Social media as a digital data source for wildlife ecology studies

Anna Willoughby
3rd Annual Digital Data Conference
New Haven, CT
June 11, 2019
Making data and images of millions of biological specimens available on the web

119,163,881
Specimen Records
30,380,997
Media Records
1,614
Recordsets

Search the Portal

Why digitize?
Why digitization matters
More about what we do and why

Digitization
Learn, share and develop best practices

Sharing Collections
Documentation on data ingestion

Working Groups
Join in, contribute, be part of the community

Proposals
New tool and workshop ideas

Citizen Scientists
How can you help biological collections?

Researchers
Learn about research directions

Collections Staff
Learn how your collection can benefit from our work

Teachers & Students
Download lesson plans about using digitized specimens
Bias in species records

Meyer et al. 2016, Global Ecology and Biogeography
The new era of social media

**DIGITAL AROUND THE WORLD IN 2019**
THE ESSENTIAL HEADLINE DATA YOU NEED TO UNDERSTAND GLOBAL MOBILE, INTERNET, AND SOCIAL MEDIA USE

- **TOTAL POPULATION**: 7.676 billion
  - Urbanisation: 56%
- **UNIQUE MOBILE USERS**: 5.112 billion
  - Penetration: 67%
- **INTERNET USERS**: 4.388 billion
  - Penetration: 57%
- **ACTIVE SOCIAL MEDIA USERS**: 3.484 billion
  - Penetration: 45%
- **MOBILE SOCIAL MEDIA USERS**: 3.256 billion
  - Penetration: 42%

## Top Social Media Sites

<table>
<thead>
<tr>
<th>Monthly Users</th>
<th>Data Type</th>
<th>API</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 2 billion</td>
<td>Video, Photos, Audio, Text</td>
<td>✔</td>
<td>350 million photos uploaded per day</td>
</tr>
<tr>
<td>1.9 billion</td>
<td>Video, Photos, Audio, Text</td>
<td>✔</td>
<td>&gt; 100 billion videos</td>
</tr>
<tr>
<td>1 billion</td>
<td>Photos, Video, Audio, Text</td>
<td>✔</td>
<td>&gt; 50 billion photos</td>
</tr>
</tbody>
</table>
Discussion Objectives:

• Horizon scan of social media in wildlife ecology studies

• Discuss opportunities and obstacles of social media data

• Brainstorm possible avenues for data curation into established pipelines of natural history data
Biodiversity patterns in pet amphibians

Measey et al. 2019, *Frontiers in Ecology and Evolution*

**FIGURE 1** Taxonomic patterns in amphibian (A) orders, (B) superfamilies, and (C) families present in the species identified in captive YouTube amphibian videos. The median (green line) and 95% confidence intervals (brown lines above and below), adjusted for multiple comparisons, were estimated from the hypergeometric distribution. The points that fall between the 95% confidence intervals are not significantly over or under-represented, relative to the number of amphibian species worldwide. Those labeled orders, superfamilies or families that fall above the 95% confidence intervals are over-represented and those below are under-represented in our sample of YouTube videos.
Coyote-domestic dog behaviors

Table 2. Numbers (and percentages) of coyote–dog (*Canis latrans, C. lupus familiaris*) dyadic interactions per ethogram category (see Table 1) for a total of 49 coyote–dog dyads involving 39 coyotes and 45 dogs (30 large, 7 medium, 8 small). Dyadic interactions were observed in 35 video clips containing 1–3 dyads per video. Number of videos here sums to 36, because 1 video contained a predatory and an agonistic dyad.

<table>
<thead>
<tr>
<th>Ethogram category</th>
<th># Dyads (%)</th>
<th># Coyotes</th>
<th># Dogs</th>
<th># Videos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Large</td>
<td>Medium</td>
</tr>
<tr>
<td>Predatory</td>
<td>5 (10.20)</td>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Agonistic</td>
<td>13 (26.53)</td>
<td>10</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>Social play</td>
<td>13 (26.53)</td>
<td>9</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Other/undetermined</td>
<td>18 (36.73)</td>
<td>15</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>

Boydston et al. 2018
Human-wildlife sentiments

• These pics were taken outside the Trafalgar Sports Bar this Thursday at 4 pm. They were mingling with tourists and passersby, eyeing everyone for possible goodies (even one on lookout duties up the lamppost!). (Anon, January 2014)

• Mount Alvernia is under siege by monkeys, but you probably already know that. (Anon, March 2014)

• Despicable vandals! (Anon, January 2014)

Radford et al. 2018, *Folia Primatologica*
Ringtails
Human-ringtails interactions
Expansion of species occurrence data

- 1398 videos
- 207 videos of wild ringtails
- 65 unique GPS locations
- Complements in
  - data deficient areas
  - edges
  - cities
Expansion of species occurrence data

- 1398 videos
- 207 videos of wild ringtails
- 65 unique GPS locations
- Complements in
  - data deficient areas
  - edges
  - cities
Ecological and behavioral insights

• Activity times
• Vocalizations
• Interspecies interactions:
  – prey, predators
• and so much more!
How to scale with computational methods?

**Table 1** Commonly used tools for computer vision application to ecology

<table>
<thead>
<tr>
<th>Name</th>
<th>Reference</th>
<th>Task</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenCV</td>
<td>Bradski (2000)</td>
<td>Description, Counting, Identity</td>
<td>Source library for computer vision algorithms in python/java/C++</td>
</tr>
<tr>
<td>ImageJ</td>
<td>Abràmoff et al. (2004)</td>
<td>Description, Counting</td>
<td>Segmentation and thresholding</td>
</tr>
<tr>
<td>BISQUE</td>
<td>Kvilekval et al. (2009)</td>
<td>Description, Counting</td>
<td>Also serves as a hosting platform for image analysis tools</td>
</tr>
<tr>
<td>Agisoft Photoscan</td>
<td>-</td>
<td>Description</td>
<td>Commercial software for 3D model reconstruction from images</td>
</tr>
<tr>
<td>StereoMorph</td>
<td>Olsen and Westneat (2015)</td>
<td>Description</td>
<td>R package for 3D reconstruction and image calibration</td>
</tr>
<tr>
<td>NaturePatternMatch</td>
<td>Stoddard et al. (2014)</td>
<td>Description</td>
<td>Comparing features among ecological images</td>
</tr>
<tr>
<td>Google Cloud API</td>
<td>-</td>
<td>Identity</td>
<td>Classification of image content using Cloud Vision API, deep learning source library using TensorFlow</td>
</tr>
<tr>
<td>Merlin</td>
<td>Van Horn et al. (2015)</td>
<td>Identity</td>
<td>Bird identification app for iPhone and Android</td>
</tr>
<tr>
<td>Wildbook</td>
<td>Crall et al. (2013)</td>
<td>Identity</td>
<td>Individual identification and data management tools</td>
</tr>
</tbody>
</table>

Discussion Questions

• How to account for bias in researcher effort and social-media use?

• How to overcome the bottleneck of image processing?

• How to better capture collector and animal behavior?