

ASIS&T 2011 PLENARY SESSION

How to Identify Ducks in Flight: A Crowdsourcing Approach to Biodiversity Research and Conservation

by Steve Hardin

EDITOR'S SUMMARY

Volunteers engaged in crowdsourcing are providing valuable information to help understand bird biodiversity and migration. Attendees at the 2011 Annual Meeting heard Steve Kelling, director of information science at Cornell University's Ornithology Lab, explain how humans play a critical role by providing knowledge used to identify, differentiate and track birds around the world. Kelling's eBird project is a global citizen science project involving birders who report bird sighting locations, numbers, dates and time. The My eBird web pages engage birders, supporting collection and sharing of personal experiences. Smartphone apps and data-backed dot indicators on Google Maps provide extra technology support for participants. The end goal is broad scale analysis with fine resolution about species occurrence and biodiversity, applying the kind of information that only humans can provide.

KEYWORDS

biology
crowdsourcing
volunteers
data collection
research methods

Steve Hardin is the interim chair of reference/instruction at Cunningham Memorial Library, Indiana State University, in Terre Haute. He can be reached at Steve.Hardin@indstate.edu.



Making good use of volunteers who enjoy bird watching is leading to the creation of a worldwide data network to help conserve the avian population and provide important environmental indicators. On October 10, 2011, Steve Kelling, director of information science for the Lab of Ornithology at Cornell University (stk2@cornell.edu), outlined some of his ongoing projects for the second plenary session at the 2011 ASIS&T Annual Meeting in New Orleans.

Kelling showed slides of Canada geese in flight, followed by a slide of a mallard duck, which is identified by its green head, yellow bill and orange feet. He then showed the common goldeneye duck, which is identified by the white spot in front of its eye. Female ducks all look different from males. With 100 species of ducks in the United States, the estimated 70 million bird watchers in the country really have 200 types of ducks to identify.

Visual recognition software has not been developed to identify ducks at the species level. Only people can do that. Birders throughout the world record their findings, and Kelling and his colleagues would like to organize these expert sensors into a network that will permit identification throughout the continent, hemisphere and globe.

Indeed, as Edith Law and Luis von Ahn discuss in their book, *Human Computation*, there are many things people can do that computers cannot. You have probably encountered squiggly characters you need to identify in order to log on to certain secure sites on the web – characters optical character



readers cannot identify. CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart, www.captcha.net/) is one program that will generate such strings. Another example exploiting human computation is Foldit (<http://fold.it/portal/>), a game that combines computers with human skills to identify ways that proteins can be folded,

while Galaxy Zoo (www.galaxyzoo.org/) is a similar project where humans can help classify galaxies by shape using a visual sky survey.

Kelling's project is called eBird (<http://ebird.org/content/ebird/>).

Members of the public are the sensors. It addresses several questions:

1. What tasks cannot be performed adequately by machines?
2. How do we leverage the complementary abilities of humans and machines to improve our ability to collect and process information?
3. How do we engage in developed communities of practice to collect sufficient quantities of information and work with it?
4. How do we deal with messy, noisy data?

eBird is a global citizen science project. Kelling said the project partners with some 50 people and organizations to run things. They ask people to do very boring things:

1. Tell where they bird.
2. Provide date and effort information: follow standard protocols the birders have to use, such as time, date and time spent birding.
3. Report what they see or hear. There is a long list of species, and they report an estimated number for each one.
4. Consider whether they are reporting all the birds they observed and identified. This question helps birders determine absence as well as presence.

At the time of Kelling's presentation in October, there had been 1.7 million checklist submissions in 2011 from every country in the world. Indeed, the project collects data on 99% of the 11,000 species of birds.

There is a real time video of checklist submissions. eBird collects about 5000 checklists and 75,000 observations each day that all go into a single

standard database. The project uses the data to engage the public – people love to see their points on the maps.

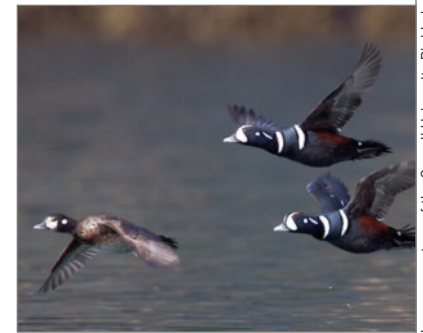
Kelling said they spend a lot of time understanding their audience. Some are nerdy birders complete with floppy hats and books. But they are also engaging with a younger community of really savvy users of the Internet.

The project has developed a set of pages, “My eBird,” on which users can keep track of observations they make. There are top-100 observer lists for states, counties, regions and countries. They hold little competitions within communities. They get people to repeatedly submit observations from the same site, important for bird observations. They also provide opportunities to view and explore the data. For example, Kelling was in Hawaii where he saw sooty shearwaters, which we seldom see because they spend most of their time out at sea. He went to the database to see how many were showing up. By creating an area where birders can create their own stories, the project invites participation. Simple histograms can tell people when to go to see particular species – a very powerful tool for wildlife managers. For greater mobility, eBird is creating apps for smartphones. Managers also put dots for sightings on Google Maps. Users can click on a sighting balloon and get all the information an observer has reported from a site over an extended period. These kinds of tools have built the community.

Now, Kelling said, they are collecting more data every month than they did in the first three years of the project. In 2011 eBird contributors volunteered more than 1.3 million hours collecting bird observations. The project uses 400 volunteers to vet the data and help ensure quality.

There are many observational data sources, including satellite imagery and LIDAR. But the only sensor that collects anything about biodiversity occurrence is human.

There are other similar projects. For example, the National Science Foundation's DataONE – Data Observation Network for Earth -



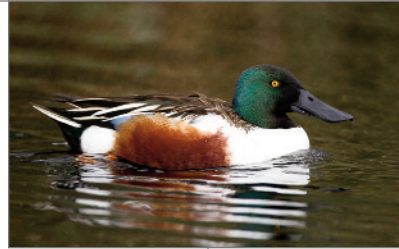
Images courtesy of the Cornell University Bird Lab

(www.dataone.org) program organizes observational networks on the environment. NEON – the National Ecological Observatory Network (www.neoninc.org/) – collects intensive data over a small area.

Data interoperability is a big challenge, Kelling said. How do we put together the hundreds of thousands of individual location observations of birds provided by human volunteers and the aerial mosaic raster landscape image provided by satellite? The eBird project does database processing and uses Darwin Core to integrate data. The Darwin Core is based on standards developed by the Dublin Core Metadata Initiative (DCMI) and should be viewed as an extension of the Dublin Core for biodiversity information

(rs.tdwg.org/dwc). The database created from data from around the world is published annually as a reference data set. The project provides access to the data through maps on the Avian Knowledge Network site (www.avianknowledge.net). The goal, he said, is analysis at a broad scale with fine resolution to understand species occurrence and biodiversity. He showed a map of wood thrush occurrences over the eastern United States. These kinds of detailed maps, he said, allow us to really understand the data.

The Secretary of the Interior issues the “State of the Birds” report (www.stateofthebirds.org) to present broad messages about the status of birds in the United States. Birds are bioindicators. Last year, the report focused on how birds use public lands across the country. Kelling said he and his colleagues worked with the U.S. Geological Survey’s data on public land, overlaid their species distribution maps and generated a metric of ownership and occurrence of species in particular bioregions. It was the nation’s first



assessment of the distribution of birds on public lands and waters. Researchers were able to drill down in the data and see which land management agency was most responsible for species in a particular land type – arid lands. They identified it as the Bureau of Land Management (BLM). But the problem is that the BLM’s primary goal is energy extraction, not preservation of species. Huge solar collector projects on public lands will have great implications for biodiversity.

With the Gulf oil spill in 2010, they generated four billion observations over three months and did a Google mashup to overlay brown pelican breeding sites with oil-spill-spread locations. It was a big effort, and actually, Kelling said, very few of these birds died in the spill –2011 was one of the best in terms of nesting success. They will develop a Gulf environmental health report card – similar to Chesapeake EcoCheck (www.eco-check.org/) already underway.

Where are they going with this project? Kelling said researchers will try to forecast bird migration, using NEXRAD weather radar to keep track of birds migrating. They can get estimates of bird populations using NEXRAD. They can also obtain information on what kind of birds are involved by recording their calls. He played recordings of different kinds of birds flying over his house at about 3:00 a.m. People have a hard time identifying species from those sounds, but in this case there are machine algorithms that can do it with 97% accuracy. They will use different kinds of information that indirectly tells them how birds migrate. They hope the information will allow them to protect birds by, for example, forecasting when birds will migrate, helping migrating birds avoid colliding with skyscrapers or determining when wind turbines on migration routes should be turned off. ■