Wildlife Insights Data Use & Citation Guide

Wildlife Insights provides access to one of the largest and most diverse camera trap datasets in the world. Before using data from Wildlife Insights in an analysis or for any other purpose, please read through this Data Use and Citation Guide to learn more about the data.

Anyone who downloads data from Wildlife Insights must first agree to the Terms of Use, which require a data user to provide attribution to the dataset creator/creators as required by the dataset license. Continue reading to learn more about the licenses available for Wildlife Insights projects and how to provide attribution.

NOTE: Many projects in Wildlife Insights are ongoing and are continually updated. If you download the same project(s) at different points in time, the resulting dataset may differ. If you plan to publish a scientific manuscript using data from Wildlife Insights, we recommend publishing the dataset used in an open-access public data repository.

Table of Contents

Data available for download

Every download from Wildlife Insights will include:

- **Projects.csv:** metadata about project methodology and objectives, including the type of project (sequence or image) and whether count was recorded in the project;
- **Cameras.csv**: metadata about the devices (cameras) used in the project;
- **Deployments.csv:** metadata about the placement of a camera, including start date, end date, coordinates and other camera settings;
- **Either an Images.csv or Sequences.csv:** Data about the animals detected by the camera traps are reported in one of two ways depending on how the data was recorded (denoted

by *project_type* in the projects.csv). The download package may include both the images.csv and sequences.csv if the request includes both image and sequence projects:
    - The **images.csv** contains data about each individual image, including species identifications and timestamp.
    - The **sequences.csv** contains data about each sequence of images. A sequence is a group of photos that are taken within some time frame (one minute) of each other. All photos in a sequence are tagged as a single detection of an animal or animal group. Data about a sequence of images including species identifications and date/time. The sequences.csv does not include records of every image nor does it include links to images within the sequence.
- **Data use & citation guide**;
- **Data Dictionary**: definitions for each field in the files provided; and
- **Wildlife Insights Terms of Use**.

Note: Wildlife Insights places certain restrictions on data to protect sensitive or private information, which includes locations of sensitive species, images of humans, and embargoed projects. Read more about how sensitive species data are protected by Wildlife Insights.

Verifying data

All users contributing data to Wildlife Insights retain ownership of their data and are responsible for reviewing and editing the identifications made by the computer vision model in their project(s). While Wildlife Insights provides tools for users to process and edit data as needed, Wildlife Insights cannot guarantee the accuracy of identifications available to the public.

How to verify whether a record has been reviewed

Every download package from Wildlife Insights includes a record of the most recent identification associated with an image. The record will detail the identification (class, order, family, genus and species), individual animal details (age, sex), and the name of the most recent identifier, along with other metadata.

In many cases, the most recent identifier recorded will be the name of a user associated with that project. This means that the image has been reviewed and/or edited by that user. This information can be found in the *images.csv* file in the column titled *identified_by.* In the example below, the reviewer's name is *Nicole Flores*.

In other cases, the most recent identifier in the *identified_by* column is *Computer Vision*. This means that the image has been uploaded, has passed through the computer vision model, but has not been reviewed by a user. In the example to the right, you can see the Computer Vision model has predicted there is one image of a Margay and one

F	M
identified_by	common_name
Nicole Flores	blank
Nicole Flores	Tiger
Computer Vision	Margay
Computer Vision	Human
Nicole Flores	Dark-winged Trumpeter
Nicole Flores	Dark-winged Trumpeter
Nicole Flores	Human
Nicole Flores	blank

image of a Human. These identifications have **not** been confirmed by a user and may be incorrect.

You can assess the accuracy of the computer vision results by referencing the *images.csv* file and the column titled *cv_confidence.* Read more about interpreting computer vision results on our About AI page.

Downloading Images

If you have been granted access to a private Project as an Owner, Editor, Contributor, Tagger or Viewer, individual images can be easily downloaded from the Wildlife Insights website by clicking on the image in the Identify or Catalogued tab and then clicking on the *Download* button.

If you wish to access images from a public download or are looking to bulk download images from a private Project, you can access the images in a Google Cloud Platform bucket (i.e, folder) by following the instructions below.

Note that the instructions to access images from sequence-based projects and image-based projects differ. Read more below for details.

Using a web browser

You can use your web browser to navigate within the Project bucket, filter and download files.

1. **Login to Google Cloud Platform** (GCP) with the same email you used to register for a Wildlife Insights account. This email must be associated with a Google Cloud Platform account in order to access your project's bucket. If you'd like to associate another email (i.e., a gmail account) with your Wildlife Insights project, please contact info@wildlifeinsights.org with your request.

2. **Once you are logged in to GCP, you can browse the project's bucket** through the project's bucket url: https://console.cloud.google.com/storage/browser/{ PROJECT } where { PROJECT } is the bucket name of the project. You can find the bucket name by following the instructions below:
    a. **For image-based projects**:

        You can find this name in the images.csv file under the location column. For example, given the following entry:

        gs://camera_trap_project__main/deployment/2019497/37f3aa87-a36e-4ca4-8fe6-ceca57977bf6.JPG

the name of the bucket is the string contained between 'gs://' and '/deployment/'. In this example, the project name is 'camera_trap_project__main'. So, for this example, the full bucket url is:

https://console.cloud.google.com/storage/browser/camera_trap_project__main

Using gsutil software

gsutil is a Python application that lets you access Cloud Storage from the command line and is the best option for batch downloads. gsutil is easy to install - follow the Google Cloud Platform instructions: https://cloud.google.com/storage/docs/gsutil_install.

Once you've installed gsutil, you can download images using the command line. Below are examples. Replace the values between the brackets with your project name (see details in the section above to find your project name):

To download all images of the project to the current directory

```
gsutil -m cp -r  gs://{ PROJECT }
```

To download all images of one specific deployment to the current directory

```
gsutil -m cp -r  gs://{ PROJECT }/deployment/{ DEPLOYMENT_ID }
```

To download a single image, find the full url in the images.csv file under the *location* column

```
gsutil -m cp -r  gs://{ PROJECT }/deployment/{ DEPLOYMENT_ID }/{ IMAGE }
```

Licenses

Creative Commons provides standardized licenses and guidelines that make it easy to share work and for users to provide attribution to the work. Each project in Wildlife Insights can be licensed under the following options (please read the Terms of Use for full details).

- Images (recorded data) can be licensed under **CC0, CC BY** or **CC BY-NC**.
- Metadata can be licensed under **CC0** or **CC BY**.

These licenses are described below:

- **Creative Commons Zero (CC0)** permits a user to share, adapt and modify the work, even for commercial purposes, without providing attribution (summary, full legal text);
- **Creative Commons Attribution 4.0 (CC BY),** permits a user to share and adapt material with appropriate attribution, including for commercial purposes (summary, full legal text);

- **Creative Commons Attribution-NonCommercial 4.0 (CC BY-NC)**, permits a user to share and adapt material with appropriate attribution, only for noncommercial purposes (summary, full legal text).

Citing data downloaded from Wildlife Insights

All projects licensed under **CC BY** or **CC BY-NC** require a data user to provide attribution. Wildlife Insights makes it easy to provide attribution by providing a list of data citations for each project in a download request. Projects in Wildlife Insights are also assigned an Archival Resource Key (ARK), which is a persistent, permanent link to a project and dataset. Please refer to the projects.csv file to view the recommended citations and licenses for the projects in this download.

Wildlife Insights suggests citing projects using the following format:

References:

> Author(s) (Year accessed from Wildlife Insights). Project name. DOI. Accessed via Wildlife Insights on dd-mm-yyyy.
>
> *Example:* Ahumada J, Schipper J (2020). Cafe Fauna. https://n2t.net/ark:/12345/bcd987 accessed via Wildlife Insights on 03-10-2020.

In-text citations:

> *Example:* Data used in this study were accessed from Wildlife Insights on *date* (Ahumada & Schipper, 2020)
>
> For additional information on citations, Wildlife Insights recommends reading GBIF's citation guidelines.