Real time massive online citizen science biodiversity programs: lessons from butterflies

Kathleen Prudic (University of Arizona), Kent McFarland (Vermont Center for Ecostudies), Rebecca Hutchinson (Oregon State University), Jeffrey Oliver (University of Arizona), Jeremy Kerr (University of Ottawa), Maxim Larrivée (Montreal Space for Life) and Elizabeth C. Long (Mohonk Preserve)
Computers change our world

Think different.
Computers have not increased productivity.

https://www.marketplace.org/2016/05/02/world/explaining-americas-productivity-downturn
Increasing productivity requires innovation

It took manufacturing over 50 years to figure out how to use electricity to boost US worker productivity

90% of the world’s data was collected in the past year (Age of Big Data)

Source: IDC, 2014

* One zettabyte is equal to a billion terabytes, about 12.5 billion 80 gigabyte hard drives
Science is about scaling up
Data Science is emerging to tackle big scientific questions with this data.
Big Data is changing biodiversity research quickly.

Environmental Sensors

Citizen Science
Both provide huge quantities of data to scale on the cheap

And wow, we need it

Research funding is decreasing

The need for biodiversity data is growing
Historically, biodiversity data has been collected two ways by two different groups:

- **Museum Collections** - presence only data
- **Checklist Surveys** - presence/absence data
Some QA/QC challenges as data is aggregated

- Inconsistent metadata
- Mostly non-digitized
- Collection bias
Past may not predict present and future

All mass extinctions have been correlated with increases in atmospheric CO2
New web platforms are collecting biodiversity data at unprecedented scale

iNaturalist (https://www.inaturalist.org)  
Presence-only data like museums

eButterfly (http://www.e-butterfly.org)  
Presence-absence data like surveys
Collaborate with citizen volunteers

72 million US residents watch wildlife for fun

10 million US residents watch butterflies a minimum of 85 hours a year each

~ $1,700,000,000 in butterfly volunteer hours a year
Citizen Science data is getting big, and fast!
What do we do with this data?

iNaturalist (https://www.inaturalist.org) 

~200,000 North American butterfly records

eButterfly (http://www.e-butterfly.org)

~300,000 North American butterfly records
Where do species live?

- Observational Map
- Monthly Species Distribution Maps
- Annual Species Distribution Map
Where are the biodiversity hotspots?

Biodiversity hotspots can be estimated by combining SDMs
QA/QC problems that impact our analyses at scale

- Participants vary in their ability to identify organisms
- Participants vary in their desire to travel
- Participants vary in what organisms they like to engage with
Using a Human-AI combination
The core butterfly analytics team!

Rebecca Hutchinson
Oregon State University

Jeff Oliver
University of Arizona

Ben Hickson
University of Arizona

Ryan Farrell
Brigham Young University
Visual recognition of butterfly photos helps ID

iNaturalist AI suggests IDs

eButterfly uses host data to inform butterfly ID
ID bias is based on expertise

Experts are more likely to mis-ID a common species for a rarer one

Beginners are more likely to mis-ID a rare species for a common one
Most observations are close to home
Use models to identify under sampled locations
Use predictive models to generate potential coordinates for butterfly habitats in those areas.
Then partner with other groups to sample those locations

http://www.adventurescientists.org/pollinators.html
Photographs are biased toward larger butterflies

Adding other collection methods such as Malaise traps can help document the small b’flies (BIOSCAN Long et al)
More data is coming

Build processes and tools to increase understanding and conservation of biodiversity in fine scale across continents

Precision conservation!
Many thanks to ...