Deep Lake Explorer Update

8/27/2018

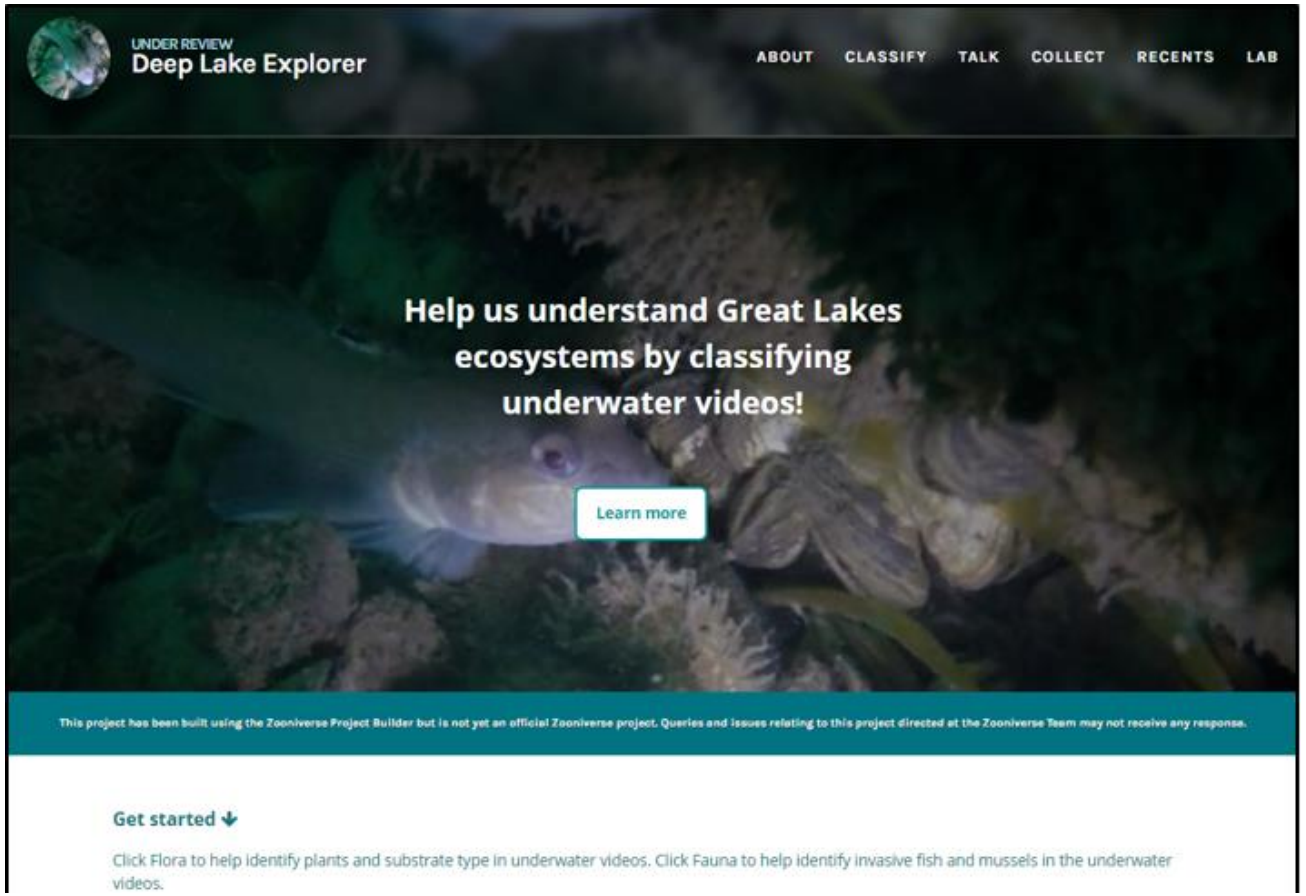
Molly Wick, wick.molly@epa.gov
Mari Nord, nord.mari@epa.gov

EPA Office of Research and Development • EPA Office of Water • EPA Region 5
EPA Great Lakes National Program Office • Oak Ridge Institute for Science • GDIT

https://www.zooniverse.org/projects/usepa/deep-lake-explorer

Project goals

- Evaluate a web-based citizen science approach to analyzing underwater videos in the Great Lakes
  - How does the precision and accuracy of data produced by citizen scientists compare to the data produced by experts?
  - What effects if any does video quality and attribute selection have when comparing analysis of experts and citizen scientists.
- Can crowdsourcing analysis of underwater video…
  - Reliable, reproducible data
  - Be cost-effective
  - Timely results
  - Educate and foster stewardship for the Great Lakes
  - Meet the data needs of managers



Outline

- Beta-test overview
- Results
  - Feedback
  - Quantitative Results
- Next Steps

Video preparation for beta-test

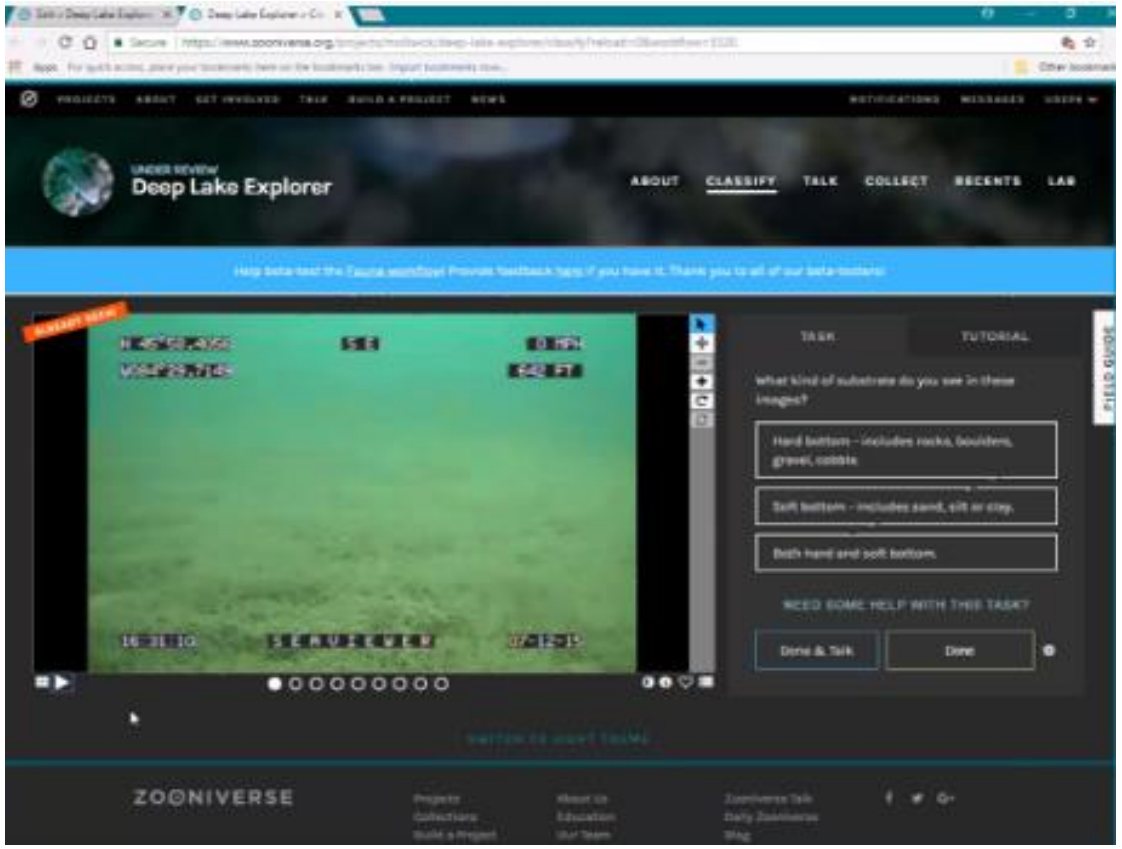
- 52 videos selected for both workflows
  - Representative of range of conditions in full dataset
    - Future – select primarily videos that contain species of interest
- Videos clipped to exclude
  - Moving down and up through water column
  - Periods of turbidity/technical issues

Flora workflow & beta-test

1. Is substrate mostly hard, soft, or mixed?
2. Is vegetation present?

- 52 sites in beta-test, 1 video per site
- 9 screen captures per site/video
- Required 15 classifications of each subject

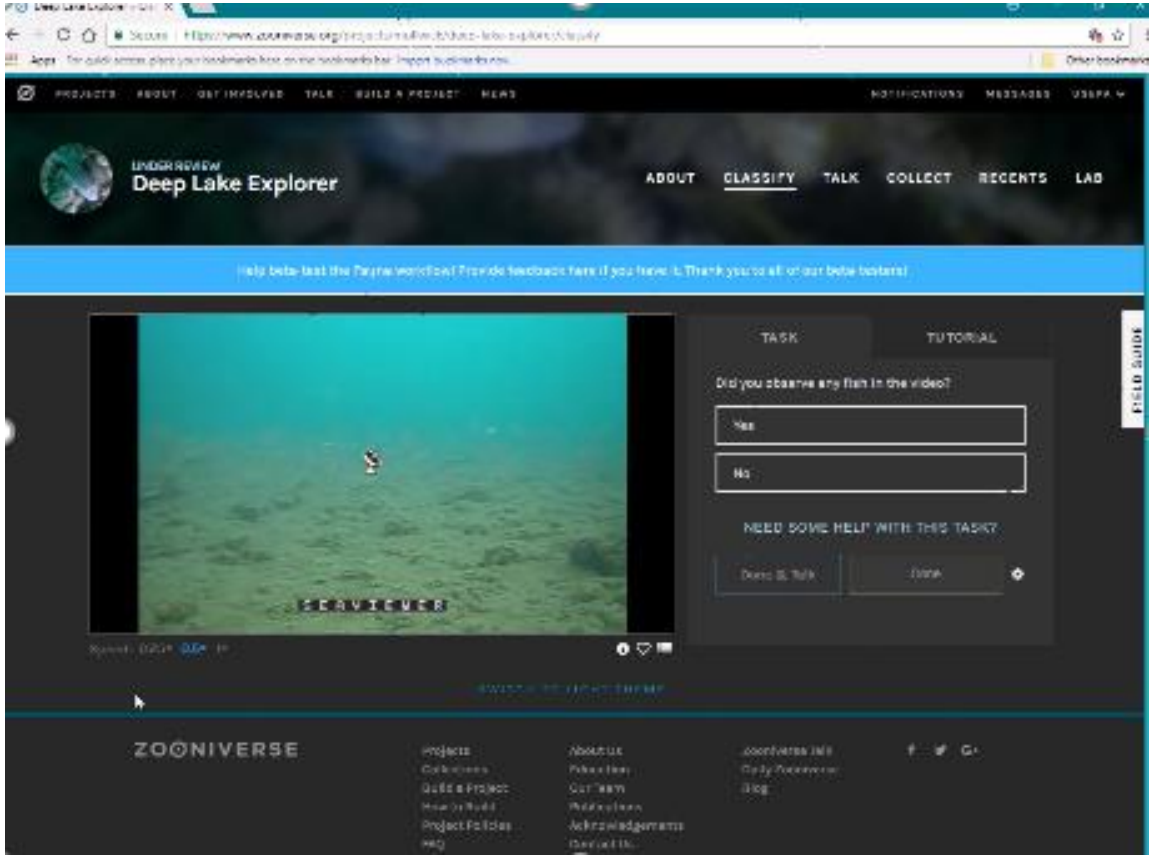
- Sent to testers in April, 2018
- Completed within 2 weeks



Fauna workflow

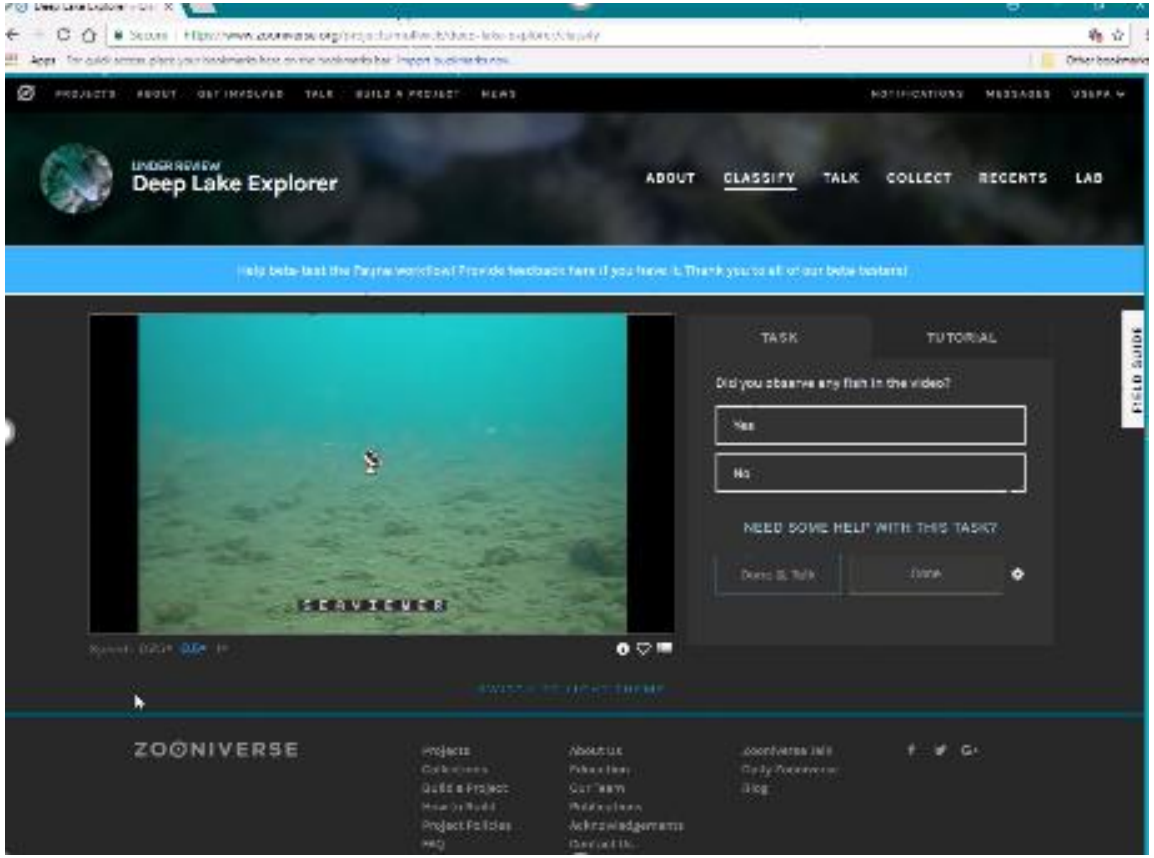
1. Do you see fish?
2. If so, are they round gobies?
3. Do you see invasive mussels?

- 15 second clips - several clips per video/site
- 52 sites/255 clips to begin

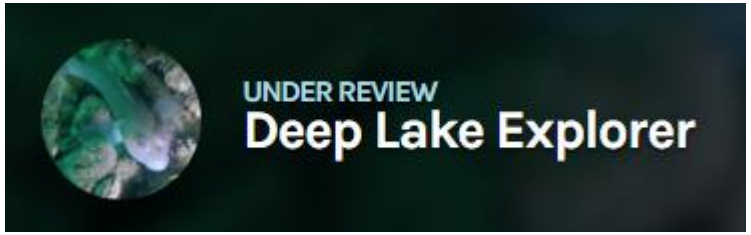


Fauna beta-test

- Sent to testers in April 2018 with Flora workflow
- Stalled at 10% complete. Potential reasons:
  - Takes more time to analyze clips
  - Drop-outs due to challenge interpreting video – poor quality or long loading time
  - Large number of clips to analyze
- Modified beta-test:
  - Reduced number of clips - 28 sites (150 clips) → reduced diversity of sites
  - Reduced number of required reviews to 10
- Re-sent to Zooniverse testers in July, 2018
- Completed within the week it was sent out.



Beta-test user feedback



Technical feedback

- Browser matters
  - Issues in Internet Explorer
  - Microsoft Edge or Chrome both worked well

Workflow feedback

- Users concerned if they did not observe fish/dreissenids –
  - Inclusion of blanks engages users longer (Bowyer et al. 2015)
- Users were unsure how to account for seeing a fish or other animals in the Flora workflow
- Requested an “I don’t know” option
- Poor video quality
  - Videos’ water clarity, lighting, camera movement, field of view, etc.
  - Files compressed for posting on Zooniverse (requires mp4) & resizing to fit in viewer
  - Limited ways to address – Videos must load quickly to maintain engagement

Updated tutorials, Help, FAQs, and other supporting materials

Flora workflow beta-test results

- Identify minimum threshold for percent agreement of users & impact on N
  - Cutoff of 75%, ~35% of videos dropped
  - Cutoff of 80%, ~50% of videos dropped
- 73% user agreement threshold
  - Maximizes expert agreement and minimizes the number of videos eliminated
  - ~1/3 of sites are still eliminated
  - Still relatively low agreement

Expert comparison

Experts are the "control", but agreement between two experts also ranges 70 – 80% for each attribute

Substrate Type			Vegetation Presence		
User Agreement Threshold	**N**	**User-Expert Agreement**	**User Agreement Threshold**	**N**	**User-Expert Agreement**
none	52	69%	none	52	75%
73%	38	74%	73%	37	89%
80%	28	75% (maximum)	85%	27	93% (maximum)

Fauna workflow beta-test results

- Better user agreement for fish presence than Flora
  - Dominated by agreement about *absence*
    - Low Ns for videos *with* fish/gobies/dreissenids
  - Beta-test videos should have mostly videos containing species
  - Only 28 sites included in beta-test
- Expert analysis comparison – Used all subjects (no user agreement threshold)
  - 100% user-expert agreement on fish presence (N = 4)
  - Users identified round goby in the one video where experts identified round gobies.
  - Users (with at least 73% agreement) missed dreissenids in 4/6 sites where experts identified dreissenids.
    - Mostly low dreissenid cover

User agreement threshold	Percent of all clips eliminated		
	Fish Presence	Round Goby Presence	Dreissenid Presence
73%	4%	2%	12%
80%	8%	5%	17%
90%	17%	13%	43%

←Clips reaggregated to represent site

(Experts analyzed entire video not clips)

Beta-test summary

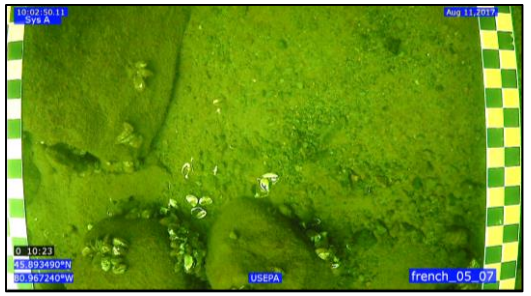
- Obtained approval for EPA to use Zooniverse
- Developed Deep Lake Explorer
- Learned how to set up beta-test and expert analysis
- Lower agreement among users and with experts than ideal, but we did find users could identify fish, round gobies, substrate, vegetation in **high quality videos**.
  - Video-quality is critical
- User agreement and agreement with experts varies by attribute

Next Steps based on lessons learned

- No public launch with 2010-2015 data

- Obtained Innovation Office funding for “Phase 2” to improve approach utilizing high quality video
  - Adapt workflow for a two camera approach
  - Beta-test and launch in 2018
  - Goal of improving data collection and analysis workflow in preparation for 2020 NCCA

High quality video



- Improvements:
  - 2 cameras – downlooking and oblique
  - 1080p resolution
  - Improved lighting
  - Stable frame – fixed field of view, stable camera
  - Ruler for scale
  - Affordable for implementation at NCCA scale

- Datsets:
  - ~ 75 sites in Lake Huron's nearshore in Aug 2017
  - ~ 100 sites in the Apostle Islands National Lakeshore in Sept. 2017
  - ~ 70 sites in the Niagara River in July 2018
  - Hard-bottom sites (number of sites TBD) in Lake Ontario's nearshore in Sept. 2018
- Long-term focus on Great Lakes-wide NCCA assessment using video
- We can test/demo video in your lake – get in touch with me

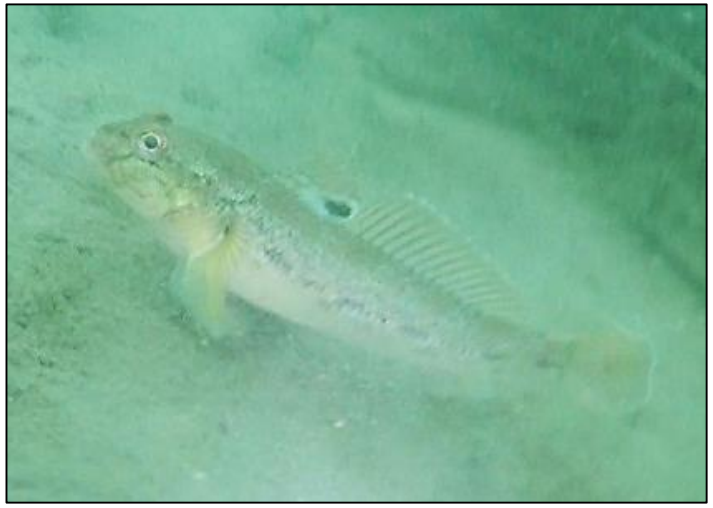
Phase 2 Timeline

- Sept. 2018 - Closing out Phase 1
  - Report to innovation office with results from development and beta-test
- Sept. 2018 – Collect videos on Lake Ontario
- Oct. – Dec. 2018 – Develop new workflows, clip/process/upload videos, expert analysis of new videos
  - Get your input – by email or next call in early Oct.
- Jan. – April 2019 – Beta-test new workflow, review/compile results and prepare for public launch
- May – Sept. 2019 – Deep Lake Explorer goes live
- Fall 2019 - Reporting

Workflow development

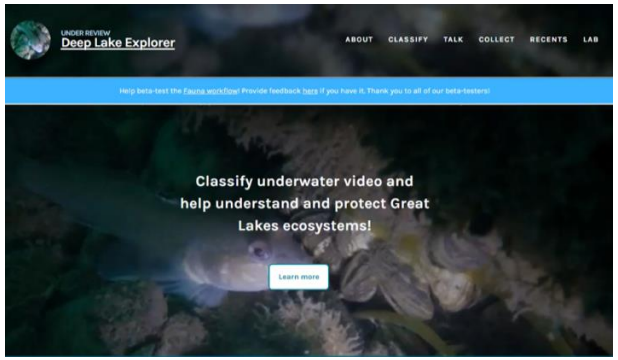
- Priority data needs – Reevaluate?
  - Dreissenids and round goby presence
  - How invasives relate to basic habitat characteristics
  - Keep it simple: presence/absence

- Requesting your input…



Workflow development considerations

- How to address 2 camera system?
  - Evaluate together to get full picture of site
  - Independently to compare methods
- Clip length?
- Type of data for each camera
  - Substrate: dominant/secondary / Point counts for size classes / Hard/soft
  - Vegetation: Presence / Percent cover
  - Dreissenids: Presence / Percent cover
  - Fish: Presence
  - Round Goby: Presence / Counts / Counts for each 10s interval into video
- Forming workflows: Stage easy/challenging / Flora/Fauna again / All together
- Expert analysis will mirror Zooniverse analysis
- Beta-test dataset will focus on videos containing fish and mussels



Thank You!

- Action items:
  - Contact me to explore potential to test/demo video in your field/project area
  - Provide feedback via email or on next call about how to design future workflows
    - Data priorities (slide 15)
    - Priority dataset/s to incorporate (slide 13)
    - How to setting up workflows (slide 16)

- Next stakeholder call in October – date/time TBD

extra

Beta-test Analysis/QA Methods

- If a single user analyzed same subject twice, we removed both classifications if they were different or removed all but one if they were the same
- Summarize multiple users analysis by subject
  - Find percent agreement among users
  - Evaluated target agreement – hoping to aim for 90% agreement or similar
  - Eliminate subjects not meeting target agreement
- For Fauna workflow, reaggregate subjects (clips) to videos (sites) to summarize attributes by site
- Compare results to expert analysis
  - Expert analysis was done by video not by clip – in future should be done by clip