Open Ecoacoustics

End to end workflows in Ecoacoustics: verifying AI output

Anthony Truskinger (QUT), Hudson Newey (QUT),
Nelli Holopainen (QUT), Paul Roe (JCU)

What is [Open] Ecoacoustics

- Ecoacoustics
  - Use sound to understand the environment
  - Sensors: cost effective, greater coverage
  - Analyze data → understand environment → make decisions
- Open Ecoacoustics
  - A platform for continental-scale ecological monitoring and research
  - Part of the Machine Observation Data Processing Infrastructure program within the ARDC Planet Research Data Commons

What is the Acoustic Workbench

- Open source web management software for Ecoacoustics
- White label application, 2 instances
  - Ecosounds: general data, user managed, public or private https://www.ecosounds.org/
  - The Australian Acoustic Observatory: one big distributed experiment, all CC BY 4.0 https://data.acousticobservatory.org/
- Project owners control permissions
  - Individual access, public access, downloading of audio
- Upload, store, share, and download audio data
- Import, create, verify, and download annotations (species detections)
- Run analyses automatically*

Stats

- 261 Projects
- 5600 Sites
- 1001 TB
- 6.5M files
- 713 years of audio
- 4.9M annotations
  - 42k verified

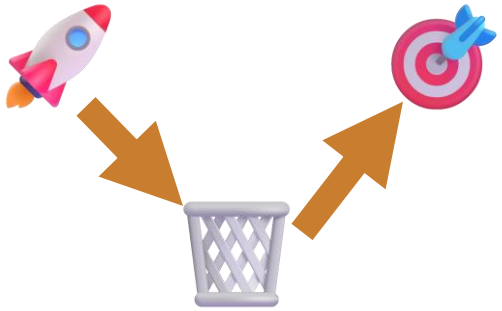
Workflow

BirdNet,
Perch,
custom classifiers,
etc..

...not interested in exporting junk

Create → Upload Audio → Analysis → Verify → Report → Export

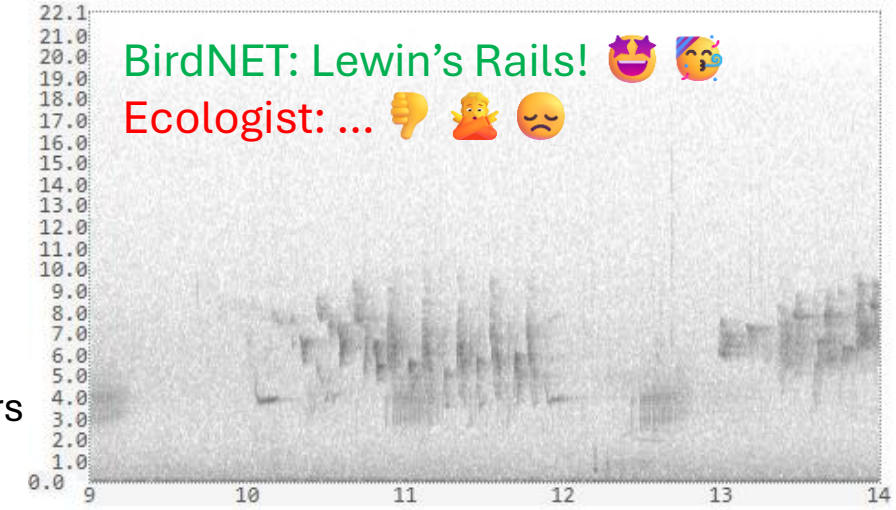
✅ ✅



Millions of false positives

> 100 detections/hour ✕ > 6 million hours

≋ 600 million events???



Verification

- Verifying labels are correct is important
  - Agile learning loops & high-quality data still valuable
  - It must be *fast*
  - It must work for experts and novices
  - Is a distinct task from labelling
  - To integrate with our platform
- But existing software…

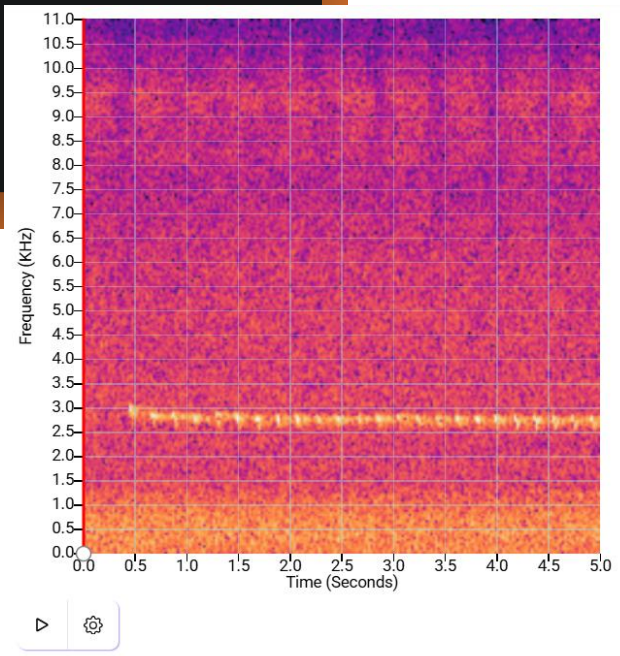
1. Via browser
2. Claimed to be open source but only distributed a Windows binary
3. Top request for expert users
4. Labelling can be customised to a verification task, but it's a hack
5. Reuses labelling functionality
6. Raven is made to work on one file at a time, users need to make a "super cut" of source audio segments – huge hack
7. Files needed locally
8. Files need to be pre-cut
9. "Collaboration" achieved by shipping hard drives of pre-cut audio files and CSV metadata and results collated by scientists manually

	Raven	Kaleidoscope	ARIEL	We need…
Web	✗	✗	✗	✓
Mobile	✗	✗	✗	✓
Desktop	✓	✓	✓	✓ [1]
Open source	✗	✗	? [2]	✓
Keyboard shortcuts[3]	✓	✓	?	✓
Made for verification	✊ [4]	✊ [5]	✓	✓
View any segment from any file	✊ [6] [7]	✊ [7]	✓ [8]	✓
Multi item (grid view)	✓	✗	✗	✓
Single item view	✓	✓	✓	✓
Collaboration	✊ [9]	✊ [9]	✊ [9]	✓

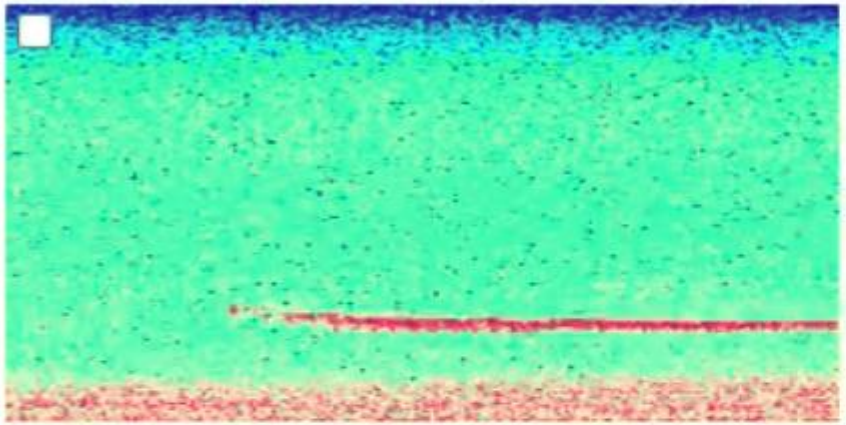
Verification: methodology

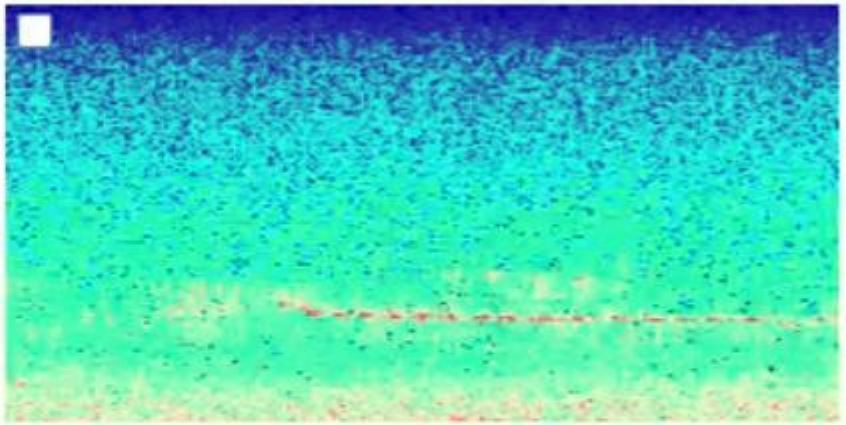
- Persona identification
- Requirements gathering
- Codesign, iterative development
  - >4 workshops, > 20 check ins, with ≈ 20 people, from 4 organisations
  - Fortnightly sprints
  - Features prioritised by greatest common need
- Web components
  - Web native UI components
  - We were already making them
    - See talk/rant “reimplementing a spectrogram component too many times”
  - Questionable
    - Is it really reusable?
    - Is the added complexity required?
    - Do the developers just like greenfield projects more?
  - Responsibility split:
    - Behaviour belongs to the component
    - Persisting state belongs to the host application

```
<oe-axes
    x-title="Time (Seconds)"
    y-title="Frequency (KHz)">
    <oe-indicator>
        <oe-spectrogram
            id="playing-spectrogram"
            src="/public/example.flac"
        ></oe-spectrogram>
    </oe-indicator>
</oe-axes>
<oe-media-controls
    for="playing-spectrogram"
></oe-media-controls>
```



Verification: iteration

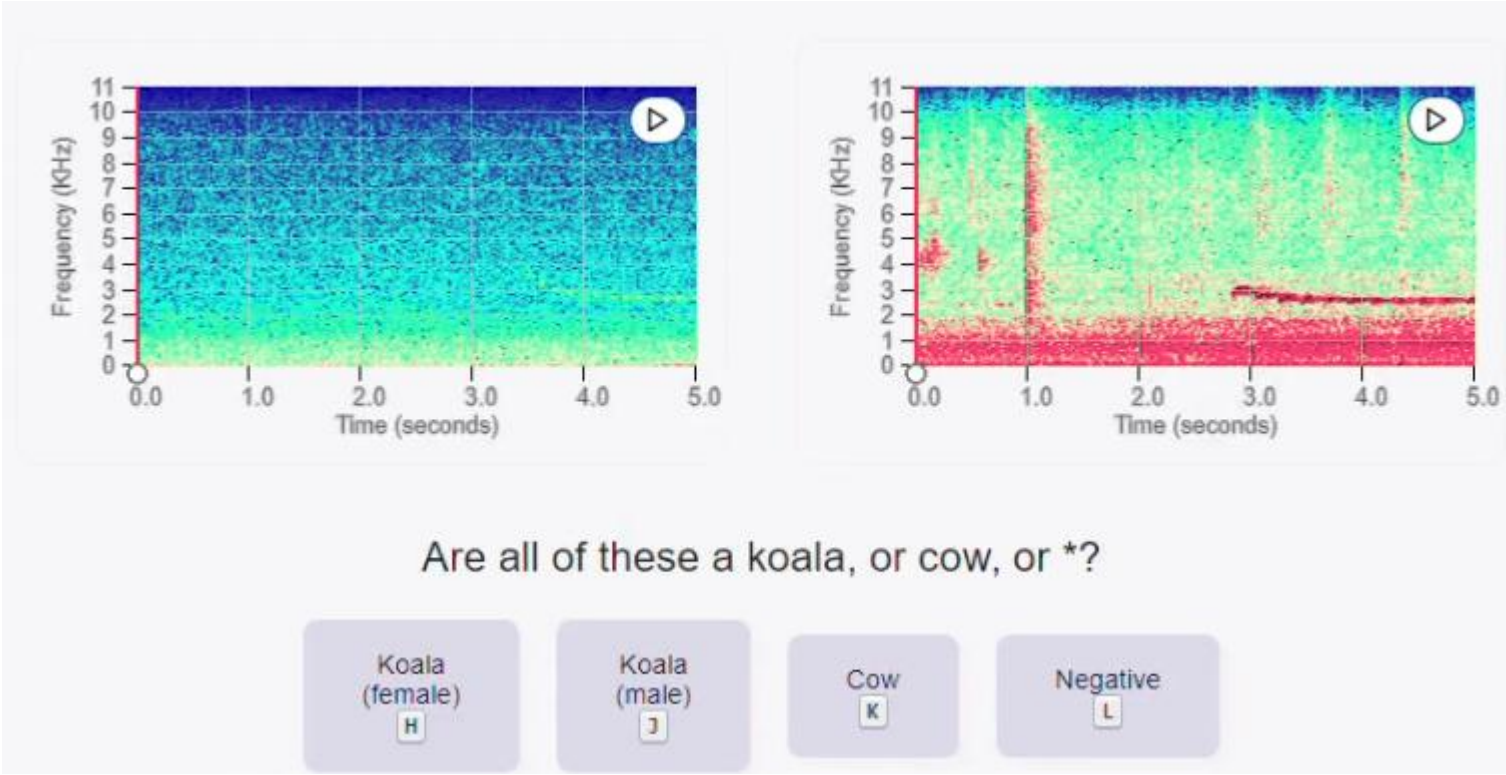




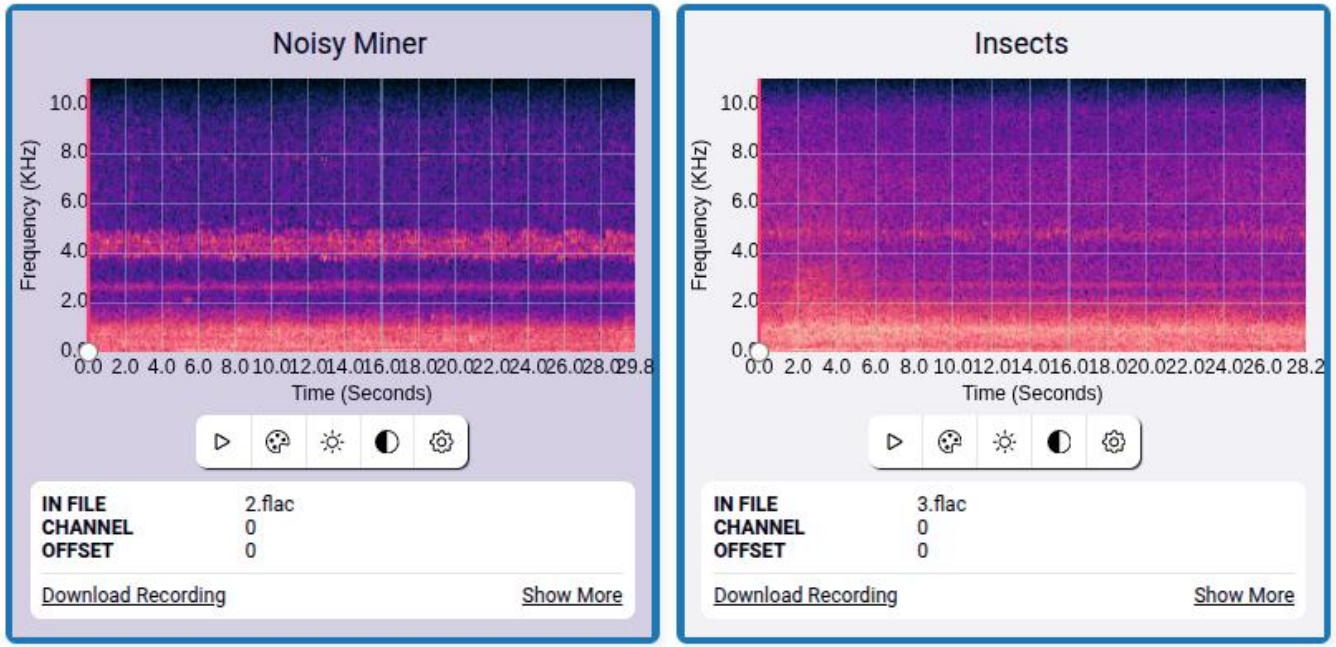
Are all of these a Koala?

Koala (J) | Not a Koala (K)

Verification: iteration



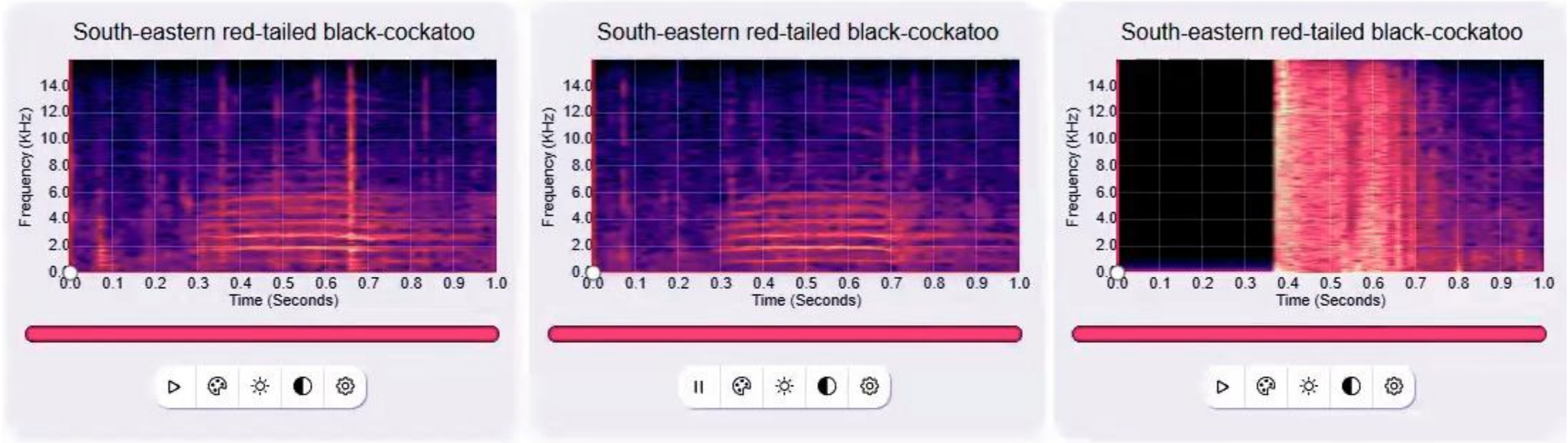
Verification: iteration



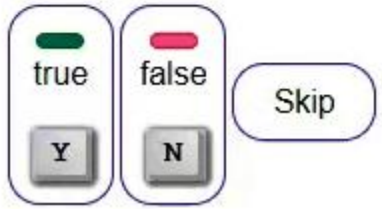
Do all of the selected audio segments have the correct applied tag

Yes	No	Skip
Y	N	`

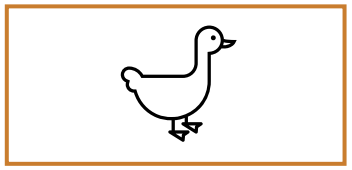
Verification: iteration



Do all of the selected audio segments have the correct applied tag

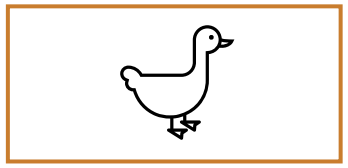


Verification: showing task progress

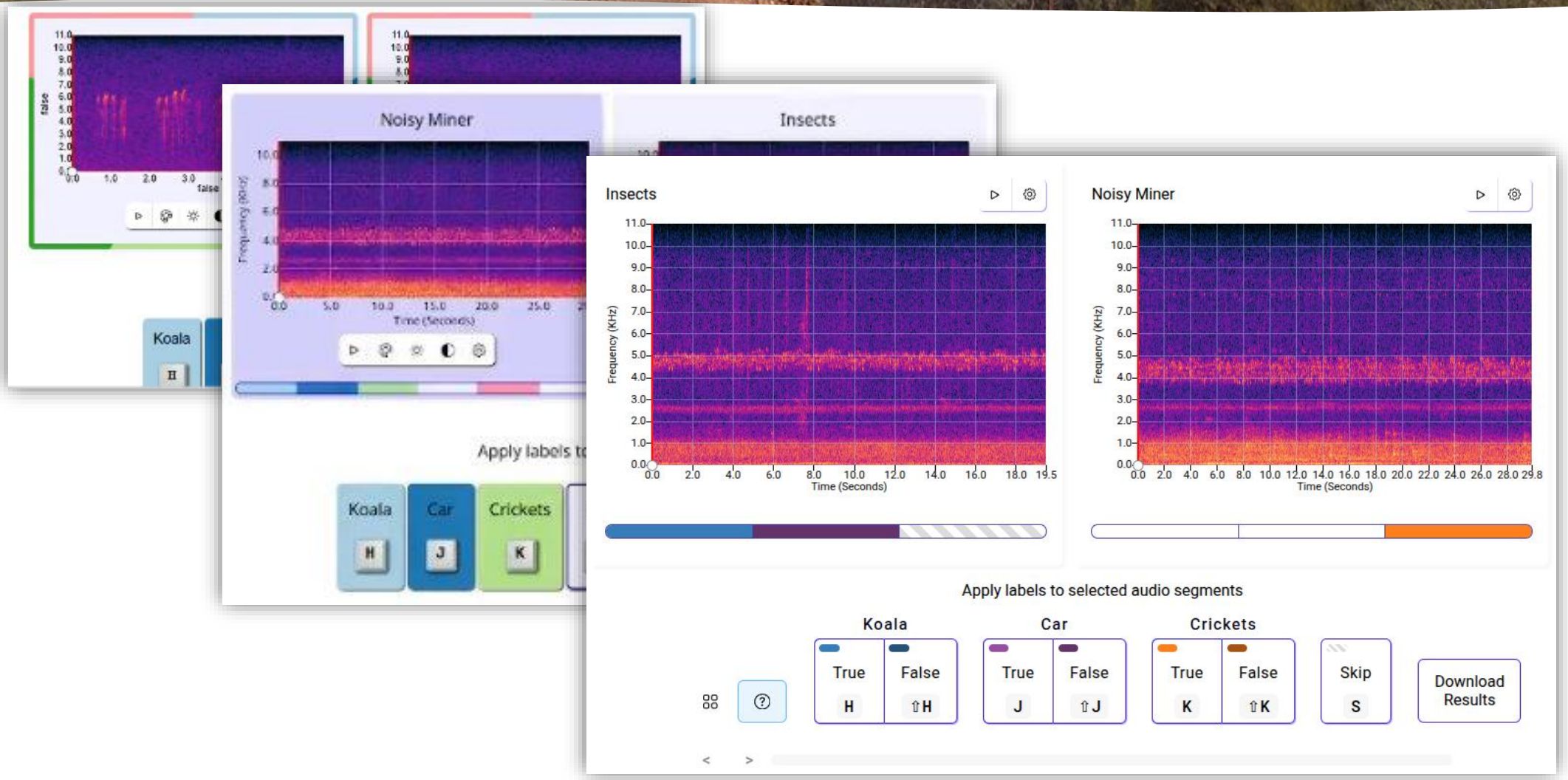


NO YES

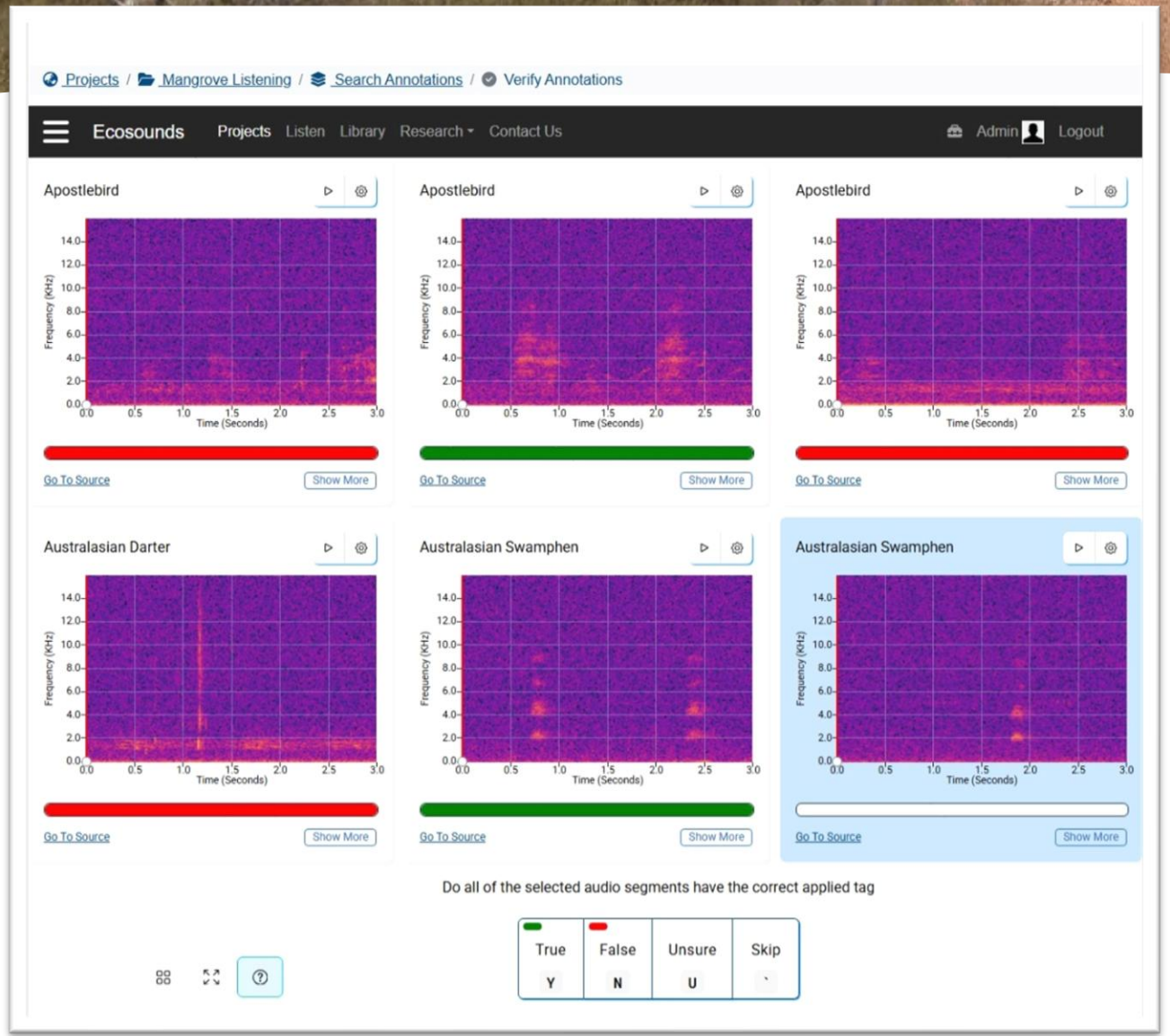
OR

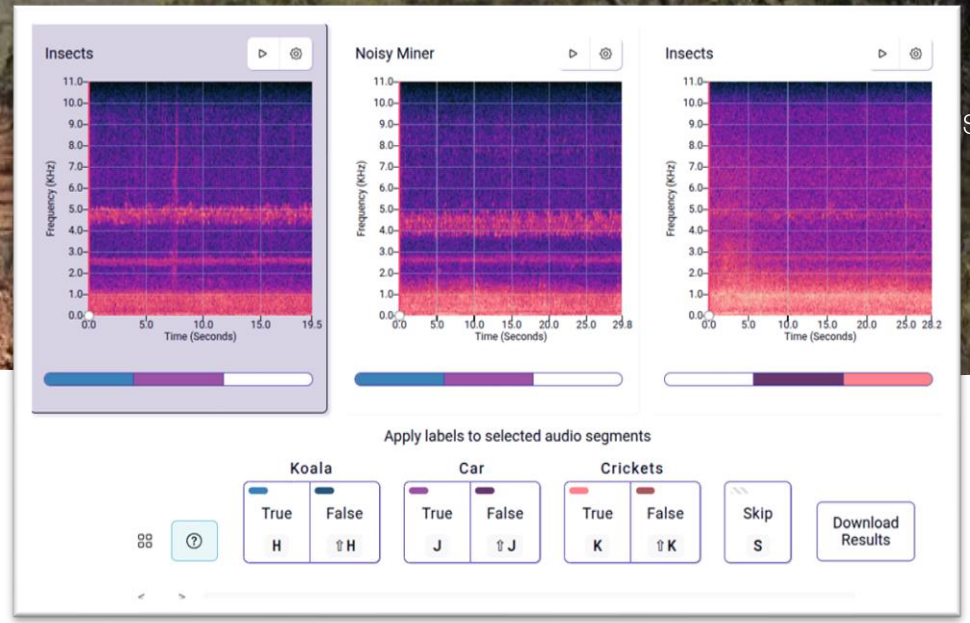


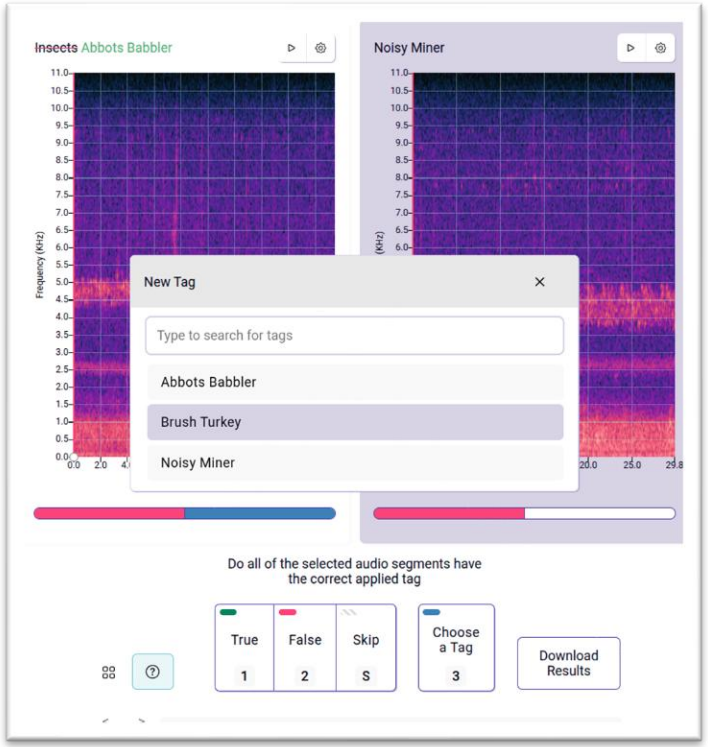
NO YES

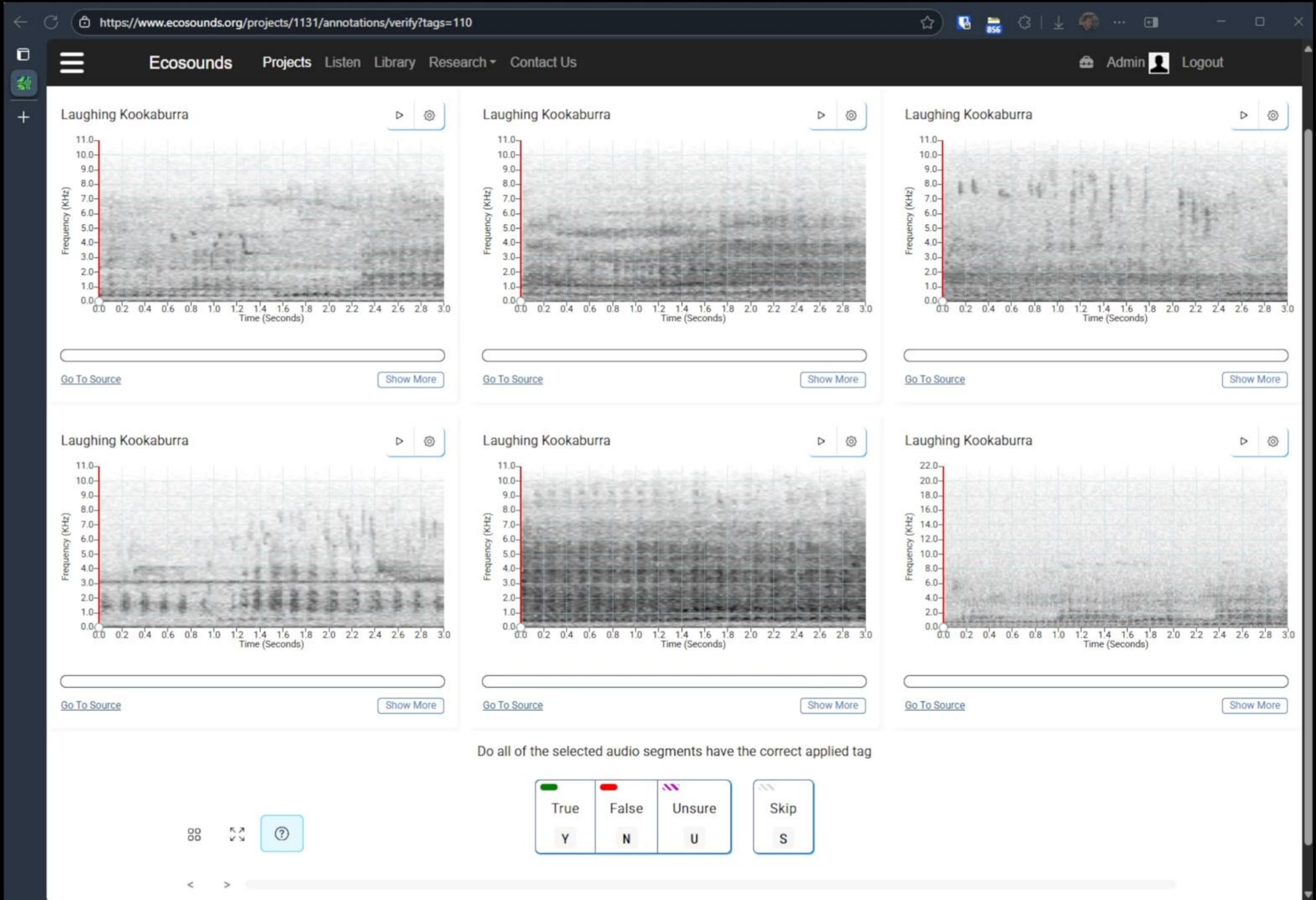


Verification: result








https://www.ecosounds.org/projects/1131/annotations/verify?tags=110
Ecosounds
Projects
Listen
Library
Research
Contact Us
Admin
Logout
Laughing Kookaburra
Frequency (KHz)
Time (Seconds)
Go To Source
Show More
Laughing Kookaburra
Go To Source
Show More
Laughing Kookaburra
Go To Source
Show More
Laughing Kookaburra
Go To Source
Show More
Laughing Kookaburra
Go To Source
Show More
Laughing Kookaburra
Go To Source
Show More
Do all of the selected audio segments have the correct applied tag
True
Y
False
N
Unsure
U
Skip
S

Verification: objectives

Objectives achieved

- ✅ Web
- ✅ Mobile
- ✅ Desktop
- ✅ Open source
  - ecoacoustics/web-components
- ✅ Keyboard shortcuts
- ✅ Made for verification
- ✅ Load any segment from any file
- ✅ Multi item (grid view)
- ✅ Single item view
- ✅ Collaboration
  - Simply Share a URL

Bonus features

- We record *incorrect*, *correct*, *unsure*, and *skip* responses
  - → rich data
- Support overrun to form consensus
- Customisable tasks
  - Verify only
  - Verify and correct label
  - Classification tasks
- Reusable
- Has a basic standalone mode
- Themeable

Deployed in July 2025

- ≈12 months of work with 1.2 FTE
- 42,000 verifications already
- 55 codesign features complete, 67 more…
- Already used for agile learning for AI

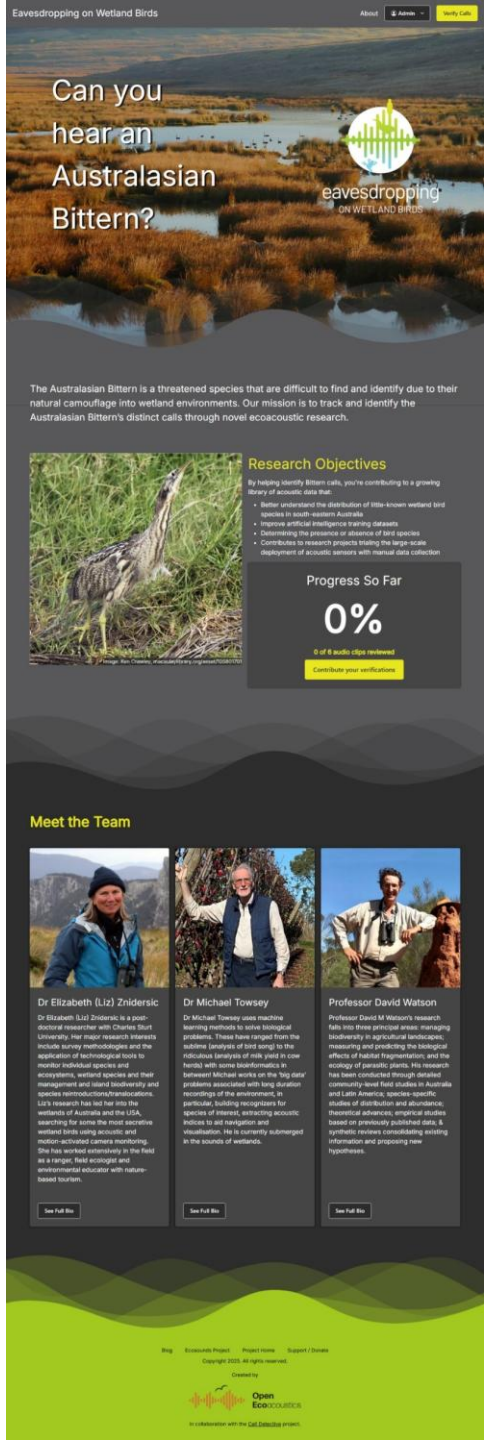
Future work

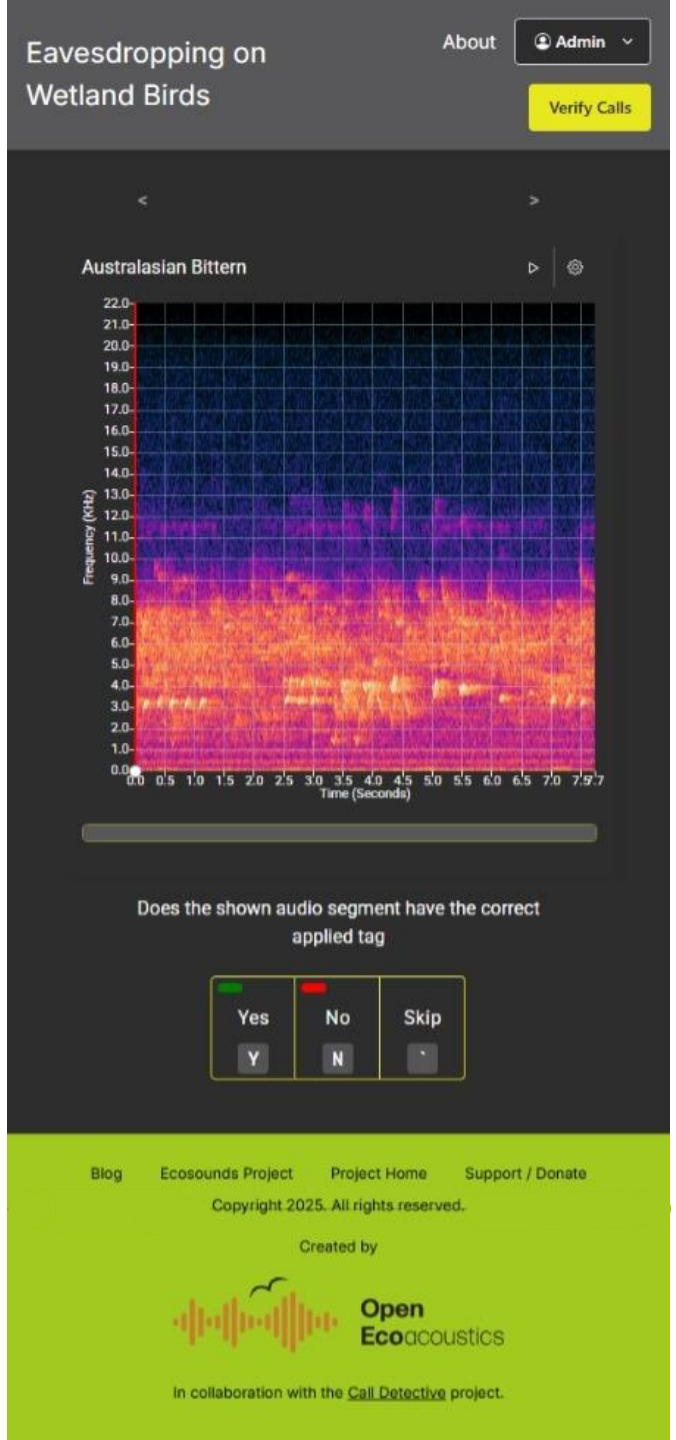
- Sampling strategies
  - e.g. find one correct label per day then don't show me anything else from the current day
- Showing history / resuming sessions
- Enhance export of verification data

Microsites & Call Detective

Dedicated websites for citizen engagement

- Simple
- Content heavy
- Focused
- Hides management platform (Ecosounds)
- Primarily targeting verification tasks
- One built, more to come
- <u>Reuses the verification component and customises it for audience</u>





Questions

Open Ecoacoustics

Special thanks to:

- Hudson Newey, primary developer for this project
- Our codesign users for feedback
- QUT eResearch

QUT

ARDC Australian Research Data Commons

Australian Government

NCRIS National Research Infrastructure for Australia

Open Ecoacoustics is a co-investment partnership with the Australian Research Data Commons (ARDC) through the Planet Research Data Commons (DOI: 10.3565/ts8c-ee10). The ARDC is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS).