SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

eBird: Curating Citizen Science Data for Use by Diverse Communities

Carl Lagoze (University of Michigan)

Andrea Wiggins, Brian Sullivan, Steve Kelling (Cornell Laboratory of Ornithology)

IDCC 2014
February 26, 2014

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Acknowledgements

- NSF SoCS program
- Other Collaborators
  - Weng-Keen Wong (Oregon State CS)
  - Jun Yu (Oregon State CS)
  - Soumik Chatterjee (UM Statistics)
- Entire eBird community and team

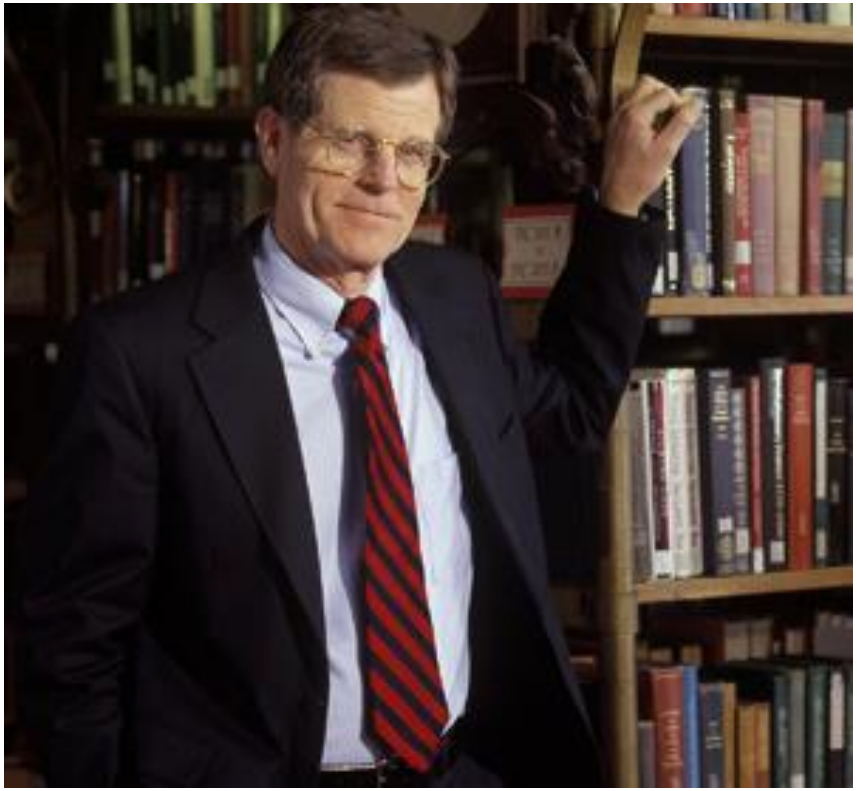
CROWDSOURCING

Citizen Science

Harnessing human intelligence to solve computational problems that are beyond the scope of existing machine intelligence:

- ESP Game – image labels
- FoldIt – protein structure
- Galaxy Zoo – galaxy classification
- reCAPTCHA – OCR of old print material

The porous “control zone”



R. Atkinson, “Library Functions, Scholarly Communication, and the Foundation of the Digital Library: Laying Claim to the Control Zone,” Libr. Q., vol. 66, no. 3, 1996.

Data quality <-> Data curation

eBird

The Cornell Lab of Ornithology

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Role of citizen science for recording ecological phenomena

- Essential because of scope of observation space and complexity of identification
- But …
  - No ground truth
  - No repeatability
- *Sensor* vs *Processor* role

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

eBird

- More than (3rd quarter 2013 figures) …
  - **160** million observations submitted
  - 10 million checklists entered
  - 150 thousand contributors
  - >95% species coverage
  - 220 countries

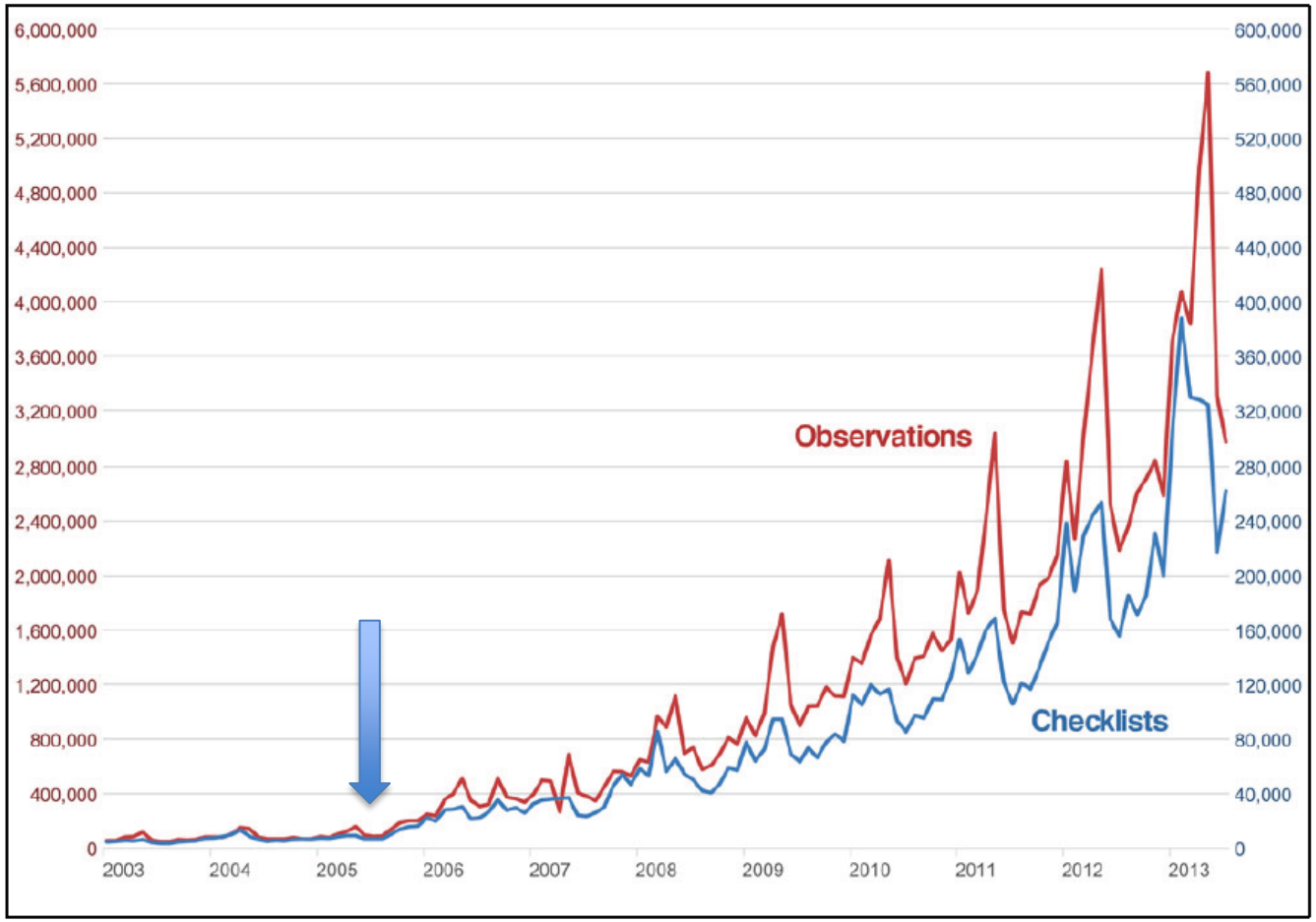
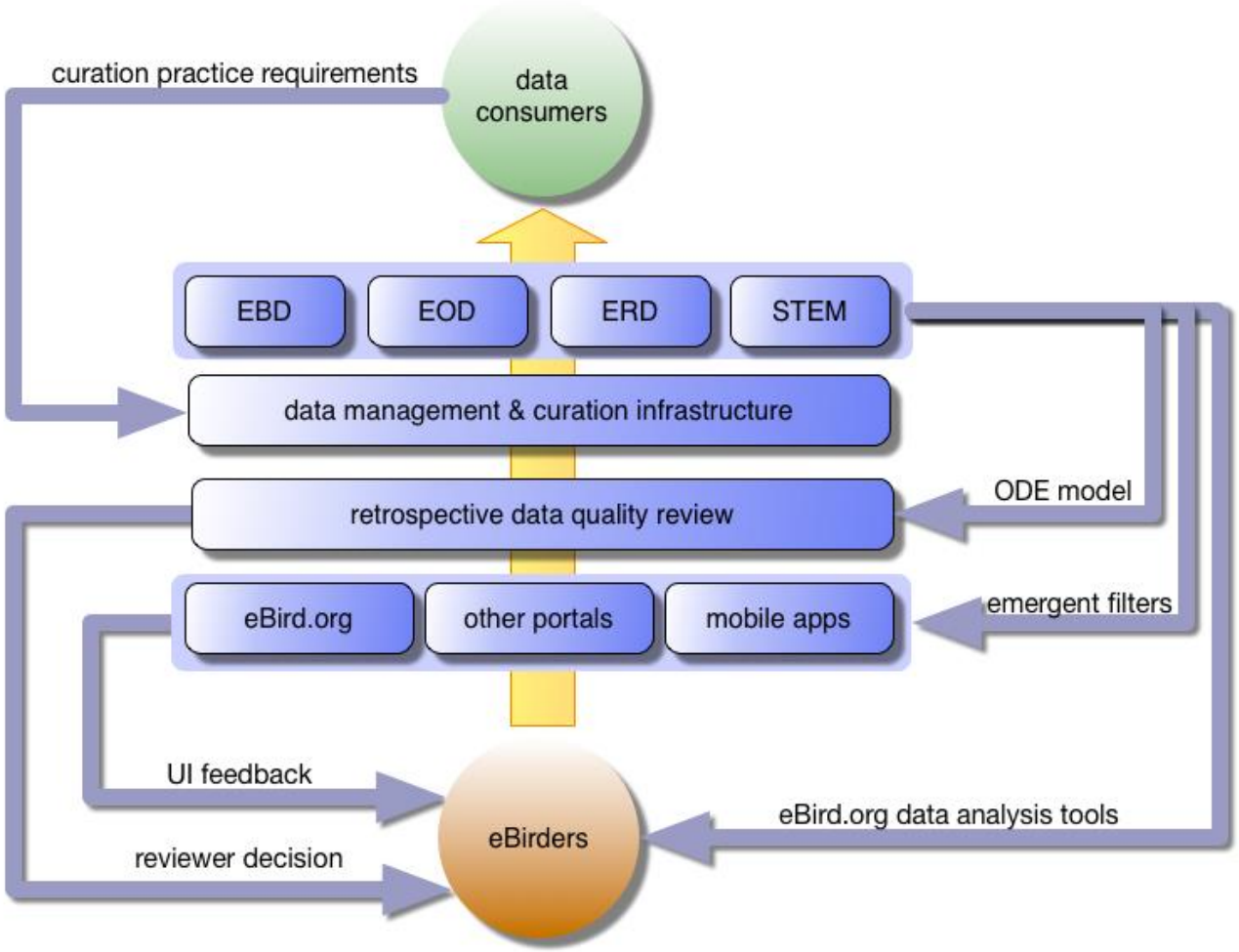


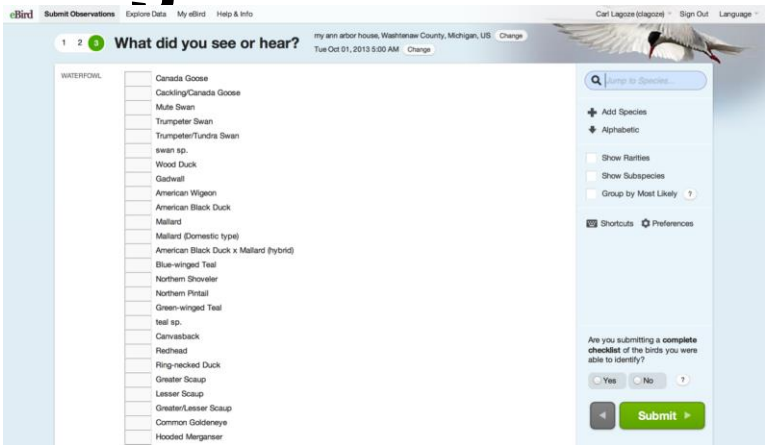
Figure 2. Growth in eBird. The number of observations and checklists submitted monthly since 2003. Note within-year fluctuations, with contributions dropping in summer and increasing through May.

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

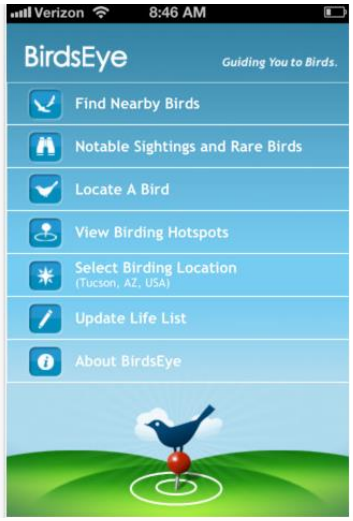


Keys to eBird success among birders

- Ease of use, use anywhere



- No joining requirements
- Tools for improving personal skills and expertise
- Appeal to birders “benevolent competitiveness”
- Use common metaphors (checklists, observations)

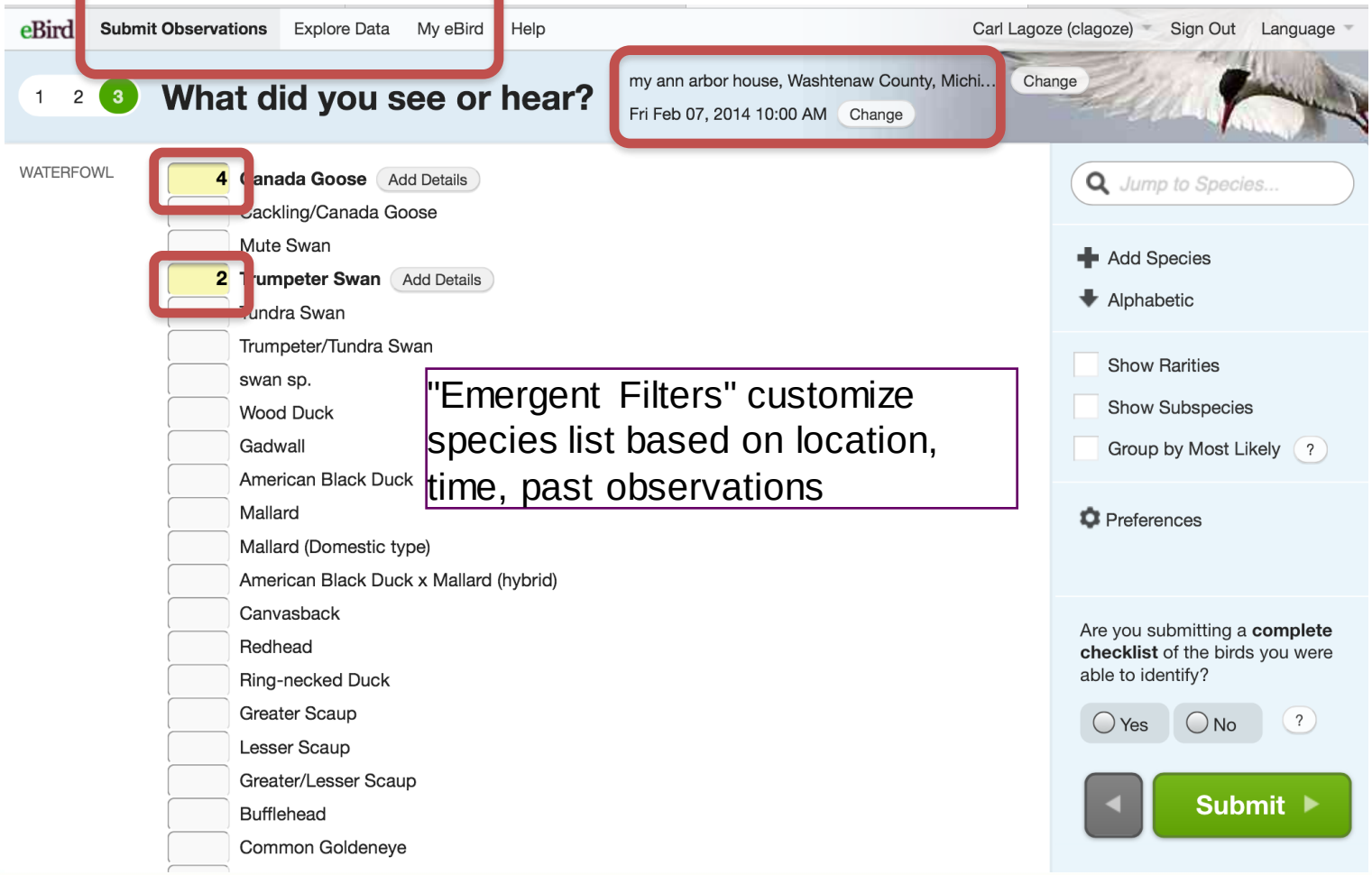


An expert
Not an expert
Can you spot the difference ?
© Bev Webb 2012

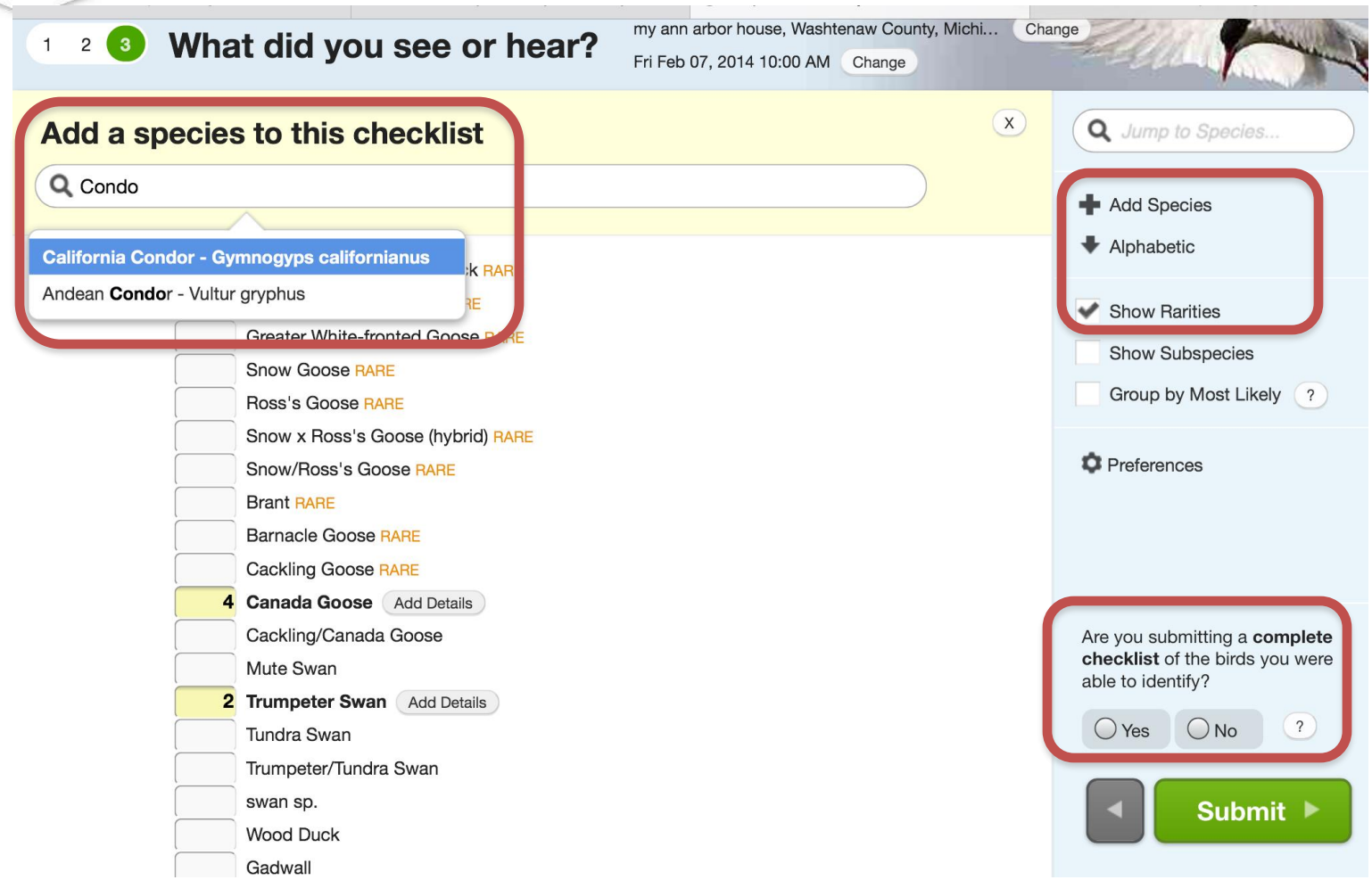
QUALITY
DATA
QUALITY

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

"Quality control" in the UI



SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

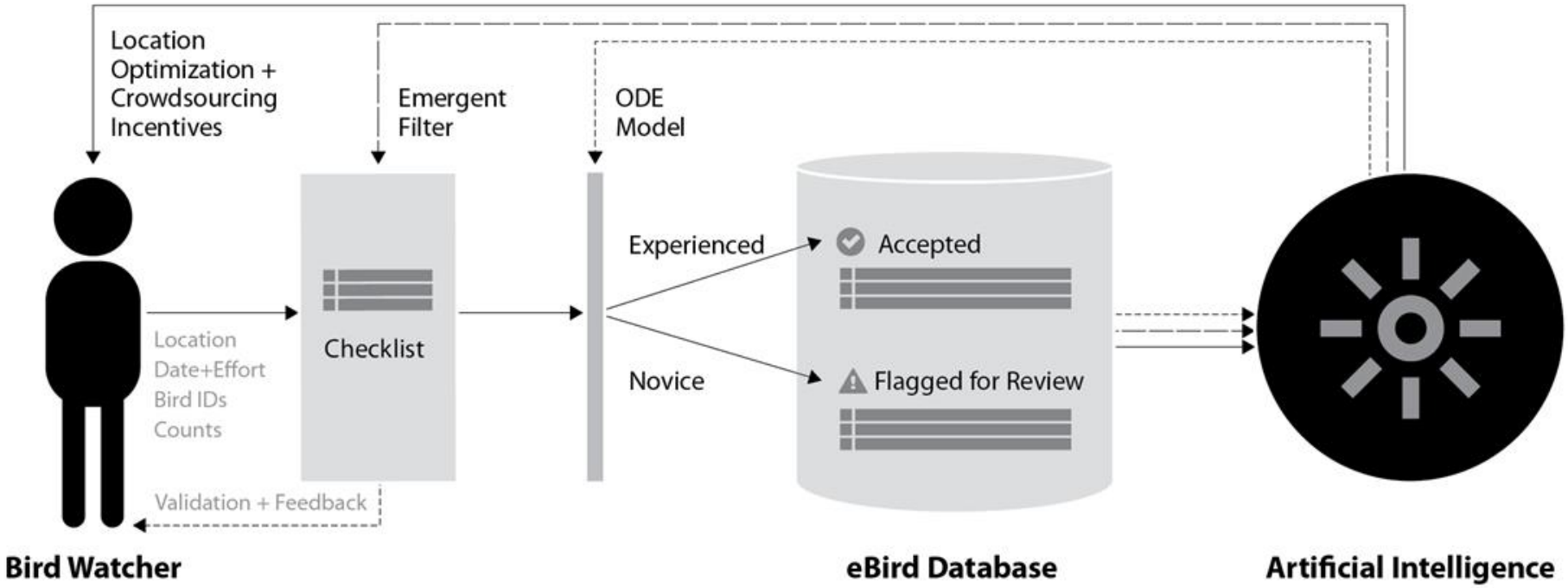


SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Retrospective quality control through human intervention

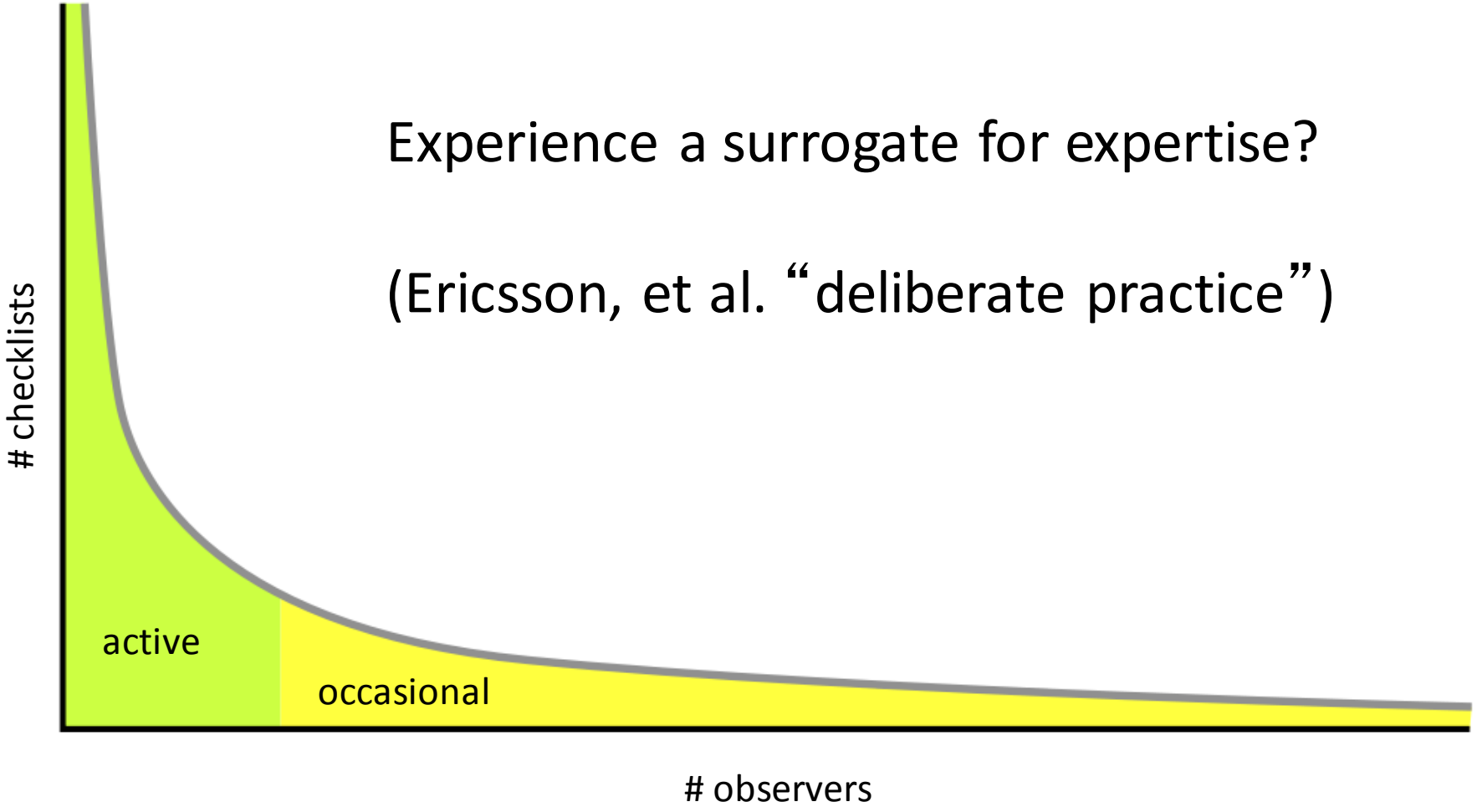
SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

NSF SoCs: Human Computational Approach for Improving Data Quality in Citizen Science Projects

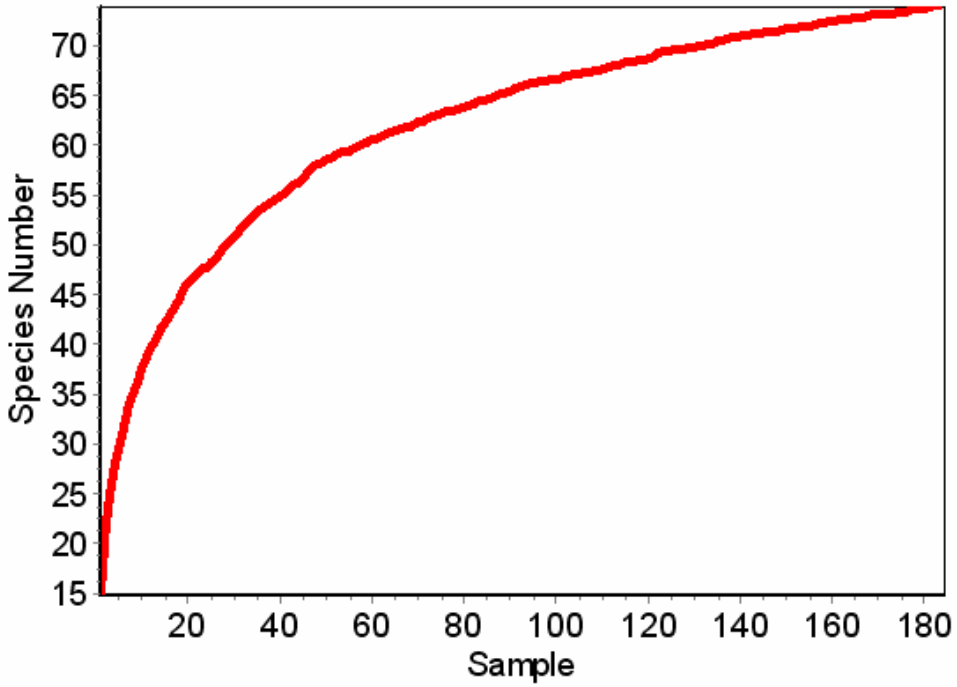


Experience a surrogate for expertise?

(Ericsson, et al. “deliberate practice”)



Species Accumulation Curve



SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Experienced observers more efficient?

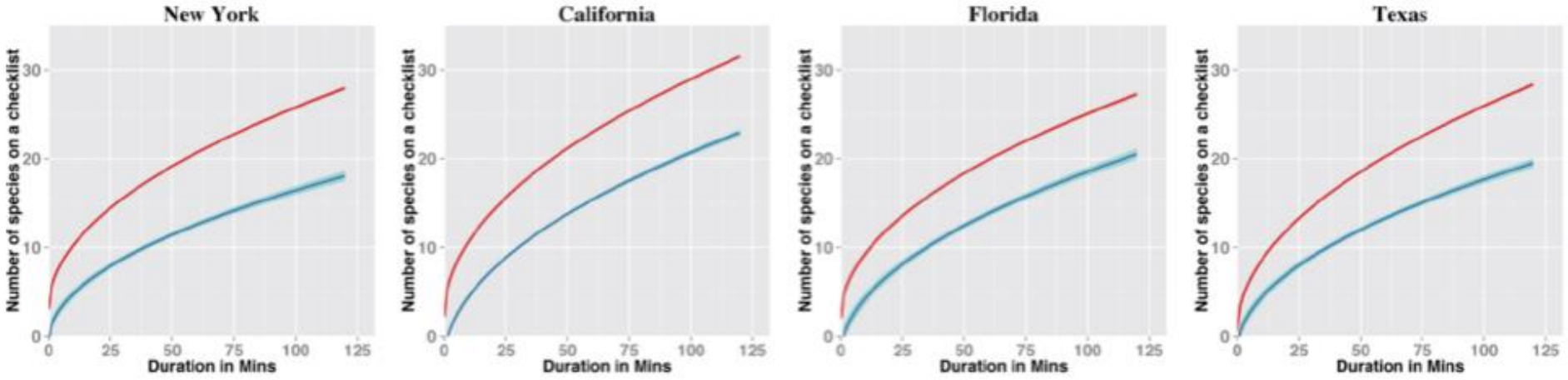


Figure 4. Species accumulation curves by state: red curve represents active users, blue for occasional users. Shaded regions around each curve represent 95% confidence intervals.

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

eBird data access policy

- Must register with eBird
- Must provide a set of basic metadata about their intended use including their name, country of residence, project type, title and abstract project, as well as their affiliation
- Must agree to the eBird Terms of Use policy, which prohibits commercial use of the data and requires them to properly attribute the Cornell Lab of Ornithology in whatever results from the use of the data

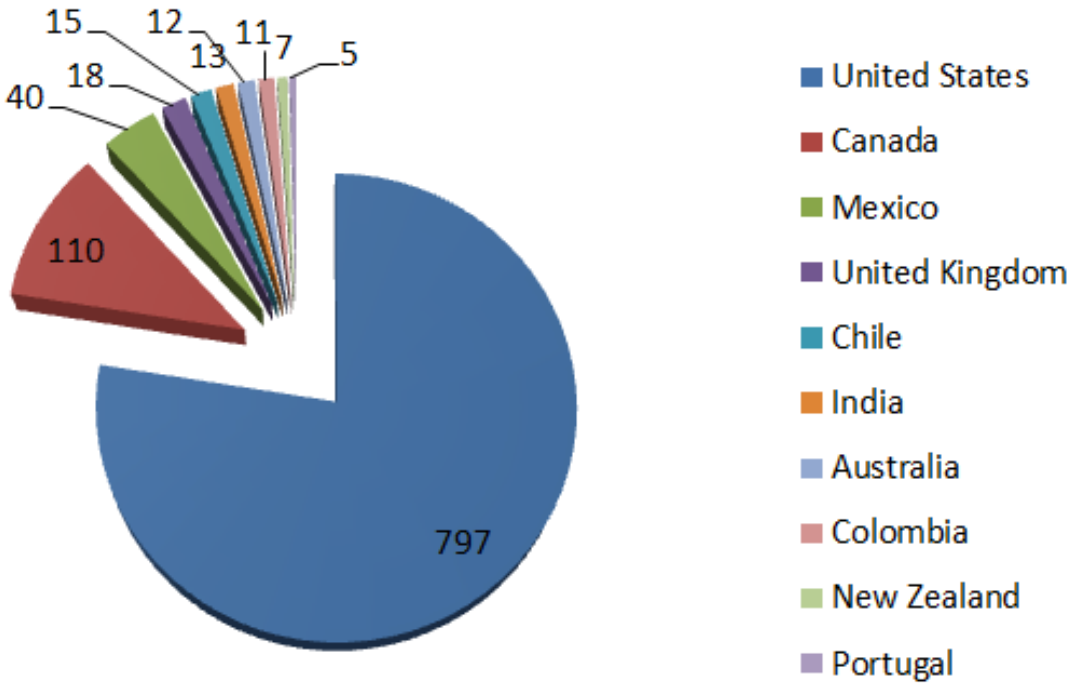
SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

eBird data products

- From clearing houses (GBIF, DataONE, no registration, no downloader data collection)
  - **eBird Observational Dataset** (EOD): primary species-occurrence data
- From eBird according to data use policy (registration required, downloader data collected)
  - **eBird Basic Dataset** (EBD): checklist data
  - **eBird Reference Dataset** (ERD): checklist data plus environment variables, e.g., land use, climate, population density
- From eBird on special request
  - **Spatial-Temporal Exploratory Models Dataset** (STEM): statistical product derived from ERD projecting species distribution

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

eBird data user community



SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

eBird data user community

Project titles and descriptions

Organization names

The State of the Birds 2011

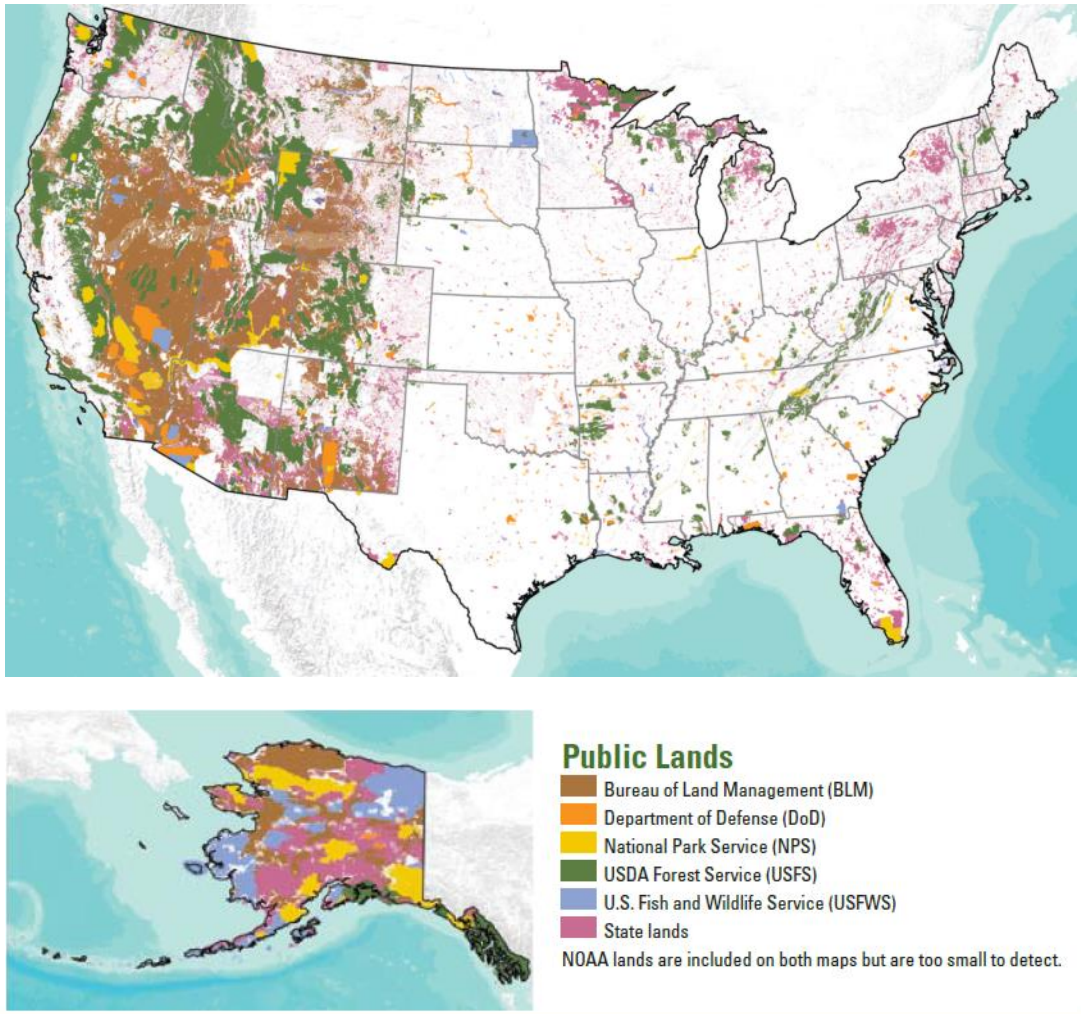
Report on Public Lands and Waters

United States of America

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

U.S. Public Lands

- Protected Areas Database for the U.S. (PAD-US 1.1; USGS-GAP)
- Includes Federal and State agencies
- Biodiversity protection status
- Bird distributions overlaid on public land ownership map

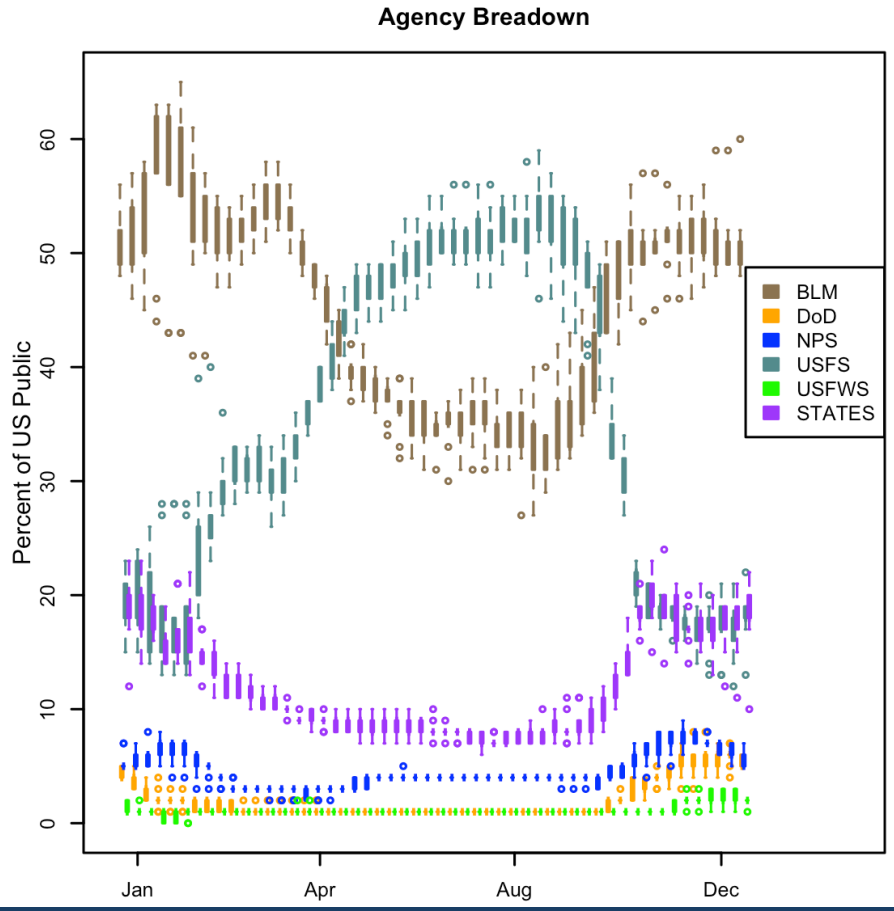


Maps produced by National Gap Analysis Program

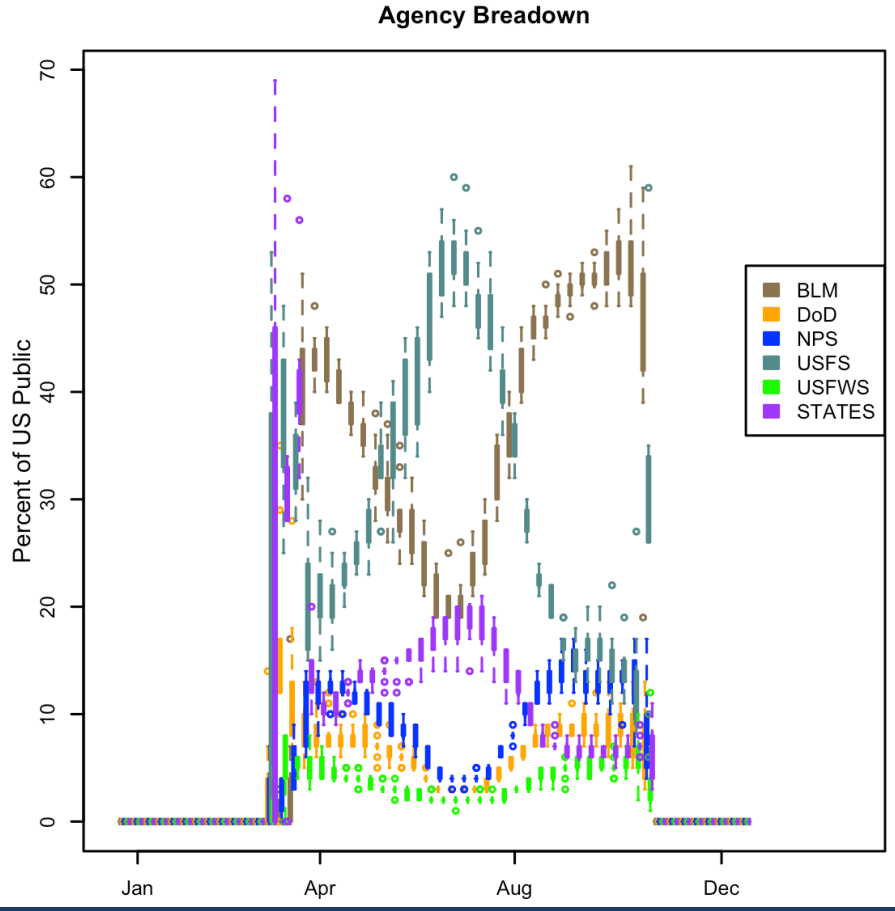
SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Stewardship Responsibilities for Birds

Mountain Bluebird



Willow Flycatcher



SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Wrap-up

- *Citizen science data quality* is not an oxymoron
- Data (and metadata) quality is a complex research issue in the mash up, crowd sourcing age
- Data curation requires a wholistic view of the data lifecycle from input through access
- Open data + usage metadata -> more effective curation

SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

Thanks!!

Questions??