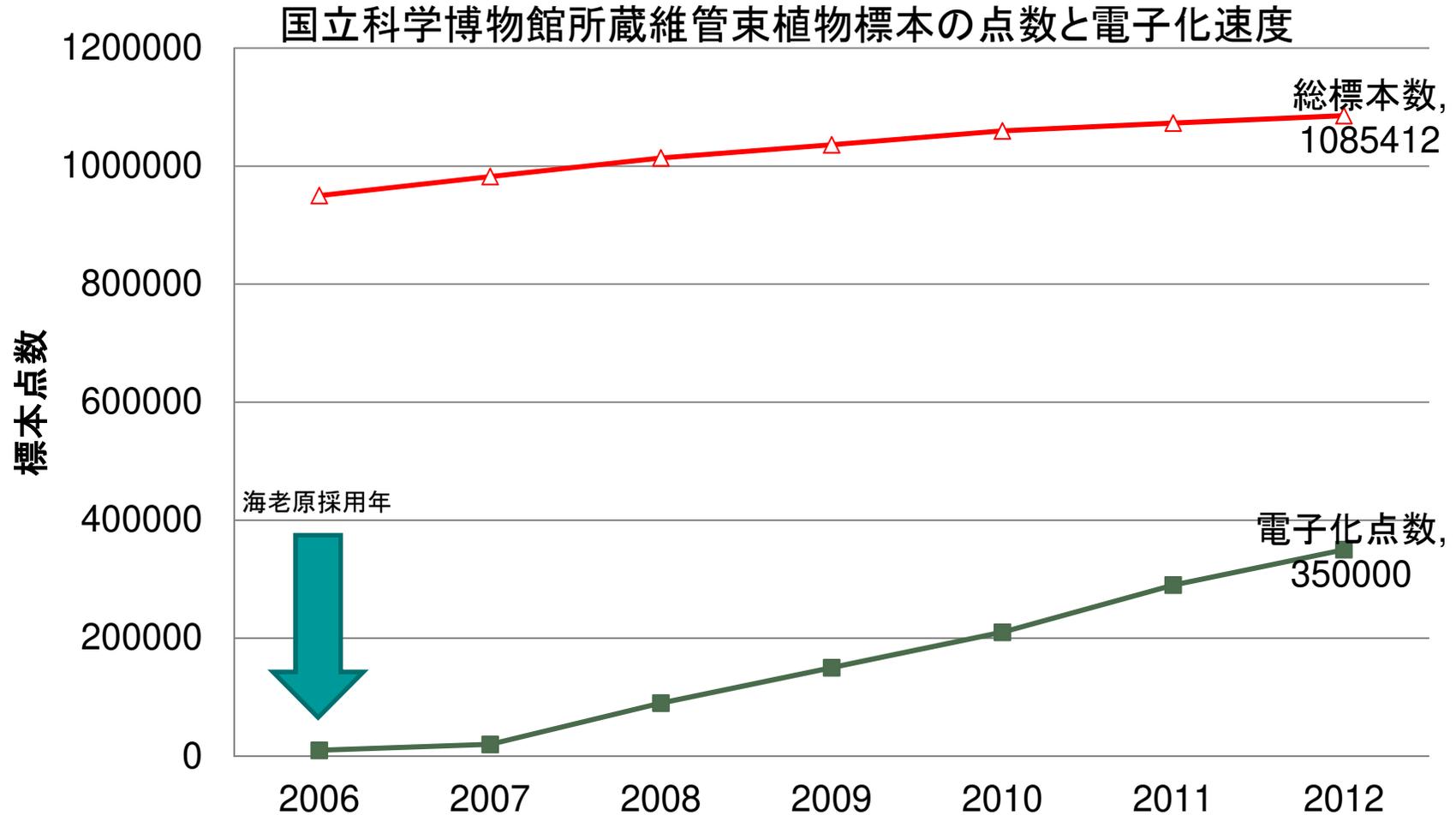
国立科学博物館の維管束植物標本の現状

全1,313,547点(2020年度末時点)



約150年分を遡及してデータベース化する困難

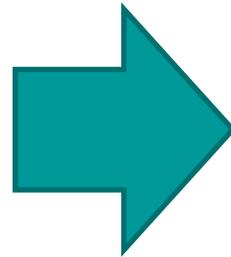
国立科学博物館の維管束植物標本の現状



手法的革新がなければ、半永久に終わらない

発想の転換

- 標本収蔵場所で、
標本を見ながら
データ入力。

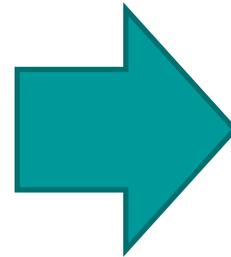


- 標本収蔵場所で
ラベルを写真撮影。
↓
画像をクラウドに置く。
↓
遠隔地・好きな時間に
作業者が画像を開い
て、データ入力。

植物のさく葉標本に適している方法。
(ラベルの撮影自体に手間がかかる標本には不向き)

発想の転換

- 標本収蔵場所で、
標本を見ながら
データ入力。



問題点
・スペース不足
・人材不足

- 標本収蔵場所で
ラベルを写真撮影。
↓
画像をクラウドに置く。
↓
遠隔地・好きな時間に
作業者が画像を開い
て、データ入力。

数名に入力を依頼するだけであれば、
Dropbox等に画像を置いて、Excel等にデータを入力してもらうだけ。
特別な仕組みは不要。

さらに大人数で入力をするためには、クラウド作業用のシステムが必要

人文系での成功事例

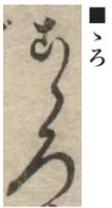
みんなで翻刻

<https://honkoku.org/>

市民が参加して、歴史資料の文字を読み解くプロジェクト。
(AIも併用)
2017年～

テスト：変体仮名 その1 (あ～か行)

■の文字は何で読むかな？ 1文字で答えてね



回答を入力

閉じる 回答



みんなで翻刻

MINNA DE HONKOKU

歴史資料を読み解く 歴史資料を解読するために お問い合わせ EN 参加する

みんなで翻刻

MINNA DE HONKOKU

仲間と、AIと協力して、
古文書解読に参加しよう！

JOIN
参加する

※現在新バージョンの試験運用中です。
安定稼働するまで旧バージョンも併せてご利用ください。

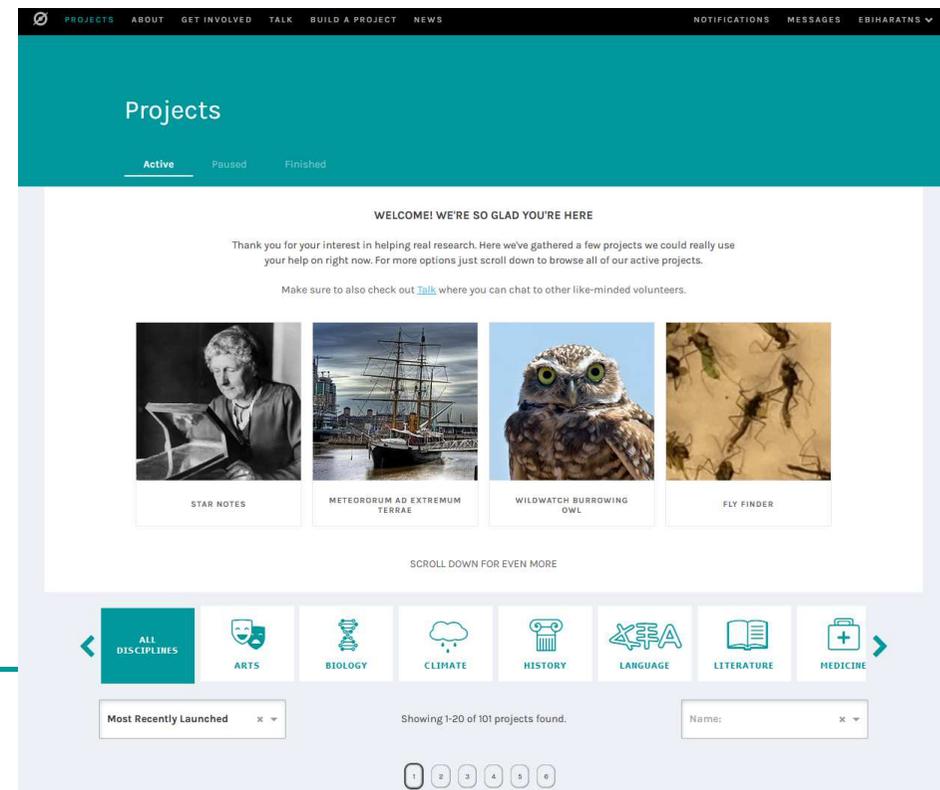
いいね! ツイート ブックマーク

完成度の高いシステムだが、自然史標本に流用するのは無理がある

Zooniverse (ズーニバース)

<https://www.zooniverse.org/>

- クラウドソーシングによる市民科学のプラットフォーム。
- 銀河の画像をボランティアが分類するプロジェクトが母体。
(英国と米国が中心で運営)
- 現在では、無数の自然科学系・人文系のプロジェクトが活動。
自然史系は
”Notes of Nature”の名称。
- ボランティアは、画像を開いて、指示された作業を行う。



Zooniverseのプロジェクトの例

“Digging Into Burrowing Owl Conservation”

減少しているフクロウの生態を調べるために、

- ・フクロウの個体数
- ・行動
- ・獲物

の情報を

95317枚の画像から読み取って入力。

1枚の画像に対して10人が
独立に情報を入力する。

2022-03-14 - WORKING

Retirement limit: 10

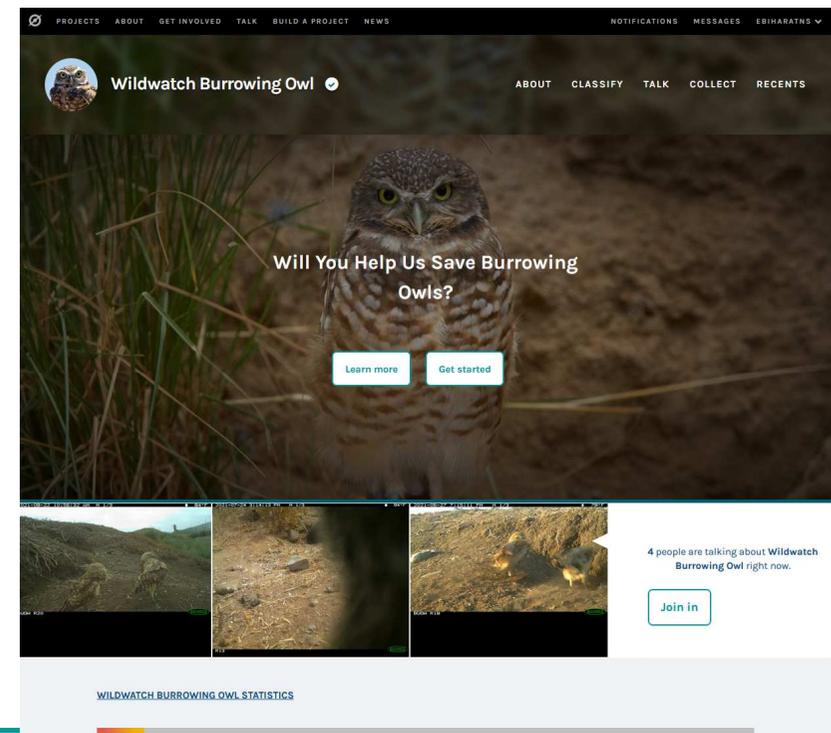
Images retired: 6,878 / 95,317

Classifications: 560,847 / 953,170

ETC* 70 days



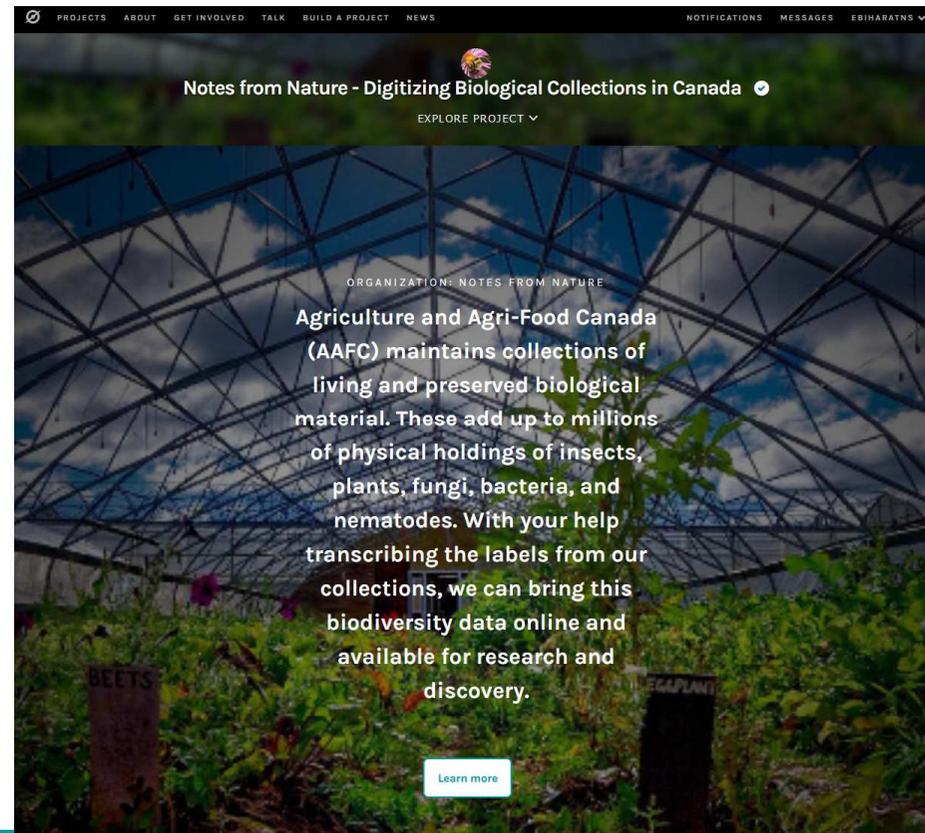
2022年3月に開始。
現在59%完了。



Zooniverseのプロジェクトの例

“Digitizing Biological Collections in Canada”

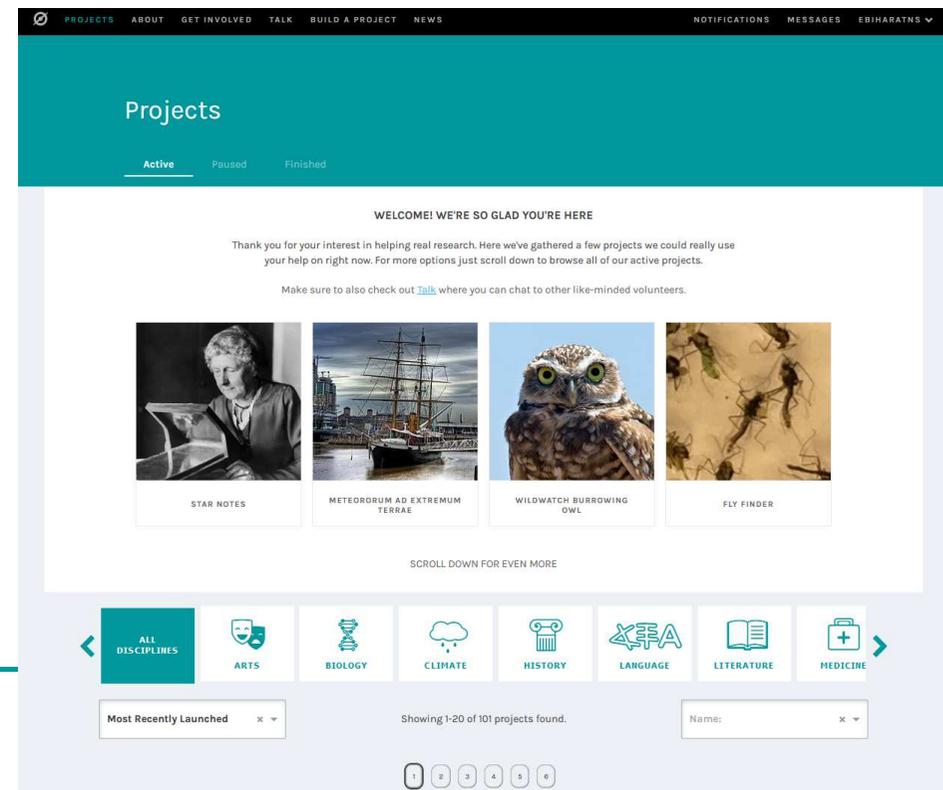
カナダにある生物標本のラベル情報を読み取って入力する。



Zooniverseの長所

- 多くのユーザーと歴史を持つシステムで、完成度が高い。
- 無料でプロジェクトの作成が可能。(画像枚数上限1万点)
- 入力の「タスク」を自由に組み合わせてプロジェクトを作成できる。

短所は、みなさんに体験していただいた後でまとめます。

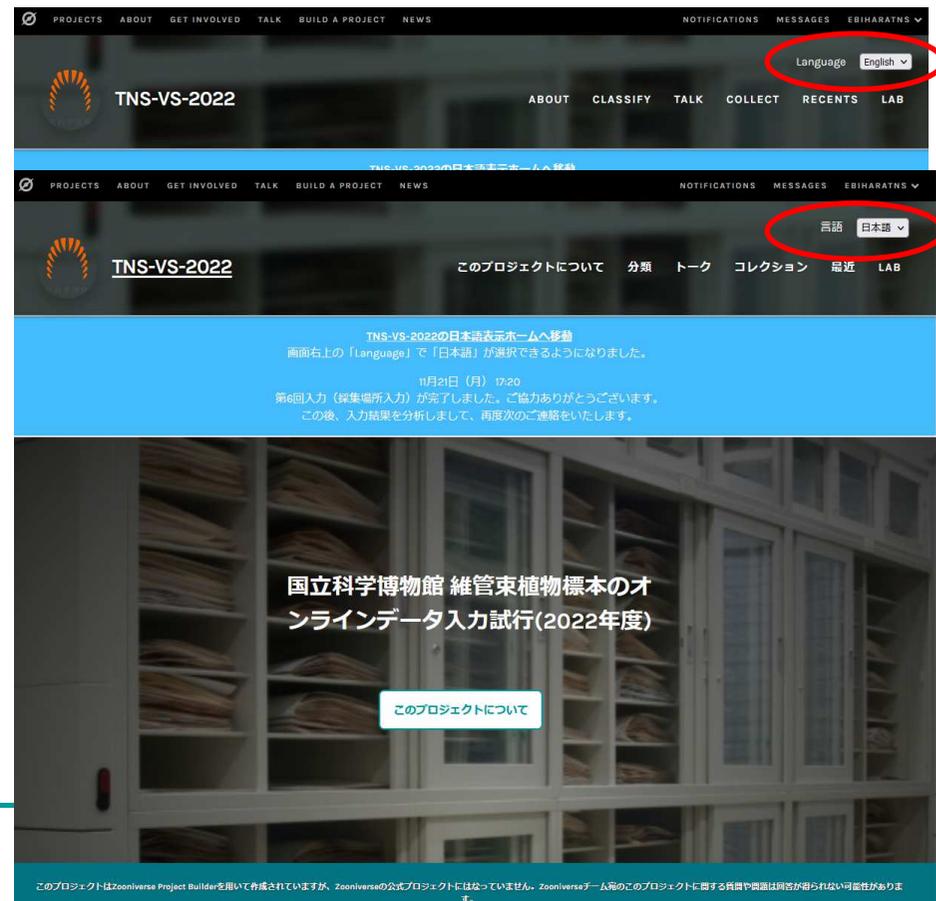


Zooniverseの長所

- 日本発のプロジェクトは見当たらないものの、一応表示メッセージの「日本語化」が可能。

(英語表示のままのメッセージも残っており、不完全)

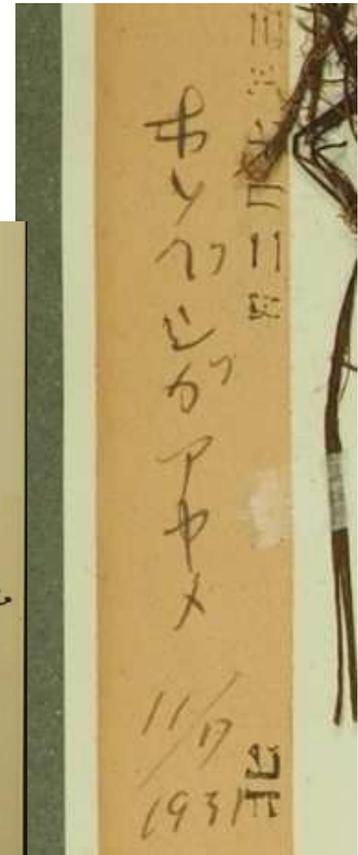
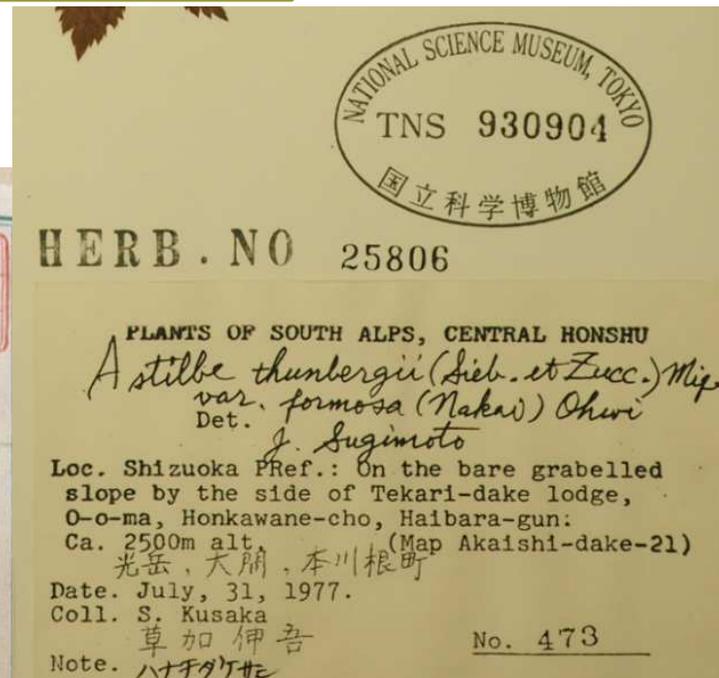
短所は、みなさんに体験していただいた後でまとめます。



今回は、国立科学博物館の維管束植物標本を例に、入力を体験していただきます

- AIによる文字認識は期待できず、人の力が頼り。
理由：情報の記載位置に一定の規則がない。

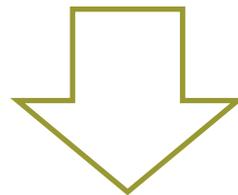
人力入力の前に、十分なトレーニングが不可欠



-
- 情報があまりに「雑多」なので、全ての情報を入力するのはハードルが高い。



- Zooniverseのシステムの習熟も兼ねて、単純な入力から、複雑な入力へ、レベルを順次上げていく



- 標本登録番号(通称「TNS番号」)の入力 から
-

登録番号 (TNS番号) の入力

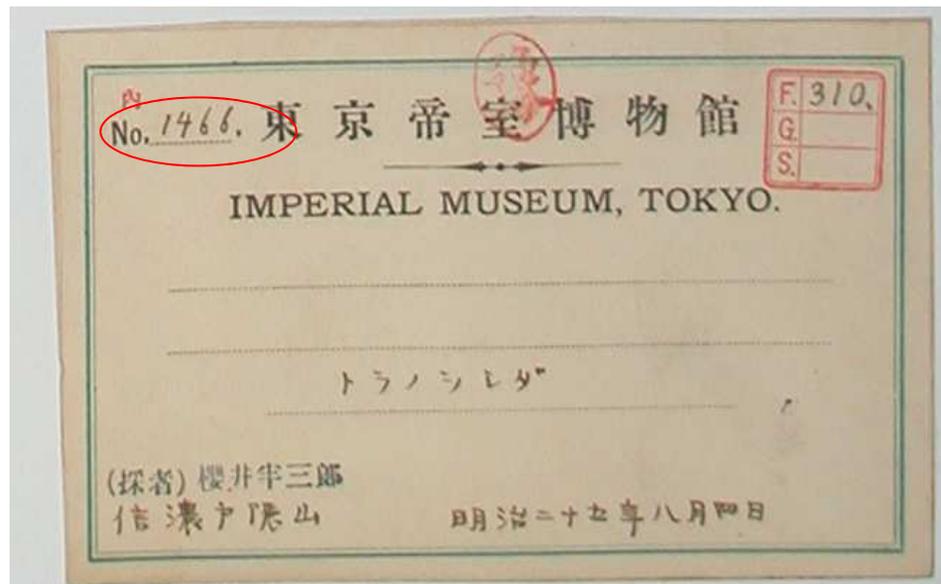
- 標本の「個体識別」ID・・・大変重要
 - 入力ミスがあると以後の作業に大きく影響するので、慎重な入力が求められる。

 - 全ての標本についているが、年代によって**番号の書式や位置が異なる**ので要注意。
-

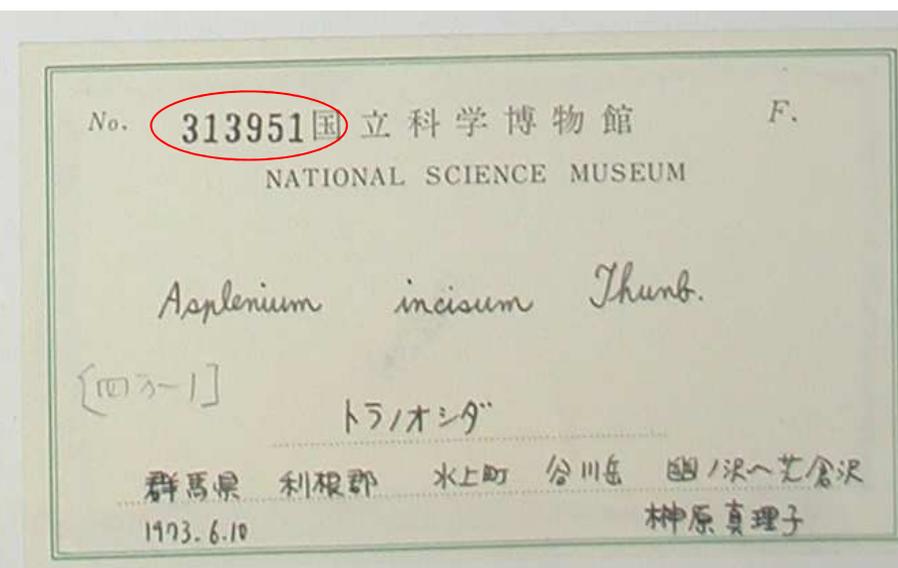
【TNS番号(1)】

ラベルの左上(初期の標本)

「No.」の後に数字が記されている

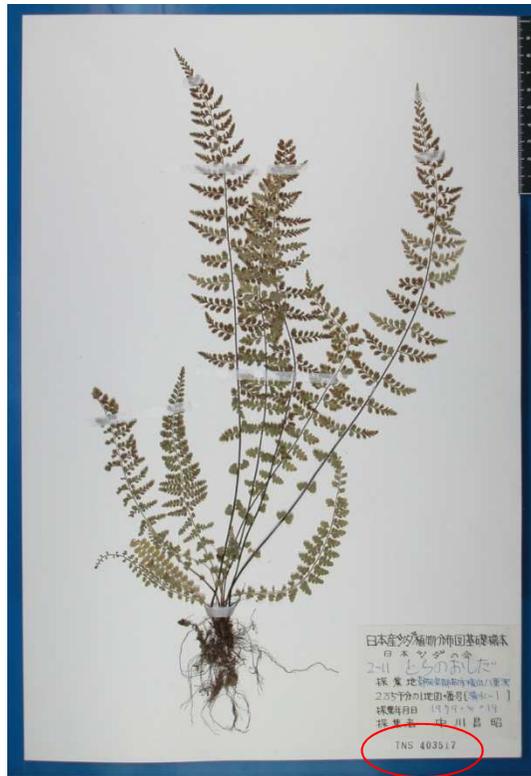


(最初期は手書き)

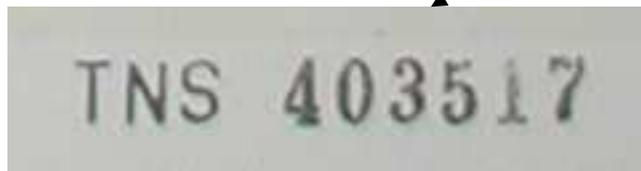


(6桁になった頃からスタンプ)

【TNS番号(2)】



フレーム内に番号があるパターン。



通常は台紙の右下*に、
「TNS」の後に数字で刻印されている
(頭に「TNS」がついていない場合もあり)

ごく稀に右上やその他の
スペースの場合あり

【TNS番号(3)】

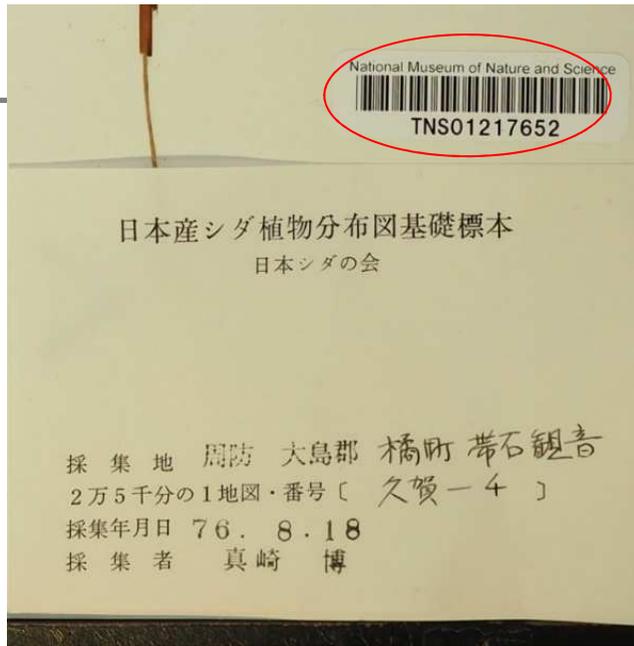
900万台のTNS番号がつけられた、旧筑波地区の標本もあり



TNS 9016728



【TNS番号(4)】



100万番以降はバーコードが使われている。

頭の「TNS0」は入力不要なので、7桁の数字部分のみを入力する。

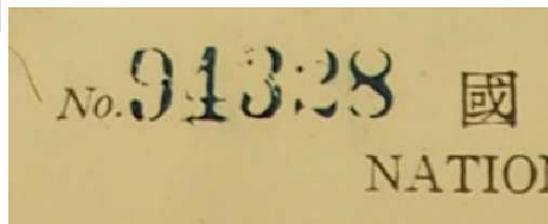


【TNS番号(5)】

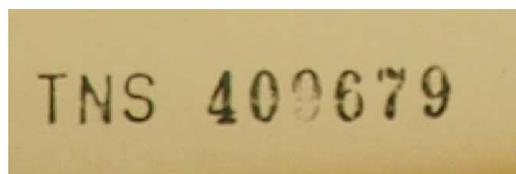


2箇所には番号がついているものがある。「TNS」が付いていたり、近くに「国立科学博物館」と書かれている方が、入力すべき登録番号である。(他機関から移管された標本では、旧所蔵機関のNo.がついていることがあるので要注意)

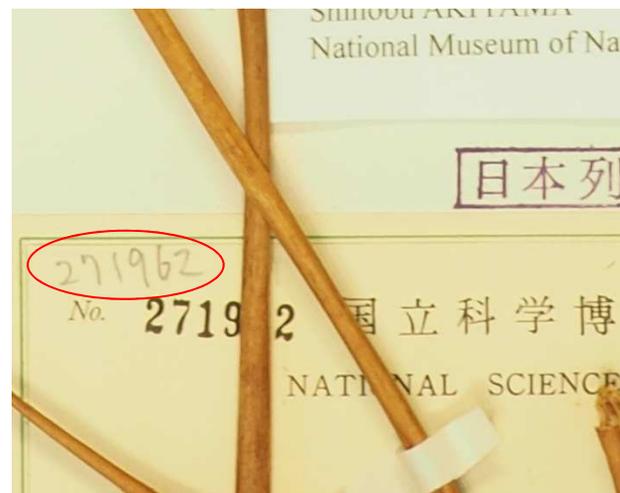
【TNS番号(6)】



94328



409679



スタンプがかすれている場合もあるので、拡大してしっかり確認する。

●多い読み間違い

0と9

1と4

3と8

TNS番号が、標本に隠れている場合には、近傍に手書きで番号をメモしている場合があります。

Zooniverseのプロジェクトページにアクセスしてみましょう

<https://www.zooniverse.org/projects/ebiharatns/tns-vs-2022?language=ja>



右上の
SIGN IN
をクリック

SIGN IN
をクリック

Get started ↓

100年以上をかけて収集された標本に記された情報の書式やレイアウトは多種多様。簡単な作業から専門員として、標本を管理する上で大変重要な登録番号（TNS番号）の入力からスタートします。登っていただく狙いもあります。材料として、豊居調査で採集された標本の画像を用います。複数の桁の数字の登録番号が右下に押印されています。

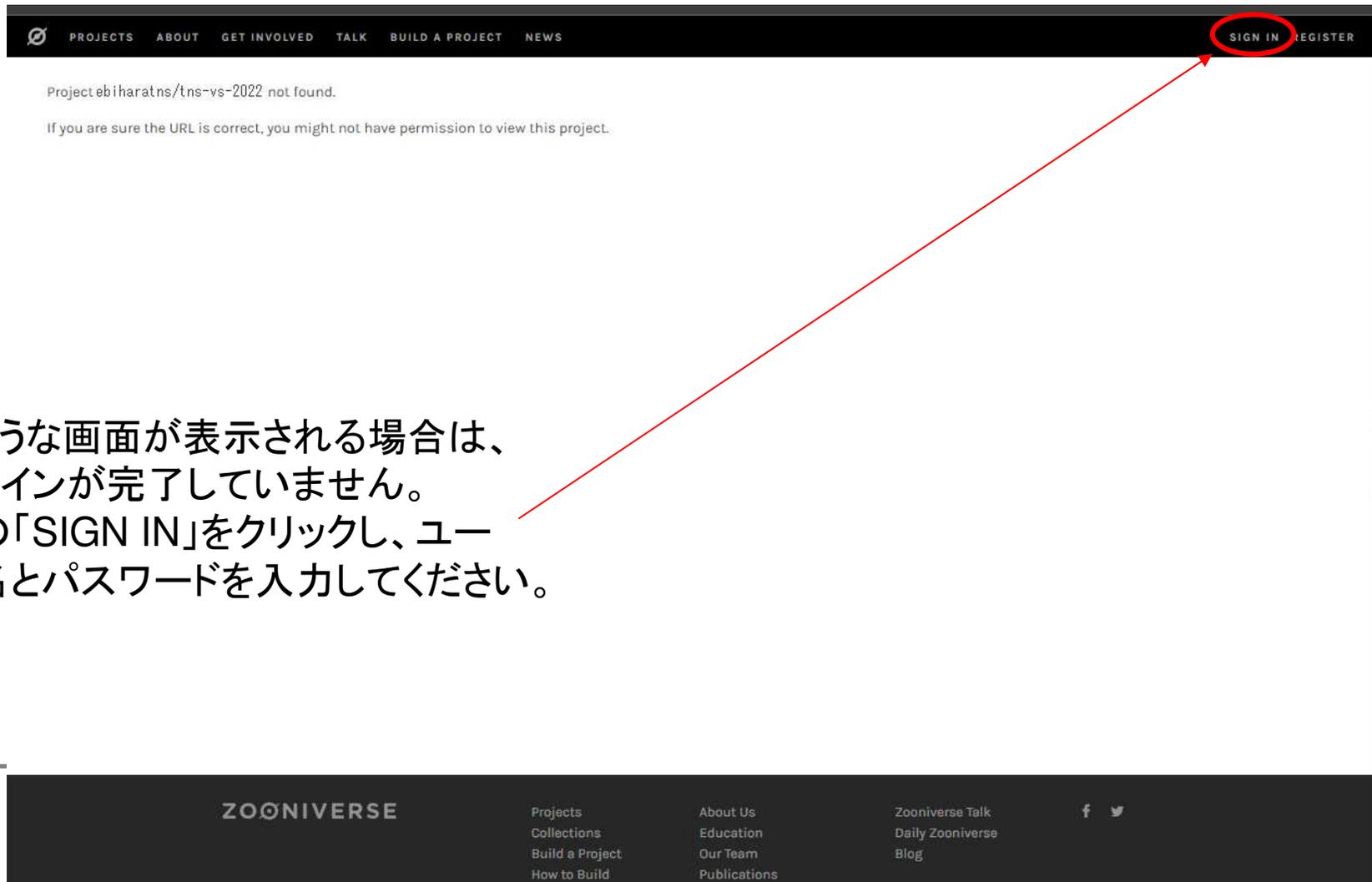
TNS番号読み取り(入門編-皇座2)

TNS番号読み取り(入門編-皇座3)

TNS番号読み取り(入門編-皇座4)

TNS番号読み取り(入門編-皇座5)

プロジェクトページが開けない場合



The screenshot shows a dark navigation bar at the top with the following menu items: PROJECTS, ABOUT, GET INVOLVED, TALK, BUILD A PROJECT, NEWS, and SIGN IN (circled in red). Below the navigation bar, the main content area displays a 404 error message: "Project ebiharatns/tns-vs-2022 not found." followed by "If you are sure the URL is correct, you might not have permission to view this project." A red arrow points from the text below to the "SIGN IN" button.

このような画面が表示される場合は、サインインが完了していません。右上の「SIGN IN」をクリックし、ユーザー名とパスワードを入力してください。

ZOONIVERSE

Projects
Collections
Build a Project
How to Build

About Us
Education
Our Team
Publications

Zooniverse Talk
Daily Zooniverse
Blog



下にスクロールすると、データ入力待ちのリストが表示されます。

This project has been built using the Zooniverse Project Builder but is not yet an official Zooniverse project. Queries and issues relating to this project directed at the Zooniverse Team may not receive any response.

Get started ↓

100年以上をかけて収集された標本に記された情報の書式やレイアウトは多種多様。簡単な作業から、徐々に複雑な作業にレベルを上げていく計画です。まず、入門編として、標本を管理する上で大変重要な登録番号（TNS番号）の入力からスタートします。登録番号（TNS番号）の読み取りと同時に、本システムの操作にも慣れていただく狙いもあります。材料として、皇居調査で採集された標本の画像を用います。複数のフォルダに分かれていますが、どれも作業難易度は同じです。6桁の数字の登録番号が右下に押印されています。

TNS番号読み取り(入門編-皇居2) TNS番号読み取り(入門編-皇居3) TNS番号読み取り(入門編-皇居4) TNS番号読み取り(入門編-皇居5)

TNS-VS-2022 STATISTICS

0% Complete

0	0	435	0
Volunteers	Classifications	Subjects	Completed Subjects

1 person is talking about TNS-VS-2022 right now.

Join in

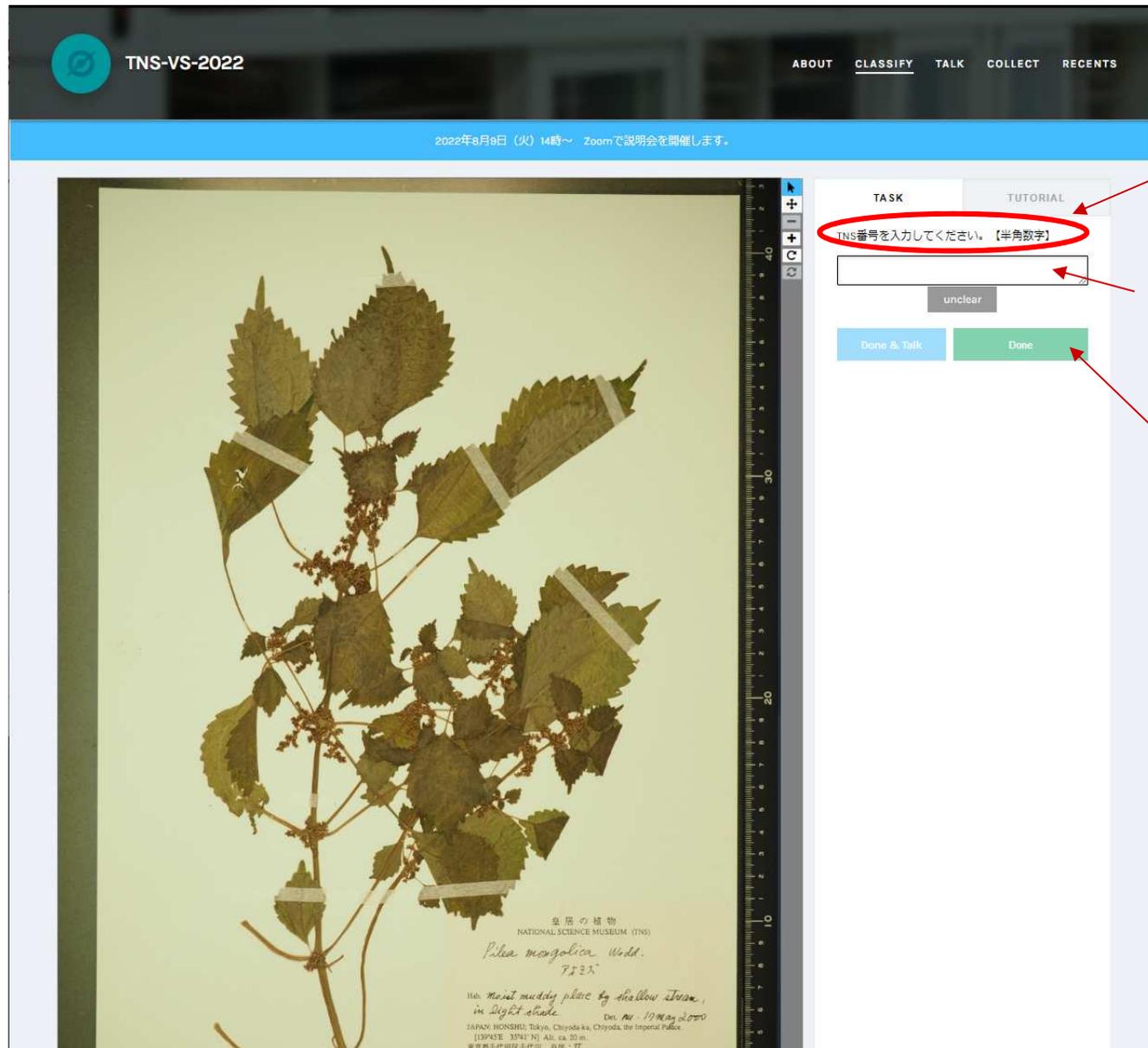
ABOUT TNS-VS-2022

日本における自然史標本情報のデジタル化へのクラウドソーシング導入に向け、国立科学博物館のボランティアさんに参加いただいてオンラインデータ入力の試行を実施します(2022年8月～)。



入力待ち
どれかを選ぶ

標本画像と、問い(入力指示)が表示されます



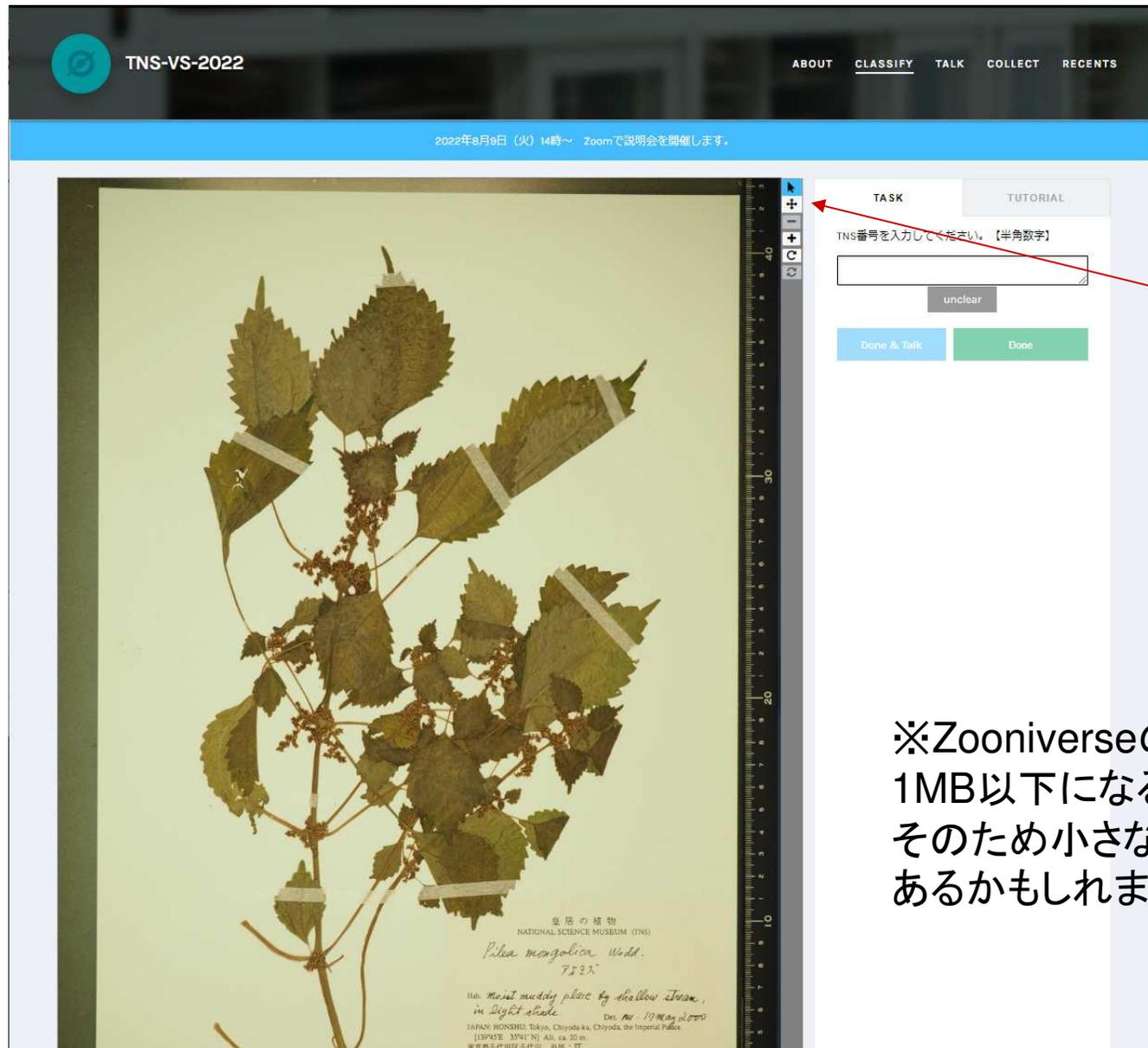
ここに問い(何を入力するか)が表示されます

この窓に答え(画像から読み取った文字列)を入力します

入力したら緑の「Done」ボタンを押します。

次の標本画像に切り替わったら、同様に入力を繰り返します。
(いつでも作業を中断して問題ありません)

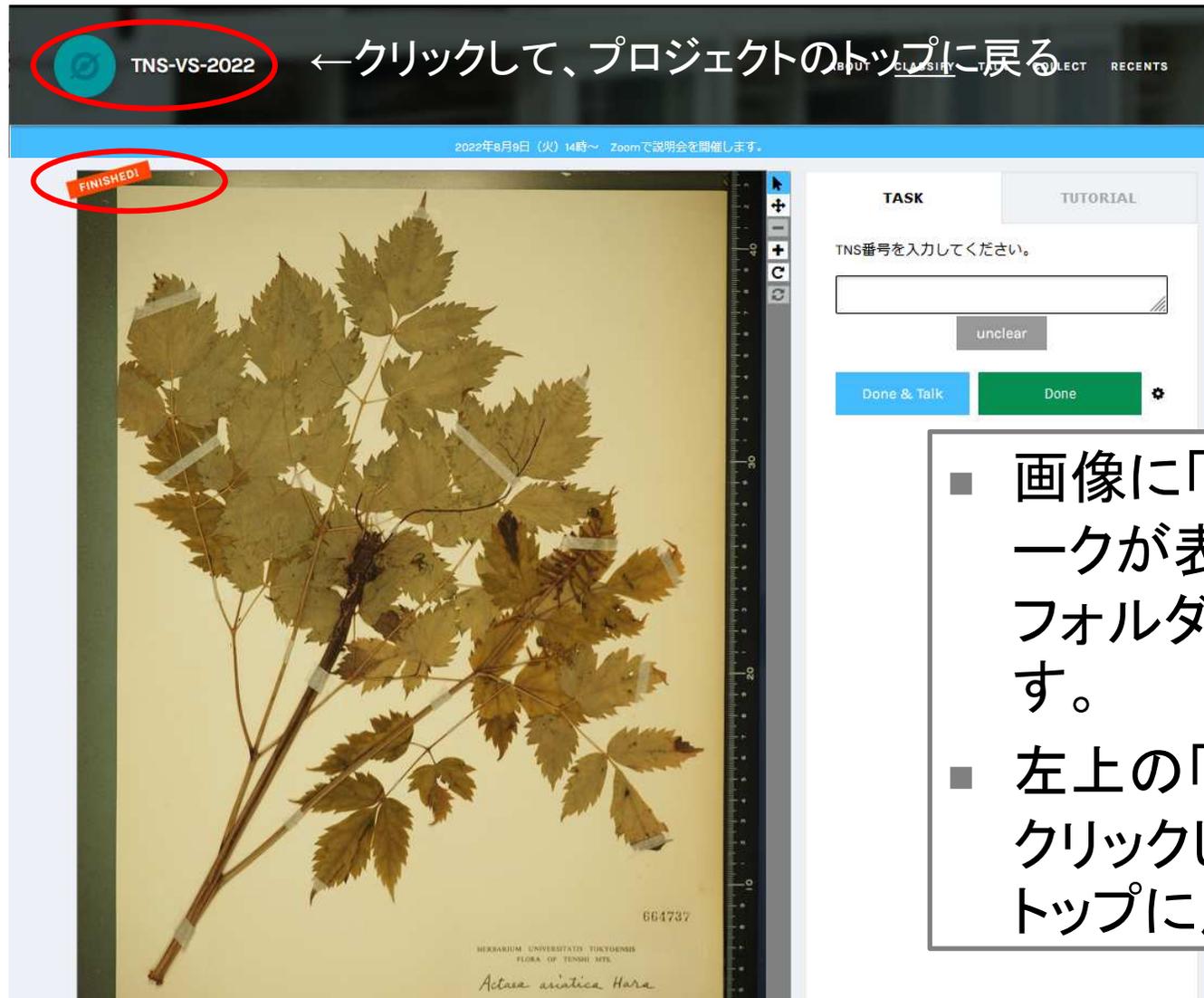
画像は拡大できます



画像を拡大したいときは、
+を押します。
(画像の上でマウスを操
作してZoomすることも
可能)

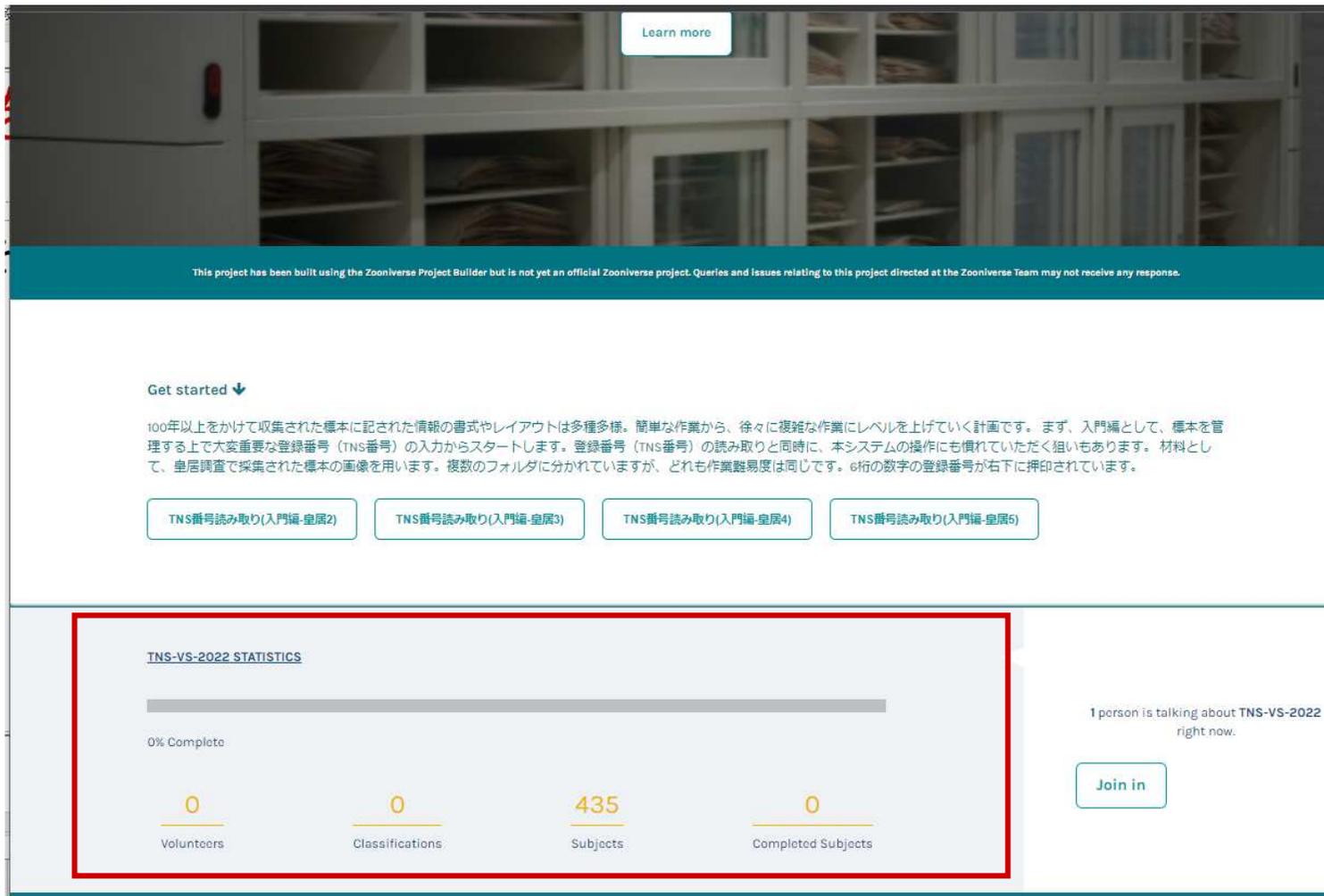
※Zooniverseの制限のため、画像は
1MB以下になるよう縮小されています。
そのため小さな文字が読み辛い場合が
あるかもしれません。

左上にFINISHEDと表示される画像は、入力が完了しています。



- 画像に「FINISHED」のマークが表示されたら、そのフォルダの作業は完了です。
- 左上の「TNS-VS-2022」をクリックして、プロジェクトのトップに戻ってください。

プロジェクトトップページの下部に 進捗状況が表示されます



Learn more

This project has been built using the Zooniverse Project Builder but is not yet an official Zooniverse project. Queries and issues relating to this project directed at the Zooniverse Team may not receive any response.

Get started ↓

100年以上をかけて収集された標本に記された情報の書式やレイアウトは多種多様。簡単な作業から、徐々に複雑な作業にレベルを上げていく計画です。まず、入門編として、標本を管理する上で大変重要な登録番号（TNS番号）の入力からスタートします。登録番号（TNS番号）の読み取りと同時に、本システムの操作にも慣れていただく狙いもあります。材料として、皇居調査で採集された標本の画像を用います。複数のフォルダに分かれていますが、どれも作業難易度は同じです。6桁の数字の登録番号が右下に押印されています。

TNS番号読み取り(入門編-皇居2) TNS番号読み取り(入門編-皇居3) TNS番号読み取り(入門編-皇居4) TNS番号読み取り(入門編-皇居5)

TNS-VS-2022 STATISTICS

0% Complete

0	0	435	0
Volunteers	Classifications	Subjects	Completed Subjects

1 person is talking about TNS-VS-2022 right now.

Join in