

2023年2月18日

第40回自然史標本の発信に関する研究会

日本の自然史標本における クラウドソーシング利用の 課題と期待

海老原 淳

(国立科学博物館植物研究部)

国立科学博物館での Zooniverse試行概要

- 2022年8月1日～11月21日
- 参加者: 科博上野本館および筑波ボランティア
(植物関係のグループに所属している方々を非公式に勧誘)
参加希望**15**名、実際の参加者**13**名
- のべ入力件数： **9929**件
(最大入力数ユーザー： **1577**件[**15.9%**])
(最少入力数ユーザー： **289**件[**2.9%**])

入力内容

対象はいずれも植物さく葉標本

- 第1回： 標本登録番号の入力 [retirement=5]
- 第2回： 標本登録番号の入力 [retirement=3](以後全て)
- 第3回： 標本登録番号の入力
+ トーク機能の試行
- 第4回： 標本登録番号 + 採集年月日の入力
- 第5回： 採集者名 + 採集者番号の入力
- 第6回： 採集場所の入力

- 徐々にステップアップして、複雑な情報の入力に進め、最終的には必要な全てのラベル情報の入力を試行した。
- それぞれの入力内容に対応したマニュアルを作成した。

Zooniverseの短所

(Zooniverse側の仕様のため、基本的には改善の余地なし)

1. ユーザー単位の達成度表示機能がない

プロジェクト全体の進捗状況のインジケーターはあるが、各ユーザーがどれだけ貢献したのかを表示する機能はない。

ユーザーに達成感を感じさせる仕組みが存在しない。

参加者のモチベーションを維持する仕掛けを、別途用意する必要あり

単なる「労働力の搾取」にならないような配慮が必要

Zooniverseの短所

(Zooniverse側の仕様のため、基本的には改善の余地なし)

2. 入力者へのフィードバック機能がない

「トーク機能」を利用すると、疑問がある標本画像について質問を入力できる。しかし、回答が得られたとしても、元の標本に戻って入力する機能はない。

管理者が「正答」を入力者に知らせる仕組みがない。

参加者の能力向上の機会を、別途設ける必要あり

Zooniverseの短所

(Zooniverse側の仕様のため、基本的には改善の余地なし)

3. 複数人の入力結果の統合が難しい

Zooniverseは、同一画像に対して複数人が繰り返し入力することを推奨しており、その繰り返し回数を“retirement”として指定できる。

同一標本に対して繰り返し入力されたデータを統合・比較する機能はZooniverseにはなく、**export**したファイル进行分析する必要がある。

年月日のような数値であれば、**retirement**の値を大きくすることで、「多数決による自動化」はある程度可能。

しかし、複雑な文字列では、**retirement**の値を大きくすると、結果の揺れが生じて収束しない可能性が高い。**Retirement=3**に設定して手作業で正答を確認する手順が最も効率的に思われた。

Zooniverseの短所

(Zooniverse側の仕様のため、基本的には改善の余地なし)

3. 複数人の入力結果の統合が難しい

数値であれば、retirementの値を大きくすることで、「多数決による自動化」がある程度可能。

The screenshot shows a data entry interface for a Zooniverse project. At the top, it displays '928138.JPG' and '採集場所読み取り(ヤマハ)'. Below this is a table with columns for '国' (Country), '地域' (Region), '旧国名' (Old Country Name), '都道府県' (Prefecture), '郡' (County), '市町村' (City/Town/Village), '区' (Ward), '島' (Island), '詳細地名' (Detailed Location Name), '標高low' (Low Elevation), '標高high' (High Elevation), and 'mesh code'. The table contains four rows of data, all with '越前' (Echigo) in the '地域' column and '練兵場' (Rienbajou) in the '詳細地名' column. The '市町村' column contains '岡本村' (Okamura) in the first two rows, '岡松' (Okasaka) in the third row, and '岡野村' (Okano) in the fourth row. A red box highlights the '市町村' column across all rows.

複雑な文字列では、retirementの値を大きくすると、結果の揺れが生じて収束しない可能性が高い。
(特に日本語の場合、全角/半角でバリエーションが増えやすい)

回答の信頼性が高い「エキスパート」を認定することで、入力結果を重みづけする処理なども考えられる。

Zooniverseの短所

(Zooniverse側の仕様のため、基本的には改善の余地なし)

4. 日本語化が完全ではない

- 主要なメッセージは日本語化できているが、随所に英語表記が残る。
- ページを辿っているうちに、言語選択が英語に戻ってしまう。
- ユーザー登録は英語しか選択できない。

英語に抵抗のあるユーザーにとっては、敷居が高い

Zooniverseの短所

(Zooniverse側の仕様のため、基本的には改善の余地なし)

5. プロジェクトをオープンにする場合、英語表記が必須

- 現在はクローズドで運用。オープンにするためには実績と審査がある。(オープンにすると世界の誰でも参加可能になる)
- オープンにするためには、英語による説明文が必須になる。

「国内向けにだけオープン」にはできない。

Zooniverse 試行入力時の苦勞

(運用で改善の余地あり)

1. 画像サイズ制限

Zooniverseのサーバーにアップロードできる画像サイズは最大**1MB**の制限あり。このサイズでは不十分な場合がある。

画像の縮小処理に手間がかかる。



外部サーバーに画像を置くことで、**1MB**以上の画像が登録可能になる。

ただし、サイズを大きくし過ぎても、読み込みに時間を要するなどの問題を引き起こす可能性がある。

Zooniverse 試行入力時の苦勞

(運用で改善の余地あり)

2. 手書き文字の読み取り

手書き文字の解読、特に人名・地名などの固有名詞で時間を要している傾向が見られる。



Zooniverseは文字認識機能は持たない。

入力済みデータに基づいた辞書の提供などで、負担を軽減することが考えられる。

Zooniverse 試行入力時の苦勞

(運用で改善の余地あり)

3. 多様なパターンの情報に対応する困難

マニュアルで対応しきれない例外の出現など



多様なパターンに対応できる完成度の高いマニュアルの作成。

情報量が多くなりすぎても読まれないので、「困ったときに容易に引ける事例集」のようなものが有用か。

Zooniverse 試行入力時の苦勞

(運用で改善の余地あり)

4. JSON形式で出力される入力結果

慣れていないと編集等が難しい形式



情報の抽出・変換はExcelのPower Query機能で可能。

```
classification_id,user_name,user_id,user_ip,workflow_id,workflow_name,workflow_version,created_at,gold_standard,expert,metadata,subject_data,subject_ids,
437253219,masaaki-sakauchi,2499607,e99da5ff1b720d36ccb7,22311,TNS番号+採集日読み取り(イブ<br>キジャコウソウ),1.5,2022-09-08 00:07:21 UTC,...,{"source":,"api_session":,"badd136c79<br>7cd2a6b6ce45f35e7cc953eb7e536813f7e877d73ef10c84acd0ac","viewport":{"width":1232,"heig<br>ht":569},"started_at":"2022-09-08T00:05:59.423Z","user_agent":"Mozilla/5.0 (Windows<br>NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/105.0.0.0 Safari/537.36 E<br>dg/105.0.1343.27","utc_offset":-32400,"finished_at":"2022-09-08T00:07:22.147Z"...<br>live_project":true,"interventions":{"opt_in":true,"messageShown":false},"user_langua<br>ge":{"ja","user_group_ids":[],"subject_dimensions":{"clientWidth":384,"clientHeigh<br>t":512,"naturalWidth":3456,"naturalHeight":4608}},"subject_selection_state":{"retire<br>d":false,"selected_at":"2022-09-08T00:05:57.867Z","already_seen":false,"selection_st<br>ate":"normal","finished_workflow":false,"user_has_finished_workflow":false},"workflo<br>w_translation_id":null},"[{"task":"T1","value":894728,"task_label":"TNS番号を<br>入力してください。【半角数字】"},{"task":"T2","value":1980,"task_label":"採集年<br>月日の表記はありますか?【半角数字】"},{"task":"T3","value":1980,"task_label":"採集年<br>月日の表記はありますか?【半角数字】"},{"task":"T4","value":24,"task_label":"採集<br>月を入力してください。【半角数字】"},{"task":"T5","value":24,"task_label":"採集<br>月を入力してください。【半角数字】"},{"task":"T6","value":24,"task_label":"採集<br>月を入力してください。【半角数字】"},{"task":"T7","value":24,"task_label":"採集<br>月を入力してください。【半角数字】"},{"task":"T8","value":24,"task_label":"採集<br>月を入力してください。【半角数字】"},{"task":"T9","value":24,"task_label":"採集<br>月を入力してください。【半角数字】"},{"task":"T10","value":24,"task_label":"採集日<br>は複数の日に渡っていますか?【半角数字】"}],{"id":"60370135","retired":false,<br>"id":"107648327","workflow_id":22311,"classifications_count":3,"created_at":"2022-09-0<br>7T23:18:34.384Z","updated_at":"2022-09-08T05:20:00.497Z","retired_at":"2022-09-08T05</pre>
```

gold_standard	expert	metadata	annotations.task	annotations.value	subject_data
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	724853	TNS番号を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	1976	採集年月日の表記はありますか?	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	1976	採集年月を入力してください【半角	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	20	採集日を入力してください【半角	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	10	いいえ、1日のみです。	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	756233	TNS番号を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	72	表記あり	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	2002	採集年月日の表記はありますか?	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	8	採集年月を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	16	採集日を入力してください【半角	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	10	いいえ、1日のみです。	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	763299	TNS番号を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	72	表記あり	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	2007	採集年月日の表記はありますか?	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	8	採集年月を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	15	採集日を入力してください【半角	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	10	いいえ、1日のみです。	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	944352	TNS番号を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	72	表記あり	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	1938	採集年月日の表記はありますか?	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	4	採集年月を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	3	採集日を入力してください【半角	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	10	いいえ、1日のみです。	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	928144	TNS番号を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	72	表記あり	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	1936	採集年月日の表記はありますか?	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]
		["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]	9	採集年月を入力してください【半	["source":"api","session":"acfb125b2e16e5d6899545e8d1a3d2f"]

※Zooniverseからのデータexportは、24時間に1回しかできない制約に注意。

標本ラベルの画像が用意できれば、
Zooniverse上に、誰でも簡単に、
クラウドソーシングのプロジェクト
を立ち上げることができます！

“Task”を組み合わせて“Workflow”を作成

採集者名・番号入力（ヤマハギ） #22327 

A workflow is the sequence of tasks that you're asking volunteers to perform. For example, you might want to ask volunteers to answer questions about your images, o

WORKFLOW TITLE

採集者名・番号入力（ヤマハギ）

If you let your volunteers choose which workflow to attempt, this text will appear as an option on the project front page.

TASKS

<input checked="" type="radio"/> 採集日は複数の日に渡っていますか？	T0
<input type="checkbox"/> TNS番号を入力してください。【半角数字】	T1
<input checked="" type="radio"/> 採集年月日の表記はありますか？	T2
<input type="checkbox"/> 採集年を西暦で入力してください。【半角数字】	T3
<input type="checkbox"/> 採集月を入力してください。【半角数字】	T4
<input type="checkbox"/> 採集日を入力してください。【半角数字】	T5
<input checked="" type="checkbox"/> 採集日の終わりの日付（年）を西暦で入力してください。【半角数字】	T6
<input type="checkbox"/> 採集日の終わりの日付（月）を入力してください。【半角数字】	T7
<input type="checkbox"/> 採集日の終わりの日付（日）を入力してください。【半角数字】	T8
<input type="checkbox"/> 採集者名（日本語）を入力してください。（表記がない場合は何も入力せずに次へ） <i>(first)</i>	T9
<input type="checkbox"/> 採集者名（ローマ字）を入力してください。（表記がない場合は何も入力せずに次へ） 【半角英数字】	T10
<input type="checkbox"/> 採集者番号を入力してください。（表記がない場合は何も入力せずに次へ） 【原則として半角英数字】	T11

[Add a task](#)

FIRST TASK

採集者名（日本語）を入力してください。（表記がない場合は何も入力せずに次へ）

A task is a unit of work you are asking volunteers to do. You can ask them to answer a question or mark an image. Add a task by clicking the question or marking buttons below.

MAIN TEXT

採集日の終わりの日付（年）を西暦で入力してください。【半角数字】

Describe the task, or ask the question, in a way that is clear to a non-expert. You can use markdown to format this text.

HELP TEXT



最初の日付と同じ年の場合でも、空欄にせず入力してください。

Add text and images for a help window.

Required

METADATA TAGS

Volunteers can attach the following tags to highlighted portions of their transcription.

Deletion

Insertion

Unclear

NEXT TASK

採集日の終わりの日付（月）を入力してください。【半角数字】

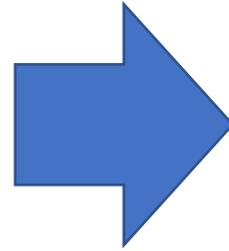
Delete this task

最後に“Subject sets”（画像）と組み合わせる

changes to a workflow, you create a new version. Big changes, like adding or deleting questions, will change the version by a whole number: 1.0 to 2.0, etc. Smaller changes,

電子化規模の拡大

- 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

撮影も加速化？

- 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。

クラウドソーシング
による加速化

将来への期待

- 日本において、自然史標本のクラウドソーシングは、まだ実績がほとんどありません。
- 未電子化標本の情報公開をスピードアップするための最有力手段と言えます。
- 当面は、無料で利用できるZooniverseを活用しつつも、将来的には「日本の自然史標本」にマッチしたシステムが欲しいところです。