

# Audubon Core: New terms for sound recordings

Dan Stowell (Queen Mary University of London)  
&  
Ed Baker (University of York / Natural History Museum)

Audubon Core Annual Meeting, 2020-09-21

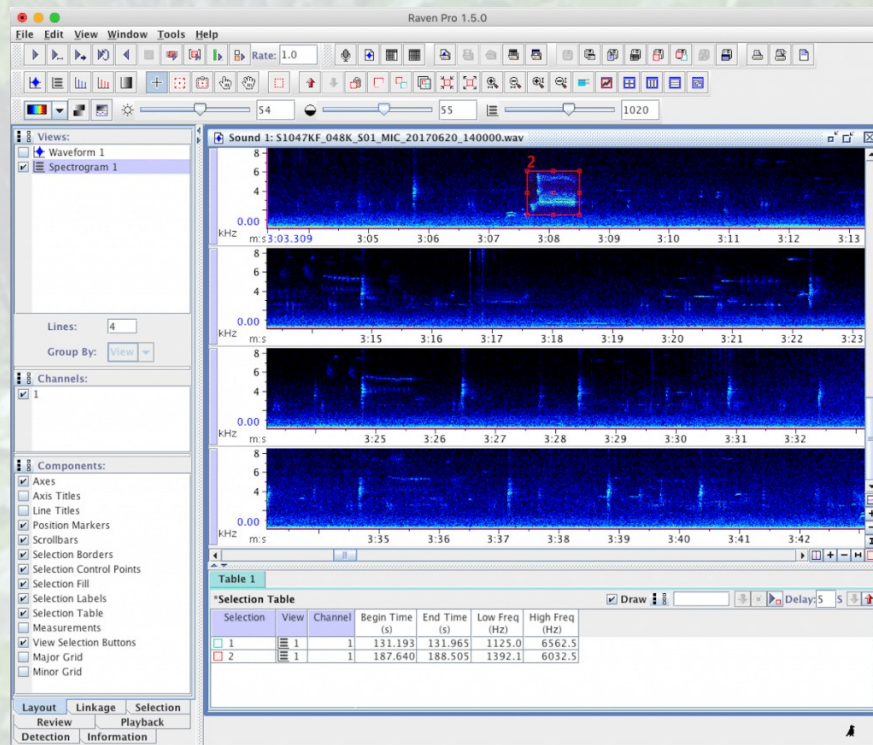
# Context

- Audubon Core
  - The **Audubon Core** is a set of vocabularies designed to represent metadata for biodiversity multimedia resources and collections.
- Starting to work on terms for **audio** collections



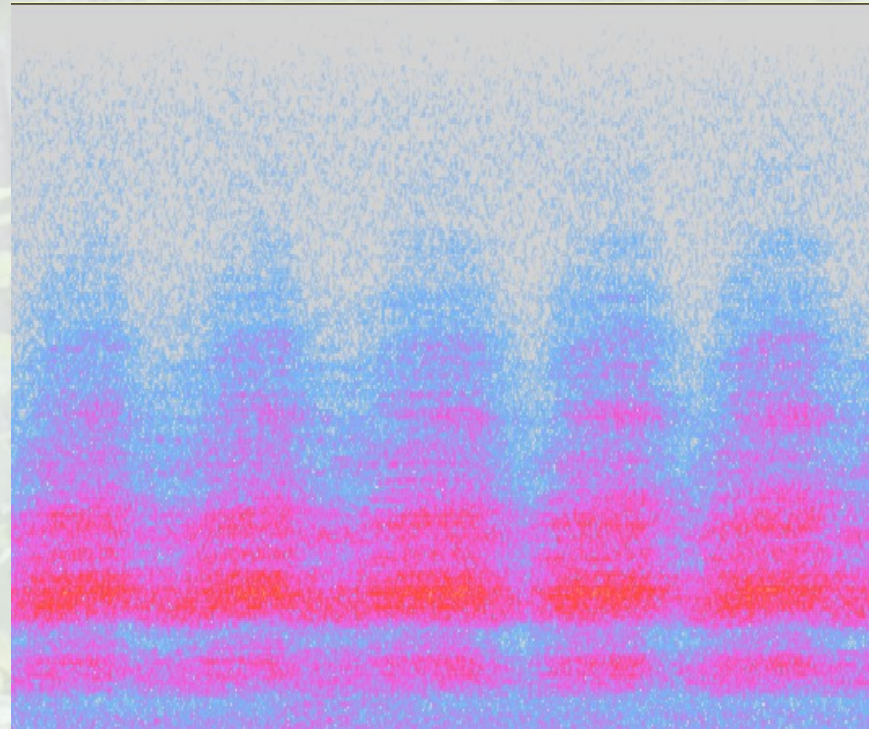
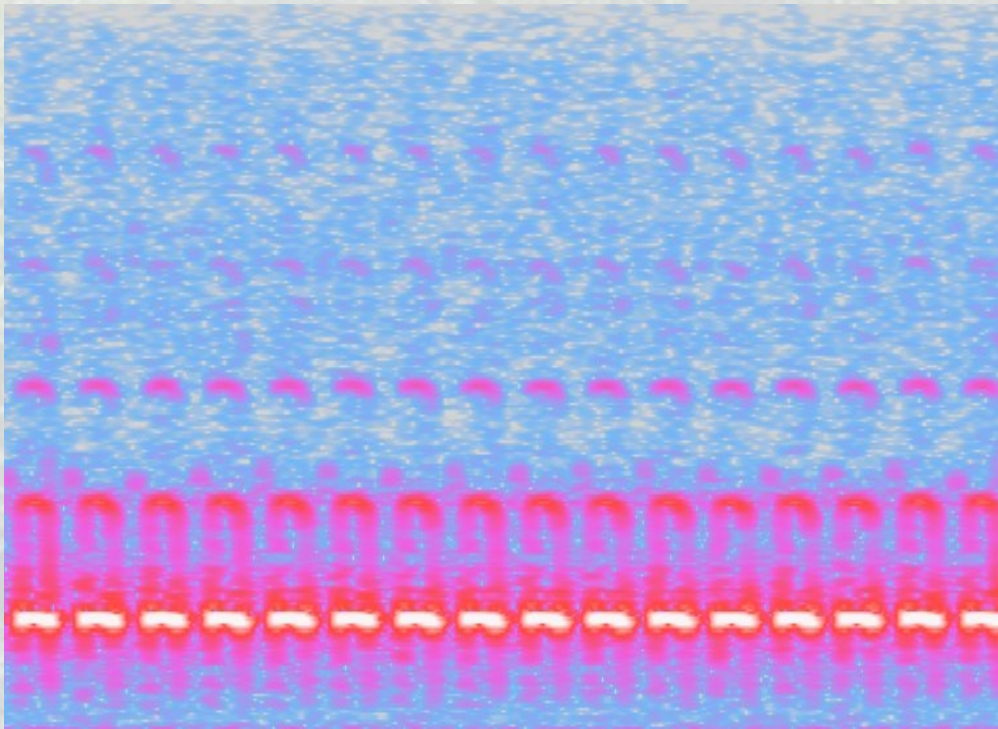
# Examples and prior work (1)

# Examples and prior work (2)





# Examples and prior work (3)





# We're presenting:

- 1) New terms proposed for AC
- 2) Annotation model (for discussion)

# Mapping of AC terms

Metadata from several sound collections compared to identify common use cases and missing terms in AC

- BioAcoustica
- Florida Museum of Natural History
- Fonoteca Neotropical Jacques Vieliard, UNICAMP
- Instituto Alexander von Humboldt Colección de Sonidos Ambientales
- Borror Laboratory of Bioacoustics
- Australian National Wildlife Collection, CSIRO
- Macaulay Library



# Example mapping (BioAcoustica)

Field	Description	AudubonCore Map
Node ID		dcterms:identifier
Title	The name used to identify the recording (typically includes original CD/tape number and sp	dcterms:title
Last revision date		dcterms:modified
Language		ac:metadataLanguageLiteral
Recording	The audio file	
Project		
Body	Comments on the recording	ac:comments
Original metadata image	A scan or photograph of the original metadata if it exists in printed or handwritten form	
Original trace images	Scan(s) or photograph(s) of paper oscillograms relating to the recording	
Original verbatim species	The species identification as recorded in the original metadata	
Original CD number	Used to associate digital record with physical collection	
Original CD track number	Used to associate digital record with physical collection	
Original tape number	Used to associate digital record with physical collection	
Copyright holder		xmpRights:owner
Licence		dcterms:rights
Species	Link to a taxon in the site's biological classification	dwc:scientificName
Requested additional species	Used to suggest a name that is not currently in the site's biological classification	
Specimen	Link the recording to a preserved or observed specimen	ac:associatedSpecimenReference or Observation
Location	Location the recording was made (may be different to where the specimen was collected)	Location content type maps directly to DarwinCore Lo



# Priorities for sound in AC

- Dealing with essential data for audio (e.g. sample rate)
- Identifying existing use cases not currently covered

# Sources consulted

- Heidelberg Bioacoustics Symposium 2019-12
  - Broad group covering many taxa (birds, insects, bats, cetaceans, terrestrial mammals)
- Baker & Chesmore (2020)
  - “Standardisation of bioacoustic terminology for insects”
- Similar data standards from other domains
  - [w3c Annotation Model](#), [IIIF Presentation API](#), [dwc:event](#), [Music Ontology](#), and more...



# Term additions: (a) dwc:individualCount

“dwc” = Darwin Core <https://dwc.tdwg.org/>


Number of individuals represented

Why in AC?

- Users of audio often want single non-overlapping sounds
- Other users want group behaviour e.g. chorusing, duetting

# Term additions: (a) dwc:individualCount

<https://github.com/tdwg/ac/issues/176>



**danstowell** commented 23 days ago • edited by baskaufs ▾

Member 😊 ...

Term Name: dwc:individualCount

Imported from: <http://rs.tdwg.org/dwc/terms/individualCount>

Type: rdf:Property

Label: Individual Count

Required: No


Repeatable: No

Definition: The number of individuals represented present at the time of the Occurrence.

Usage: Numeric number of individuals

Justification for the term addition: The behaviour of individuals may be affected by the number of individuals present (e.g. chorusing). For analysis of bioacoustic traits, recordings of a single individual are often required, rather than a group with overlapping songs.

Proposed by myself and **@edwbaker**

 1



# Term additions: (b) mo:sample\_rate

“mo” = Music Ontology <http://musicontology.com/>

- Started in 2007; used in music audio informatics


Associates a digital signal to its sample rate.  
Numeric value in hertz (Hz)

Why in AC? Can be used in queries, since:

- acts as a hard limit on represented sound frequency ranges
- often an indicator of the acceptability of a sound file for a particular purpose

# Term additions: (b) mo:sample\_rate

<https://github.com/tdwg/ac/issues/177>

 danstowell commented 23 days ago • edited by baskaufs

Member 😊 ⋮

Term Name: mo:sample\_rate ([http://purl.org/ontology/mo/sample\\_rate](http://purl.org/ontology/mo/sample_rate))

Imported from: [http://musicontology.com/specification/#term-sample\\_rate](http://musicontology.com/specification/#term-sample_rate)

Type: rdf:Property

Label: Sample Rate

Required: No

Repeatable: No

Definition: Associates a digital signal to its sample rate.

Usage: Numeric value in hertz (Hz)

Notes: For example, a Service Access Point may have a specific resolution, quality, or format. "Sample rate" is distinct from the related concept of "bit rate" for compressed files such as MP3, and is applicable to both uncompressed and compressed files.

Justification for the term addition: Audio sampling rate is an extremely common attribute of digitised sound recordings, and is important in search queries because it acts as a hard limit on the ability to represent sounds that occur in certain frequency ranges. It is often an indicator of the acceptability of a sound file for a particular purpose.

(Currently, "audio sampling rate" is mentioned in the unconstrained ac:resourceCreationTechnique, though this cannot support search queries very well. It is suggested that this use case is deprecated. See [#179](#).)

Proposed by myself and [@edwbaker](#)



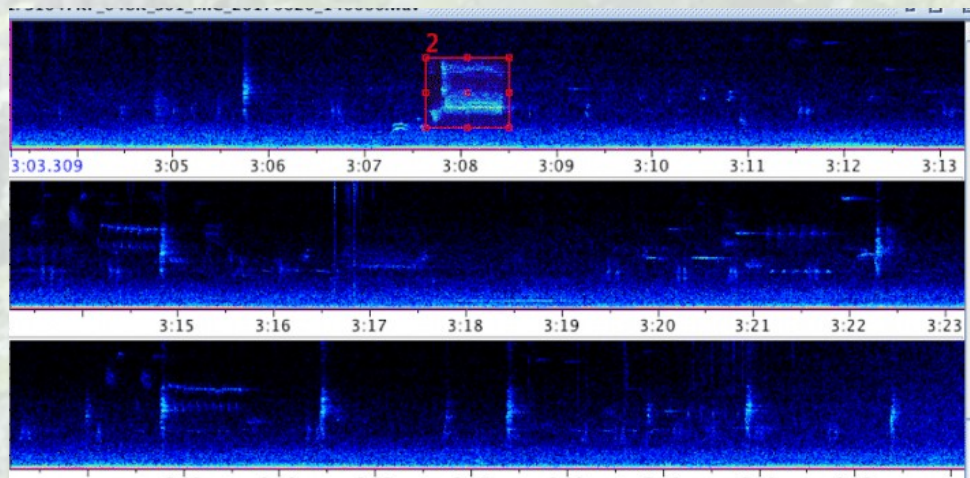
# Term additions: (c) freqLow & freqHigh

New terms.

The lowest/highest frequency of the phenomena reflected in the multimedia item.  
Numeric value in hertz (Hz)


Why in AC?

- Commonly-annotated attribute
- Can be used to validate taxon identity
- These terms may be independently present/absent



# Term additions: (c) freqLow & freqHigh

<https://github.com/tdwg/ac/issues/178>



**danstowell** commented 23 days ago • edited by baskaufs

Member 😊 ...

Term Name: ac:freqLow  
Term Name: ac:freqHigh

Type: rdf:Property

Label: Lower frequency bound  
Label: Upper frequency bound

Required: No

Repeatable: No

Definition: The lowest frequency of the phenomena reflected in the multimedia item.  
Definition: The highest frequency of the phenomena reflected in the multimedia item.

Usage: Numeric value in hertz (Hz)

Notes: These terms refer to the sound events depicted and not to the constraints of the recording medium, so are in principle independent from sampleRate. If dwc:scientificName is specified, these frequency bounds refer to the sounds of the species given in the dwc:scientificName throughout the whole recording. Although many users will specify both freqLow and freqHigh, it is permitted to specify just one or the other, for example if only one of the bounds is discernible.

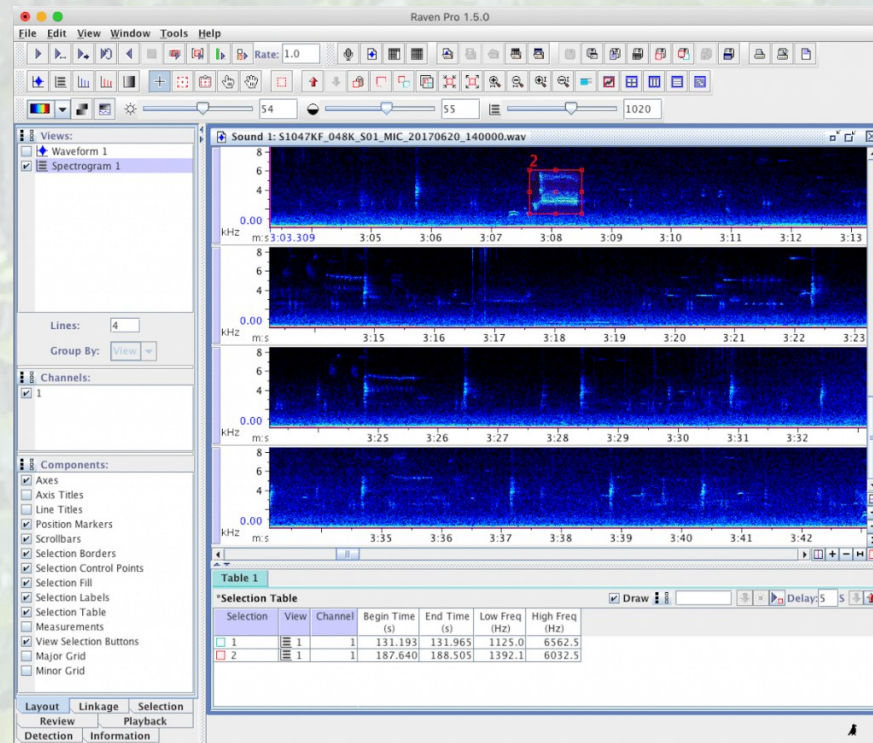
Justification for the term addition: "Time-frequency boxes" are a common approach for labelling events within bioacoustic sound recordings. Audio recordings can be used to explore and validate taxon identity for insect sounds. For many insect sounds (e.g. extended calling or chorusing), frequency bounds can be much clearer than time bounds. Thus, in a soundscape recording, it can be desirable to associate a taxon label with some sub-region of the audio, but as frequency bounds without time bounds. In search queries, it can often be desirable to target (or exclude) animal sounds that occur within specific frequency ranges.



# Annotation model (for discussion)

How to represent metadata  
for each  
“sound event” (e.g. call)  
within an audio item?

*(See also: video,  
machine observation...)*



# Example of audio annotations data

time_start	time_end	freq_low	freq_high	taxon	type
	12	2000	5000	Gryllotalpa gryllotalpa	Calling song
716.188	746			Gryllotalpa gryllotalpa	Call
0	24.157			Gryllotalpa gryllotalpa	Call
0	48.025			Gryllotalpa gryllotalpa	Call
0.602893	0.726025			Gryllotalpa gryllotalpa	Call
5	6			Gryllotalpa gryllotalpa	Call
4	34			Gryllotalpa vineae	Call
0	3				Voice Introduction



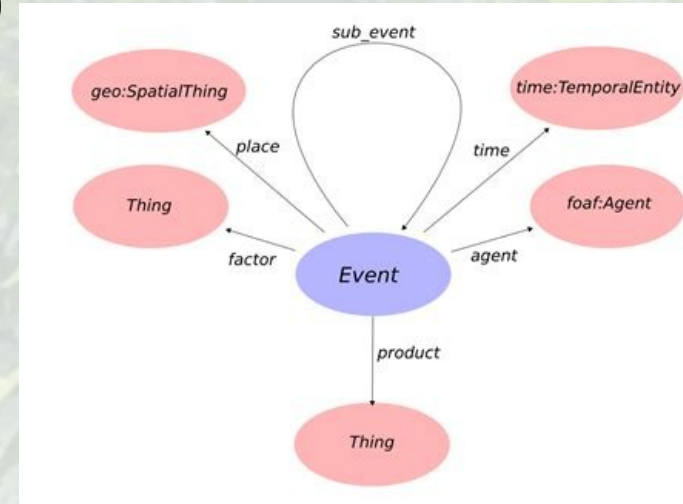
# Possible models... (Stowell et al 2016)

Output paradigms for event detection of birds

Output format		Common algorithms	Used by applications	Advantages	Disadvantages	Cmplxty
(a) Presence/ absence		Classifiers	Occupancy-models in statistical ecology; retrieval / data mining systems generally	Evaluation is straightforward; manual annotation can be efficient	Low temporal precision; multiple events merged	1
(b) Onsets		Onset detectors e.g. energy slope, per-frame classifier		Overlapping events are OK	No offset/duration information	1
(c) Monophonic segmentation / VAD		Energy thresholding; VAD HMM decoding			Overlapping events merged	2
(d) Polyphonic segmentation (multi-monophonic)		NMF		Joint estimation can reduce confusion between similar sound types; overlaps <i>between</i> species are OK	Overlaps in <i>same</i> species merged	3
(e) Polyphonic segmentation (overlappable)				Overlapping events are OK		3
(f) Time-frequency boxes		Spectrogram correlation	cross- Common where spectrogram cross-correlation used e.g. in Raven			4

# Recommended approach

- Music Ontology (again) well-developed
  - Event model uses Timeline Ontology (“tl”)
    - tl:start
    - tl:end
    - tl:duration
- Plus, adopt a set of item-level terms to be used for individual events (taxon, freqLow, freqHigh, ...)





# Summary

- Term additions
  - a) dwc:individualCount <https://github.com/tdwg/ac/issues/176>
  - b) mo:sample\_rate <https://github.com/tdwg/ac/issues/177>
  - c) ac:freqLow & ac:freqHigh <https://github.com/tdwg/ac/issues/178>
- Annotation model for events
  - Wider discussion needed (video, machine observations)
  - Please help to join up the discussion!