#DATA4WILDLIFE CHALLENGE 2022
Precedent Document
Challenge 1 (Data for Deterrence)
Create a benchmark dataset of online wildlife crime images, text, and emojis on social media and e-commerce platforms

WANTS & NEEDS. We're looking to create a benchmark dataset that can be searched and analyzed according to different types of wildlife trafficking-related information. A benchmark dataset is a public dataset which is designed and collected for studying real-world data science/research problems. The benchmark dataset should be social media platform agnostic, as wildlife is trafficking across multiple platforms such as Instagram and YouTube.

TIPS. Although there is no restriction regarding social media platforms used to collect data, we strongly encourage participants to focus on Instagram as the default social media platform since it contains all data types (e.g., photos, video, emoji, text) and is feasible to collect from.

- Emojis are sometimes used in combination with other emojis or with text; emojis and combinations of emojis can change over time.
- Text referencing trade ('how much,' 'how many available?' 'DM for price') may appear in the comments section.
- Text referencing specimen origin may appear ('wild,' 'tamed,' or 'untamed,' location of harvest).
- Text can be embedded within images, including hashtags and profile names.
- Images may or may not reflect the actual specimen for sale; stock images may be used.
- Postings should be related to online wildlife trafficking/illegal wildlife trade (e.g., advertisement, selling, shipping) NOT animal abuse or animal welfare

POTENTIAL PITFALLS. Sometimes videos are posted of animals as a form of clickbait. The videos are not necessarily about anything criminal or harmful. The comments section associated with the video is where the advertising, selling and buying information is located.

EXAMPLES.

- How Instagram Celebrities Promote Dubai's Underground Animal Trade, Bellingcat
- Global dataset for seized and non-intercepted illegal cheetah trade (Acinonyx jubatus) 2010–2019
- https://www.sapiens.org/culture/animal-trafficking-instagram/ Illegal online trade in endangered parrots: A groundbreaking investigation, The World Parrot Trust and World Animal Protection
- Online markets for African Grey Parrots, Global Initiative Against Transnational Organized Crime
- Inferring patterns of wildlife trade through monitoring social media: shifting dynamics of trade in wild-sourced African Grey parrots following major regulatory changes, Science Direct

DATA COLLECTION. The data collection guidance below will be based on Instagram. From each posting, extract images/videos by simply using an Instagram video downloader tool to download/extract the videos. For example:

- https://igram.io/
- https://bigbangram.com/
- https://toolzu.com/downloader/instagram/video/

In addition, extract posting URL; poster's username; number of likes; posting time; collecting time; text/content; evidence of trading including terms and relevant emojis; poster's profile URL; and insert the extracted information into the provided below CSV template. Each row is for a specific online wildlife trafficking posting. Create an image/video folder for each posting. You are welcome to use any APIs or external tools to collect the above information. But, make sure collected postings are related to online wildlife trafficking/illegal wildlife trade. High quality data is crucial for future analysis. CSV template: https://drive.google.com/file/d/1XJ2oVmIW0Jkr3ePcjQa_RwjkGbEh0a3j/view?usp=sharing. It contains information extracted from an example posting: https://www.instagram.com/p/Bt4CfrAlSIa/. Delete it when insert your own collected data.

OUTPUT & SUBMISSION. Submit images/videos and the csv file. Each row in the csv file represents each posting that you collected. Compress the images folders (again, each folder for each posting) and csv file (e.g., make a zip file by using 7-zip or Winzip), and then submit the compressed file to a pre-assigned Google Drive link/folder. In addition, submit statistics of the collected data. For example, how many postings did the participant/team collect?