

**Date:** October 15, 2022

**Title of Article:** A pipeline for assessing the quality of images and metadata from crowd-sourced databases.

**Manuscript:** <https://www.biorxiv.org/content/10.1101/2022.04.29.490112v1>

External Files:

**Name of the Corresponding Author:** Jackie Billotte, M.S.

**Email Address of the Corresponding Author:** [jackie.billotte@colostate.edu](mailto:jackie.billotte@colostate.edu)

Dear Reviewers,

Thank you for your careful reading of the manuscript, positive comments, and constructive critiques. I feel that the revisions and responses to the reviewers' comments substantially improve the quality of the manuscript and address some key issues.

*I have addressed your individual comments below. Major changes to the manuscript include the following:*

1. Figure 2 was modified to include a screen shot of Section 2 of the iNaturalist download function to help clarify the selections for downloading observations that were used to demonstrate the use of the pipeline.
2. An additional sentence was added to help clarify the difference between observations and records.

Again, thank you for providing your insight and recommendations and your consideration of this manuscript.

Sincerely,

Jackie Billotte, M.S.

Ph.D. student, Colorado State University

[jackie.billotte@colostate.edu](mailto:jackie.billotte@colostate.edu)

(720) 590-2373

**Detailed response to reviewers:**

1. L 9/10: avoid repetitions
  - Thank you for the suggestion. Edits have been made to avoid repetition throughout the manuscript.
2. L42: delete comma after "image-based" and L234: delete comma after "quantifiable."
  - The comma has been deleted and the rest of the document's punctuation has also been reviewed.
3. L63/64: switch "the of" to "of the"
  - Thank you for alerting me to the typo. This has been corrected.
4. Figure 2: The labeling seems incomplete. In the legend, you refer to Section 1, 2, and 3. Which sections are these?

- Labels were added to the screen grabs of iNaturalist's download function and a screengrab of Section 2 was added to help clarify the
5. Figure 3: In the legend, last sentence, it should be "observations" (plural).
    - This has been corrected.
  6. Results: please check the numbers one more time, or at least clarify. On L171, you state that you found 156,842 downloadable observations and on L176, you say that 49.91% were identified to at least family level. But on L181, you state that 156,842 out of 158,129 downloadable observations had a family-level identification. On L185, you refer to 425,950 "records". What is a record in this context, i.e. how does it differ from an observation?
    - I agree this needs clarification. Observations may have more than one image attached to them, each with separate metadata. "Records" refers to the metadata associated with each image. This information has been added to the manuscript.
  7. L223-225: use "research grade" throughout.
    - Instances of "research quality" being used instead of "research grade" have been changed to "research grade" through out the manuscript.
  8. Zizka et al. 2019 is not in the reference list.
    - This reference has been added to the reference list. Additionally, the in-text citations were checked against