Wildbook: Crowdsourcing, computer vision, and data science for conservation.

Tanya Y. Berger-Wolf
University of Illinois at Chicago
Chicago, IL
tanyabw@uic.edu

Daniel I. Rubenstein
Princeton University
Princeton, NJ
dir@princeton.edu

Charles V. Stewart
Rensselaer Polytechnic Inst.
Troy, NY
stewart@cs.rpi.edu

Jason A. Holmberg
WildMe.org
Portland, OR
jason@whaleshark.org

Jason Parham
Rensselaer Polytechnic Inst.
Troy, NY
parhaj@rpi.edu

Jonathan Crall
Rensselaer Polytechnic Inst.
Troy, NY
crallj@rpi.edu

Sreejith Menon
Bloomberg LP
New York, NY
smenon59@bloomberg.net

Jon Van Oast
WildMe.org
Portland, OR
sito.org@gmail.com

Emre Kiciman
Microsoft Research
Redmond, WA
emrek@microsoft.com

Lucas Joppa
Microsoft Research
Redmond, WA
lujoppa@microsoft.com

ABSTRACT

Photographs, taken by field scientists, tourists, automated cameras, and incidental photographers, are the most abundant source of data on wildlife today. Wildbook is an autonomous computational system that starts from massive collections of images and, by detecting various species of animals and identifying individuals, combined with sophisticated data management, turns them into high resolution information database, enabling scientific inquiry, conservation, and citizen science.

We have built Wildbooks for whales (flukebook.org), sharks (whaleshark.org), two species of zebras (Grevy's and plains), and several others. In January 2016, Wildbook enabled the first ever full species (the endangered Grevy's zebra) census using photographs taken by ordinary citizens in Kenya. The resulting numbers are now the official species census used by IUCN Red List: http://www.iucnredlist.org/details/7950/0. In 2016, Wildbook partnered up with WWF to build Wildbook for Sea Turtles, Internet of Turtles (IoT), as well as systems for seals and lynx. Most recently, we have demonstrated that we can now use publicly available social media images to count and track wild animals. In this paper we present and discuss both the impact and challenges that the use of crowdsourced images can have on wildlife conservation.

Keywords

Data Science, Computational Ecology, Biodiversity, Conservation, Citizen Science, Crowdsourcing, Computer Vision, Bias

1. INTRODUCTION

How many African elephants are left in the world and how fast are they being lost to poaching? How far do whales travel? How many turtle hatchlings survive? Answers to these basic questions are critical to saving these and other endangered species and to the conservation of the biodiversity of our planet. However, this basic data is barely available for just a handful of species. The official body that tracks the conservation status of planet's species, International Union for Conservation of Nature (IUCN) Red List of Threatened Species™ [34], currently has over 79,000 species [36]. Yet, the Living Planet report, the most comprehensive effort to track the population dynamics of species around the world, includes just 10,300 populations of just 3,000 species [12]. That's not even 4%! Scientists do not have the capacity to observe every species at the needed spatio-temporal scales and resolutions and there are not enough GPS collars and satellite tags to do so. Moreover, invasive tracking can be dangerous to the animals [48].

Images of animals and their environment, intentionally and opportunistically collected, are quickly becoming one of the richest, most abundant, highest coverage and widely available source of data. Coming from camera traps, cameras mounted on vehicles or UAVs (drones), photographs taken by tourists, citizen scientists, field assistants, scientists, and public photo streams, many thousands of images may be collected per day from just one location. Taking advantage of this rich but big and messy source of data is only possible if we leverage computational approaches for every stage of the process, including image collection, information extraction, data modeling, and query processing. We have developed algorithms and built a system, called **Wildbook**™[1], based on the state of the art machine learning and data management approaches. Wildbook is an autonomous computational system that starts with an arbitrary heterogeneous collection of photographs of animals (Figure 1). Wildbook can detect various species of animals in those images (using DCNN) and identify individual animals of most striped, spotted, wrinkled or notched species [10]. Wildbook can find matches within the database: once an animal has been identified, it can be tracked in other photographs. It stores the information about who the animals are and where and when they are there in a fully developed database, and provides query tools to that data for scientists researching population demographics, species distributions, individual interactions and movement patterns [22].

Wildbook system allows to add biological data, as simple as sex and age, but also habitat and weather information which allows to truly do population counts, birth/death dynamics, species range, social interactions or interactions with other species, including humans. An example of a Wildbook for whales, Flukebook, page is shown in Figure **??**.

Using Wildbook, it is possible to connect the information about sightings of animals (*who? where? when?*) derived from images to additional relevant data, providing the historic, current, and projected context of these sightings, thus enabling new science, conservation, and education, at unprecedented scales and resolution. By layering additional data sets, covering everything from climate change and extreme weather to habitat ecology, agricultural development, urbanization, deforestation, the exotic animal trade, and the spread of disease, a much more detailed and useful picture of *what* is happening — and *why* — can be constructed within our architecture.

2. EXAMPLES OF WILDBOOK USES AND IMPACT

Using our system, estimates of population sizes and movement patterns can be far more accurate, creating a better understanding of social structures and breeding of species, relationships between predators and prey, and responses to environmental pressures, including land use by humans and long-term climate patterns. Wildlife managers are better able to monitor the health of entire populations, discover dangerous trends, and reduce conflicts between humans and wildlife. Access to information about individual animals, particularly visual information, can also increase the public's understanding of the workings of science and its role in guiding conservation. By contributing their photographs for scientific studies, visitors to parks and nature preserves in return learn the life histories of the individual animals they photograph and become connected to research projects and to the animals. We now present several examples of real impact a system like Wildbook can make in conservation policy, science, and public engagement.

2.1 Evidence-based conservation policy: Lewa

The first deployment of Wildbook was in January 2015 at Lewa Wildlife Conservancy[2] in Kenya helping manage the endangered Grevy's zebra population. The information from Wildbook for Grevy's showed that there are not enough babies surviving to adulthood mainly due to the lion predation. This lead to a change in the lion population management policy in Lewa helping save the endangered zebra.

2.2 Crowdsourcing accurate conservation data: GZGC and GGR

Knowing the number of individual animals within a population (a population census) is one of the most important statistics for research and conservation management in wildlife biology. Moreover, a *current* population census is often needed repeatedly over time in order to understand changes in a population's size, demographics, and distribution. This enables assessments of the effects of ongoing conservation management strategies. Furthermore, the number of individuals in a population is seen as a fundamental basis for determining its conservation status.

Unfortunately, producing a population census is difficult to do at scale and across large geographical areas using traditional, manual methods. One of the most popular and prevalent techniques for producing a population size estimate is mark-recapture [33, 35] via a population count. However, performing a mark-recapture study can be prohibitively demanding when the number of individuals in a population grows too large [41], the population moves across large distances, or the species is difficult to capture due to evasiveness or habitat inaccessibility. More importantly, however, a population *count* is not as robust as a population *census*; a count tracks sightings whereas a census tracks individuals. A census is stronger because it can still produce a population size estimate implicitly but also unlocks more powerful ecological metrics that can track the long-term trends of individuals. In recent years, technology has been used to help improve censusing efforts towards more accurate population size estimates [8, 13, 44, 45] and scale up[3]. However, these types of population counts are still typically custom, one-off efforts, with no uniform collection protocols or data analysis, and do not attempt to accurately track *individuals* within a population across time.

To address the problems with collecting data and producing a scalable population census, we performed the following [31]:

- using citizen scientists [9, 19] to rapidly collect a large number of photographs over a short time period (e.g. two days) and over an area that covers the expected population, and
- using computer vision algorithms to process these photographs semi-automatically to identify all seen animals.

We showed that this proposed process can be leveraged at scale and across large geographical areas by analyzing the results of two completed censuses. The first census is

[1] `http://Wildbook.org`

[2] `http://www.lewa.org/`

[3] penguinwatch.org, mturk.com

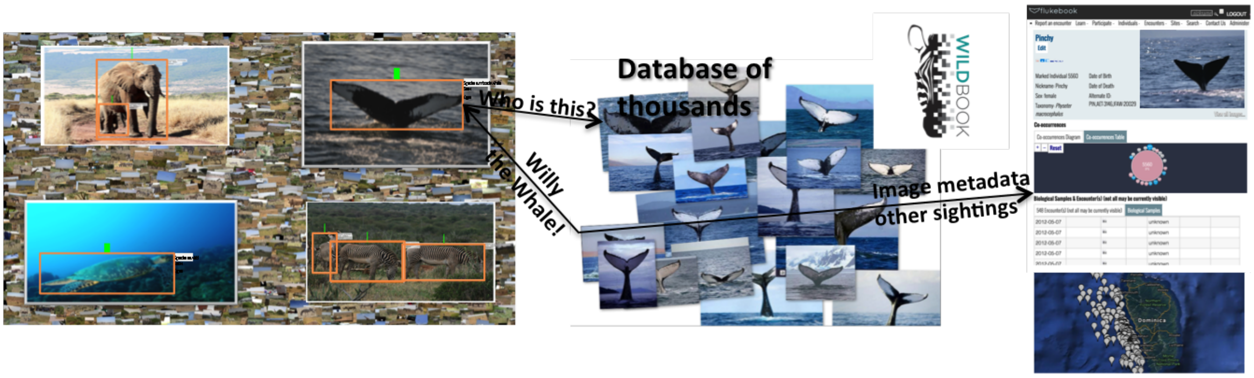

Figure 1: The Wildbook pipeline, starting with a collection of images, through species detection and individual animal identification, to the web-based data management layer and individual animal page.

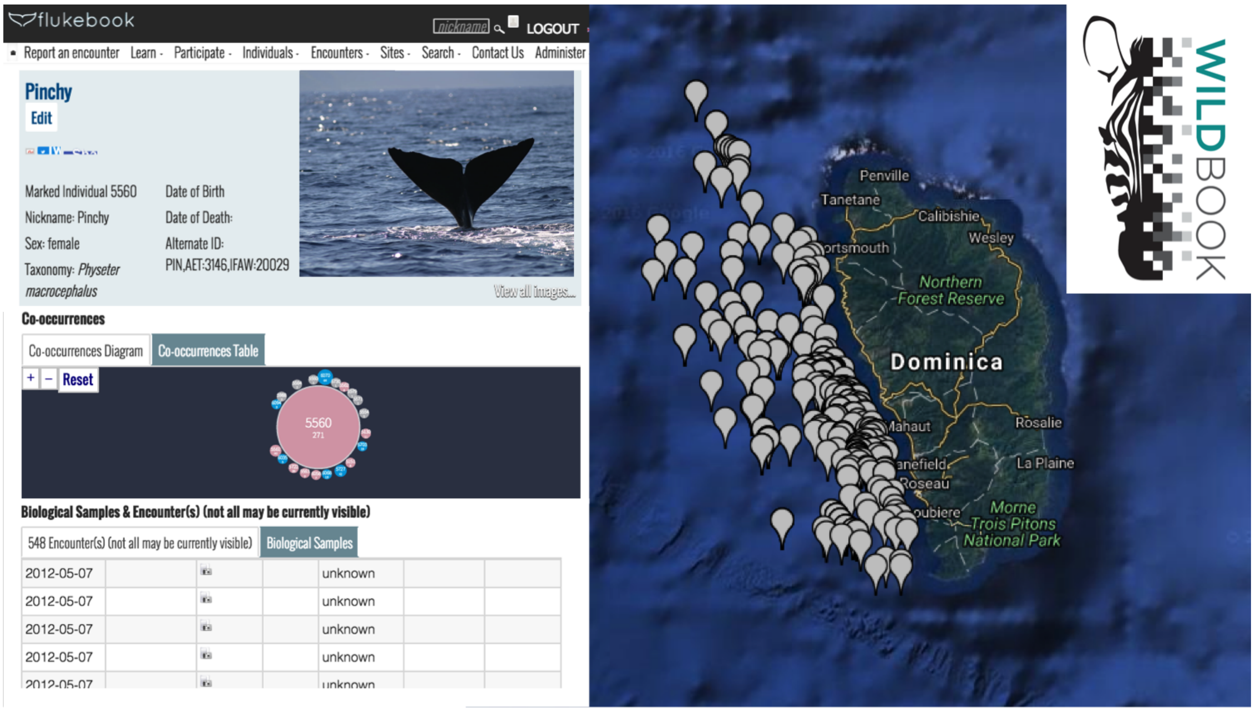

Figure 2: An example of a Flukebook (Wildbook for whales) page displaying information for an individual.

	Cars	Cameras	Photographs
GZGC	27	55	9,406
GGR	121	162	40,810

Table 1: The number of cars, participating cameras (citizen scientists), and photographs collected between the GZGC and the GGR. The GGR had over 3-times as many citizen scientists who contributed 4-times the number of photographs for processing.

	Annots.	Individuals	Estimate
GZGC Masai	466	103	119 ± 4
GZGC Plains	4,545	1,258	$2{,}307 \pm 366$
GGR Grevy's	16,866	1,942	$2{,}250 \pm 93$

Table 2: The number of annotations, matched individuals, and the final mark-recapture population size estimates for the three species. The Lincoln-Peterson estimate has a 95% confidence range.

the Great Zebra and Giraffe Count (GZGC) held March 1-2, 2015 at the Nairobi National Park in Nairobi, Kenya to estimate the resident populations of Masai giraffes (*Giraffa camelopardalis tippelskirchi*) and plains zebras (*Equus quagga*). The second is the Great Grevy's Rally (GGR) held January 30-31, 2016 in a region of central and northern Kenya covering the known migratory range of the endangered Grevy's zebra (*Equus grevyi*). While our method relies heavily on collecting a large number of photographs, it is designed to be simple enough for the average person to help collect them. Any volunteers typically must only be familiar with a digital camera and be able to follow a small set of collection guidelines.

Mark-recapture is a standard way of estimating the size of an animal population [7, 33, 35]. Typically, a portion of the population is captured at one point in time and the individuals are marked *as a group*. Later, another portion of the population is captured and the number of previously marked individuals is counted and recorded. Since the number of marked individuals in the second sample should be proportional to the number of marked individuals in the entire population (assuming consistent sampling processes and controlled biases), the size of the entire population can be estimated.

The population size estimate is calculated by dividing the total number of marked individuals during the first capture by the proportion of marked individuals counted in the second. The formula for the simple Lincoln-Peterson estimator [30] is:

$$N_{\text{est}} = \frac{K * n}{k}$$

where N_{est} is the population size estimate, n is the number of individuals captured and marked during the first capture, K is the number of individuals captured during the second capture, and k is the number of *recaptured* individuals that were marked from the first capture.

The number of cars, volunteers, and the number of photographs taken for both rallies can be seen in Table 1.

By giving the collected photographs to a computer vision pipeline, a semi-automated and more sophisticated census can be made. The speed of processing large qualities of photographs allows for a more thorough analysis of the age-structure of a population, the distribution of males and females, and the movements of individuals and groups of animals, etc. By tracking individuals, related to [21, 42], our method is able to make more confident claims about statistics for the population. The more individuals that are sighted *and* resighted, the more robust the estimate and ecological analyses will be.[4] The resulting estimates of the populations of Plains zebra and Maasai giraffe in NAirobi National Park and of the species census of the Grevy's zebra are the most accurate to date and the Grevy's zebra numbers are now used as the official numbers of the Grevy's zebra global population size by IUCN Red List [38].

2.3 Crowdsourcing conservation data at scale: Whaleshark, Flukebook, online social media

Today, Wildbooks for over a dozen species are available or are in the process of development. Wildbook for whales, Flukebook (`http://flukebook.org/`, started with just over 800 individuals less than two years ago, is fully functional and helps track, protect, and study more than 11,600 marine mammals. Wildbook for whale sharks (`http://whaleshark.org/`) is the longest running animal sighting website which started with a couple fo hundred individuals animals 16 years ago and has more than 8,000 individuals with over 50,000 sightings today. IUCN Red List uses whaleshark.org for global populatino estimates [32]. There are several projects with the WWF, World Wildlife Fund, with wildbooks for Suomi rigned seals (`http://norppagalleria.wwf.fi/`), lynx (`http://lynx.wildbook.org`), and sea turtles (the real IoT, Internet of Turtles: `http://iot.wildbook.org/`). As each Wildbook goes online, the number of identified individuals for each species grows over an order of magnitude within less than a year and Wildbook becomes the most reliable source of data for the species. Moreover, over the last year, we showed that social media can be a reliable source of information about animal populations [27] and Whaleshark.org uses public videos [46], while Flukebook is starting to use public images as a supplementary source of information.

3. WITH GREAT DATA COMES GREAT RESPONSIBILITY

3.1 Bias and accuracy

Like all data, photographic samples of animal ecology are biased. These biases may lead to inaccurate population size or dynamic estimates. For example, even for such an iconic species as snow leopard, the last population estimate was in 2003 and "many of the estimates are acknowledged to be rough and out of date" [20]. Yet, its conservation status can change depending on just a difference of a few individuals for some geographic locations. Human observers tend to overestimate population sizes since they may misidentify the same individual as different ones, while photo id provides evidence that those are indeed the same. Thus, image-based census can be used to support the more accurate population counts,

[4]Portions of the results in this section were previously reported in two technical reports: [39] for the GZGC and [2] for the GGR.

which, in turn, will affect conservation status or policies for a species. That is indeed a big responsibility.

To administer a correct population census, we must take these biases into account explicitly as different sources of photographs inherently come with different forms of bias. For example, stationary camera traps, cameras mounted on moving vehicles, and drones are each biased by their location, by the presence of animals at that location, by photographic quality, and by the camera settings (such as sensitivity of the motion sensor) [14, 16, 17, 37]. These factors result in biased samples of species and spatial distributions, which recent studies are trying to overcome [1, 23, 50].

Any human observer, including scientists and trained field assistants, is affected by observer bias [24, 25]. Specifically, the harsh constraint of being at a single given location at a given time makes sampling arbitrary. Citizen scientists, as the foundation of the data collection, have additional variances in a wide range of training, expertise, goal alignment, sex, age, etc. [11]. Nonetheless, recent ecological studies are starting to successfully take advantage of this source of data, explicitly testing and correcting for bias [47]; recent computational approaches address the question of if and how data from citizen scientists can be considered valid [49], which can be leveraged with new studies in protocol design and validation. There are multiple biases that influences the final outcome of estimating population of a certain species from images that are obtained from social media. Some of the most prominent biases which influence the data we obtain from social media are outlined in Figure 3. There are several layers of biases, accumulating in the resulting bias of estimating animal population properties from images. First, there are biases in the types of animals that people typically photograph in sufficient numbers in the first place. These may be charismatic or endangered species, or simply the ones easily observed. Second, there are biases in what images people take versus which ones they decide to share publicly on social media. These range from the Hawthorne Effect [3, 29, 40, 43] of changing behavior when knowing to be observed, to biases introduced by the demographics of the person sharing [5, 29] and the choice of the social media platform [6, 15, 28]. There are biases of our notions of beauty and aesthetics and cultural differences. Any mark-recapture model used to estimate the population size makes many assumptions and introduces its own biases. The fundamental question, however, is: *Do any of these actually affect the estimates of the population size and other parameters and if so, how?* Menon has begun to answering this question [26] but a lot of work remains to be done. Moreover, combining these differently biased sources of data mutually constrains these biases and allows much more accurate statistical estimates than any one source of data would individually allow [4].

3.2 Security

As mobile phone masts went up across the world's jungles, savannahs and mountains, so did poaching. Wildlife crime syndicates can not only coordinate better but can mine growing public data sets [18]. Tourists are now warned not to geotag photos [51] of big game but most have no idea how to comply. The push for open research data is often in conflict with conservation, and new technologies such as animal recognition from photographs bring new hazards. Privacy matters for tigers, for snow leopards, for lions, and indeed for elephants and rhinos and even tortoises or any other endangered species. The privacy problems that humans face in an online globalized world mostly have parallels for wildlife, with some of them different in fascinating ways. The issues sprawl across many of the technical and policy areas of classical security and privacy, from insider threats to jurisdictional tangles, from multilevel policies to secondary-use hazards, from covert channels to geofencing, and from security economics to usability. Conservation law is stuck in the twentieth century, and no-one seems to have started to think about information security policy. It is urgent to address the issues of privacy and security for image-based wildlife data. To start, Wildbook has worked with the Center for Trustworthy Scientific Cyberinfrastructure to design a secure system to prevent data leakage and poachers' access yet more work remains to be done.

4. CONCLUSIONS AND CHALLENGES

We have designed, implemented and deployed a prototype, and are continuing to develop Wildbook, an image-based ecological information system. Using image analysis algorithms and state-of-the-art information management infrastructure, Wildbook adds images, opportunistically and intentionally crowdsourced and scientifically collected, to the source of data about animals and provides the analytical tools to gain scientific and conservation insight from those data. As the new type of data, the images are not only augmenting the scale and resolution of existing scientific and conservation inference, but allow *new types* of questions that lead to new scientific understanding of why animals do what they do, as well as a change in the conservation policy. Moreover, we have already demonstrated that Wildbook provides a platform and a tool for a much more personal and committed public engagement in science and conservation than has been available to date. By enabling events such as the Great Zebra and Giraffe Count and Great Grevy's Rally, Wildbook both presents an instant, easily available, no-training-needed route for general public contribution to science and conservation, as well as creating a personal bond with animals and nature by providing an instant individual animal identity. Moreover, it provides data for evidence-based conservation policy at large scale and high resolution over time, space, and individual animals.

However, to achieve these new insights and engagement, many challenges need to be overcome. In addition to the many computational, scientific, and societal challenges, there are two directly related to data. First, Wildbook requires a new infrastructure that can function, synchronize, and coordinate across many platforms, from the mobile phones and GPS cameras of the citizen scientists, the bandwidth and electricity-starved research stations and conservation outposts in remote uninstrumented locations, to the cloud infrastructure containing information about entire species and regions, as well as the algorithms necessary for its analysis. The data-related aspects of this infrastructure challenges are about information aggregation, integration, synchronization, semantic complexity, and access control. The second big data-related challenge that the new data sources and the enabled use of those data present is the unknown data biases that challenge traditional computational methods and analytical tools. From the simplest population size and species range estimates, the traditional methods rely on a uniform random sampling regime. While it is not clear that the assumption is true for any of the data collection methods, it

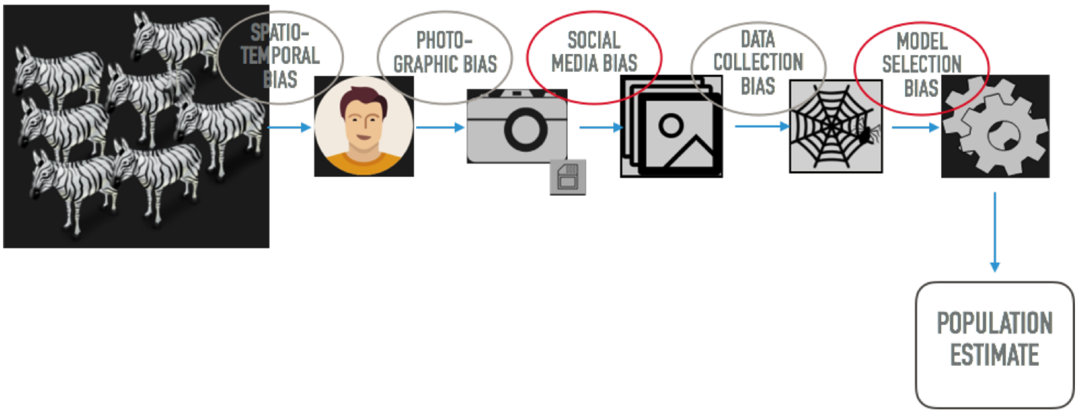

Figure 3: High level schematic representation of the problem of population estimation of wildlife species using images from social media, its challenges and biases in play.

is most definitely does not hold for the data coming from citizen scientists' and tourists' photographs. Our preliminary analysis shows that there are wide variations in the number and rate of images taken and complex patterns of camera and species fatigue that arise from demographic, cultural, and event-specific factors. Accounting for or leveraging those in designing the new generation of analytical tools is a challenge and a goal of Wildbook and it is critical for reliable conservation policy decisions.

Finally, one potential use of Wildbook that presents very different data-related challenges is as a tool in wildlife crime prevention. Wildbook ability to track individuals through photographs during their life, as well as identifying these individuals by a reasonably sized part of the body later, given that part has been previously photographed, allows photographic evidence to be used both in forensics and as a deterrence in poaching, killing, and illegal trafficking of animals. Wildlife crime is threatening to wipe out many charismatic species from the planet: rhino population (across species) is down 90% from its hight [?] and 100,000 elephants were killed over the last 3 years in Africa for ivory [?]. Many charismatic fauna species, such as leopards, elephants, tigers, zebras, snow leopards, turtles, and tortoises, have individual-level uniquely identifiable body patterns, essentially 'body-prints'. With an Wildbook app, a law-enforcement official would be able to take a picture of an animal crime victim (alive or not) and be able to find a match in the reference database if one exists. Thus, the identity, geographic origin and life history of the animal will be instantly available. There are two types of primary users. The first are conservation and wildlife managers responsible for overseeing a particular endangered species of identifiable fauna. They collect photographs and submit information to Wildbook to build up the databases. The second type of users are law enforcement officials who would take pictures of animals or their hides or carcasses, submit these to Wildbook to obtain, if known, the identity and origin of the specimen.

The use of Wildbook or a similar information system in the for the purposes of conservation of highly endangered species or wildlife crime prevention presents a unique problem in data security and privacy. Ironically, every new technology is a "double-edged sword" for wildlife and is often used by the criminals to aid in illegal wildlife trade [?], as highlighted by the recent killing of Cecil the lion (who was tracked using his GPS collar). Thus, the location, current or predicted, of criminally valuable or highly endangered species or individuals must be protected. Privacy and security protocols must be developed to protect animal data. The balance of opening data for science, conservation, and human curiosity must be weighed against the danger of exposing animals to extinction.

Acknowledgements

This work was supported in part by a gift from Microsoft Research, funding from the Department of Computer Science at the University of Illinois at Chicago, and an NSF grants CNS-1453555 EAGER: Prototype of an Image-Based Ecological Information System (IBEIS) and EF-1550853 EAGER-NEON: Image-Based Ecological Information System (IBEIS) for Animal Sighting Data for NEON. We would also like to acknowledge in-kind support, volunteer time, and collaborations with Mpala Research Centre (Kenya), Lewa Wildlife Conservancy (Kenya), WildlifeDriect.org, Grevy's Zebra Trust, and Kenya Wildlife Service.

5. ADDITIONAL AUTHORS

6. REFERENCES

[1] M. Ancrenaz, A. Hearn, J. Ross, R. Sollmann, and A. Wilting. *Handbook for Wildlife Monitoring using Camera-traps*. BBEC, Kota Kinabalu, Sabah, Malaysia, 2012.

[2] T. Berger-Wolf, J. Crall, J. Holberg, J. Parham, C. Stewart, B. L. Mackey, P. Kahumbu, and D. Rubenstein. The Great GrevyâĂŹs Rally: The Need, Methods, Findings, Implications and Next Steps. Technical Report, Grevy's Zebra Trust, Nairobi, Kenya, Aug. 2016.

[3] M. S. Bernstein, A. Monroy-Hernández, D. Harry, P. André, K. Panovich, and G. G. Vargas. 4chan and/b: An analysis of anonymity and ephemerality in a large online community. In *ICWSM*, pages 50–57, 2011.

[4] T. J. Bird, A. E. Bates, J. S. Lefcheck, N. A. Hill, R. J. Thomson, G. J. Edgar, R. D. Stuart-Smith, S. Wotherspoon, M. Krkosek, J. F. Stuart-Smith,

G. T. Pecl, N. Barrett, and S. Frusher. Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, 173(1):144 – 154, 2014.

[5] J. Brenner and M. Duggan. The demographics of social media users. *Consultado en*, 2013.

[6] A. Bruns and S. Stieglitz. Twitter data: what do they represent? *it-Information Technology*, 56(5):240–245, 2014.

[7] D. Chapman and N. Chapman. *Fallow deer: their history, distribution, and biology*. Dalton, Ithaca, NY, 1975.

[8] M. J. Chase, S. Schlossberg, C. R. Griffin, P. J. BouchÃľ, S. W. Djene, P. W. Elkan, S. Ferreira, F. Grossman, E. M. Kohi, K. Landen, P. Omondi, A. Peltier, S. J. Selier, and R. Sutcliffe. Continent-wide survey reveals massive decline in African savannah elephants. *PeerJ*, 4(1):e2354, Aug. 2016.

[9] J. P. Cohn. Citizen science: Can volunteers do real research? *BioScience*, 58(3):192–197, Mar. 2008.

[10] J. Crall, C. Stewart, T. Berger-Wolf, D. Rubenstein, and S. Sundaresan. Hotspotter - patterned species instant recognition. *Applications of Computer Vision (WACV)*, pages 230–237, 2013.

[11] J. L. Dickinson, B. Zuckerberg, and D. N. Bonter. Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1):149–172, 2010.

[12] W. W. F. for Nature. The living planet report, 2016. Accessed May 1, 2017.

[13] T. Forrester, W. J. McShea, R. W. Keys, R. Costello, M. Baker, and A. Parsons. eMammalâĂŞcitizen science camera trapping as a solution for broad-scale, long-term monitoring of wildlife populations. In *North America Congress for Conservation Biology*, Missoula, Montana, July 2014.

[14] R. J. Foster and B. J. Harmsen. A critique of density estimation from camera-trap data. *The Journal of Wildlife Management*, 76(2):224–236, 2012.

[15] S. Gonzalez-Bailon, N. Wang, A. Rivero, J. Borge-Holthoefer, and Y. Moreno. Assessing the bias in samples of large online networks. *Social Networks*, 38:16–27, 2014.

[16] A. Hodgson, N. Kelly, and D. Peel. Unmanned Aerial Vehicles (UAVs) for Surveying Marine Fauna: A Dugong Case Study. *PLoS ONE*, 8(11):e79556, 2013.

[17] V. Hombal, A. Sanderson, and D. R. Blidberg. Multiscale adaptive sampling in environmental robotics. In *In Proceedings of the 2010 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 80–87, Salt Lake City, UT, Sept. 2010.

[18] M. Hower. Poacher's delight: Technology is a double-edged sword for wildlife. *Green Biz*, 2015.

[19] A. Irwin. *Citizen Science: A Study of People, Expertise and Sustainable Development*. Environment and Society. Routledge, New York, NY, 1995.

[20] R. Jackson, D. Mallon, T. McCarthy, R. Chundaway, and B. Habib. Panthera uncia. The IUCN Red List of Threatened Species 2008: e.T22732A9381126. `http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T22732A9381126`, 2008. [Online; accessed 07 July 2017].

[21] G. M. Jolly. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1/2):225–247, 1965.

[22] R. Kondos. From binoculars to big data: Citizen scientists use emerging technology in the wild. *O'Reilly Media*, July 2017.

[23] N. W. Maputla, C. T. Chimimba, and S. M. Ferreira. Calibrating a camera trapâĂŞbased biased markâĂŞrecapture sampling design to survey the leopard population in the N'wanetsi concession, Kruger National Park, South Africa. *African Journal of Ecology*, 51(3):422–430, 2013.

[24] D. M. Marsh and T. J. Hanlon. Observer gender and observation bias in animal behaviour research: experimental tests with red-backed salamanders. *Animal Behaviour*, 68(6):1425 – 1433, 2004.

[25] D. M. Marsh and T. J. Hanlon. Seeing What We Want to See: Confirmation Bias in Animal Behavior Research. *Ethology*, 113(11):1089–1098, 2007.

[26] S. Menon. Animal Wildlife Population Estimation Using Social Media Images. Master's thesis, University of Illinois at Chicago, Chicago, IL, 2017.

[27] S. Menon, T. Berger-Wolf, E. Kiciman, L. Joppa, C. V. Stewart, J. Parham, J. Crall, H. J., and J. Van Oast. Animal population estimation using flickr images. In *In Proceedings of the 2nd International Workshop on the Social Web for Environmental and Ecological Monitoring (SWEEM 2017)*, Troy, NY, 2017.

[28] F. Morstatter, J. Pfeffer, and H. Liu. When is it biased?: assessing the representativeness of twitter's streaming api. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 555–556. ACM, 2014.

[29] A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. pages 1–44, 2016.

[30] S. W. Pacala and J. Roughgarden. Population experiments with the Anolis lizards of St. Maarten and St. Eustatius. *Ecology*, 66(1):129–141, Feb. 1985.

[31] J. Parham, J. Crall, C. Stewart, T. Berger-Wolf, and D. Rubenstein. Animal Population Censusing at Scale with Citizen Science and Photographic Identification. 2016.

[32] S. Pierce and B. Norman. Rhincodon typus. The IUCN Red List of Threatened Species 2016: e.T19488A2365291. `http://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T19488A2365291`, 2016. [Online; accessed 07 July 2017].

[33] R. Pradel. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics*, 52(2):703–709, June 1996.

[34] I. G. S. Programme. The IUCN red list of threatened species website, 2017. Accessed May 14, 2017.

[35] D. S. Robson and H. A. Regier. Sample size in petersen markâĂŞrecapture experiments. *Transactions of the American Fisheries Society*, 93(3):215–226, July 1964.

[36] A. S. Rodrigues, J. D. Pilgrim, J. F. Lamoreux,

M. Hoffmann, and T. M. Brooks. The value of the iucn red list for conservation. *Trends in ecology & evolution*, 21(2):71–76, 2006.

[37] J. M. Rowcliffe, R. Kays, C. Carbone, and P. A. Jansen. Clarifying assumptions behind the estimation of animal density from camera trap rates. *The Journal of Wildlife Management*, 77(5):876–876, 2013.

[38] D. Rubenstein, B. Low Mackey, Z. Davidson, F. Kebede, and S. King. Equus grevyi. The IUCN Red List of Threatened Species 2016: e.T7950A89624491. `http://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T7950A89624491`, 2016. [Online; accessed 07 July 2017].

[39] D. I. Rubenstein, C. V. Stewart, T. Y. Berger-Wolf, J. Parham, J. Crall, C. Machogu, P. Kahumbu, and N. Maingi. The Great Zebra and Giraffe Count: The Power and Rewards of Citizen Science. Technical Report, Kenya Wildlife Service, Nairobi, Kenya, July 2015.

[40] S. Y. Schoenebeck. The secret life of online moms: Anonymity and disinhibition on youbemom. com. In *ICWSM*. Citeseer, 2013.

[41] G. Seber. *The Estimation of Animal Abundance and Related Parameters*. Blackburn, Caldwell, NJ, 2 edition, 2002.

[42] G. A. Seber. A note on the multiple-recapture census. *Biometrika*, 52(1/2):249–259, 1965.

[43] M. Shelton, K. Lo, and B. Nardi. Online media forums as separate social lives: A qualitative study of disclosure within and beyond reddit. *iConference 2015 Proceedings*, 2015.

[44] R. Simpson, K. R. Page, and D. De Roure. Zooniverse: Observing the World's Largest Citizen Science Platform. In *In Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 1049–1054, New York, NY, 2014. ACM.

[45] A. Swanson, M. Kosmala, C. Lintott, R. Simpson, A. Smith, and C. Packer. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, 2(150026):1–14, June 2015.

[46] J. Van Oast and J. Parham. Finding Big Fish Online: How Wild Me Is Turning Vacation Videos Into Shark Science with Artificial Intelligence. `http://www.prweb.com/releases/2017/05/prweb14334385.htm`, 2017. [Online; accessed 07 July 2017].

[47] A. J. van Strien, C. A. van Swaay, and T. Termaat. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50(6):1450–1458, Sept. 2013.

[48] C. Welch. Orca killed by satellite tag leads to criticism of science practices. *National Geographic*, Oct. 2016.

[49] A. Wiggins, G. Newman, R. D. Stevenson, and K. Crowston. Mechanisms for Data Quality and Validation in Citizen Science. In *2011 IEEE Seventh International Conference on e-Science Workshops*, pages 14–19, Washington, DC, 2011.

[50] Y. Xue, I. Davies, D. Fink, C. Wood, and C. P. Gomes. Avicaching: A two stage game for bias reduction in citizen science. In *In Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, pages 776–785, Sao Paulo, Brazil, 2016.

[51] M. Zhang. Geotagged Wildlife Photos Help Poachers Kill Endangered Animals. *PetaPixel*, June 2014.