I Spy with my Little iNaturalist: Open Data, Stalkerware, and Citizen Science

Dr. Kit Heintzman

>> So, welcome back, everyone. For GUADEC 2020. First day, first track. And we continue with Dr. Kit Heintzman with I spy with my little iNaturalist: Open data, stalkerware, and citizen science. Please.

We don't hear you. What do you have at the bottom of your screen?

Okay. No, not yet. You are showing us as just listening. You should choose -- you have to click on the bottom of your screen, there's three buttons and one is kind of showing -- showing that you are being...

You're still muted.

Dr. Kit Heintzman: What about now?

>> Yes, excellent! Thank you.

Dr. Kit Heintzman: Great. Thank you. All right. Hi, everybody. I am really excited to be here. Depending on what time zone you are in this morning, this afternoon, this evening, or perhaps this night time. I am coming to you from territory in Canada, something that's important to acknowledge as a longstanding reconciliation process. So, the talk I'm going to give today is I spy with my little iNaturalist: Open data, stalkerware, ands is science. As the title suggests, there's references to stalking inside of the talk. So, it's important for people to stay in touch with their own reactions and needs around this information.

And one of the great luxuries of the digital conference is that it's so easy to bounce if you need to step away from this conversation. There will only be one moment where I discuss actual stalking as opposed to if material can be manipulated for stalking. And I will give a warning and it will be brief. If you want to leave for that, bounce for a minute or two.

So, I'm going to be talking to you about iNaturalist. A free software project that tries to turn everyday people into what's called citizen scientists. And how I, as a historian teaching a history class, and came to think about how you would use iNaturalist ain't and how I would use it with my students in a very different way than it is typically intended. Which is to teach them things about open data and human technological interfaces.

So, I'm gonna start by talking a bit about iNaturalist. What I have put on screen is what iNaturalist would call the sort of total representation of the number of observations that they have. So, the kinds of things iNaturalist does is in its soul it's a big data project meant for people overwhelmingly in the environmental sciences and biology. Broadly construe. What it does is users, users can be biologists, zoologists, whatever. Can upload their own data in a spirit of open data to share it with the world. And it's very easy to access for everyday users such as myself or humanities students can come in and use this material too to offer representations of the natural world.

A part of why iNaturalist wants this work is that it's in a long tradition of natural historians actually crowd sourcing information. And it does have a lot of scientific of weight behind it. So, as of 2018, more than a hundred research articles cited data from this website. It's credited with the discovery of a kind of poisonous frog in the Andes, if you need some nightmares, that had never been heard before. And there's a commonly told heartwarming story of a high school girl who discovered a kind of sea slug that had never been discovered in the region before and came to publish on it with a scientist.

So, this is really sort of two functions of this. One, collect data. And two, get everyday people interested in science and the environment and the ecology and in the natural world. This system absolutely requires geolocation for this data to mean anything. So, the way the data is collected and verified is someone takes a picture of something like a horned spanworm moth. Takes a picture of something like that. And then

uploads it. Provides a geolocation if they want it to be useful for scientific purposes. And they can hold that back if they choose too.

So, all users do have that freedom. But there is a tension in this project of wanting users to present this information freely and as usefully as possible and thinking about what kind of information may actually -- may users be sharing that they haven't really thought about or considered that they have been sharing. So, iNaturalist boasts more than 44 million observations as of this morning. The number changes daily. And more than a million observers. And when I was thinking about integrating this into my teaching, I was really interested in a history class thinking about the relationship between humans studying the natural world and how that changes over time. And then also, thinking about humans interacting with technology and specific data.

So, this talk brings together two iterations of myself and my online documentation. One of which is the ways in which academics are required to have pretty extensive documentation of where I am at a lot of moments. I'll be sharing a fair amount of personal information about my own movements as I've tracked them through the software. But everything I share is something that someone could learn from my CV. So, I don't feel particularly vulnerable doing it.

So, there's the kind of general imperative in our world to have an online identity which presents a lot of information. And so, we're used to this and it's really normalized. I did, when starting this -- before starting this project with my students, did what any good teacher should do, started tinkering with the program first. See how it works, why it works, what kind of community I'm inviting my students into. And the first step in this project for me was looking at other people's data and uploaded information. So, seeing how other -- what's a typical way to use this platform. And I was surprised how much I could learn about the everyday users. With relative confidence. So, things like being able to tell someone's exact home address and also that they're the owner of an exotic pet. That kind of information is really easy to find if you are adequate at research. And by research I mean I am probably the least technologically capable person in this room as evinced by the microphone problem.

I've researched like a historian does. I sleuth through information. This was not done in any kind of high-tech way. It's really low-tech research. And beyond my capacity, that's also because most people don't need to worry about being stalked by trained hackers. Most stalkers happen -- most stalking happens by people in our lives. I do have a third part of my identity, which is not online which is that I am someone who has been stalked by proxy.

What that means in terms of understanding stalking culture, and this is the sort of one description of stalking that's coming. What it means to be stalked by proxy is that when someone is stalking one person, they tend to stalk other people in their lives. So, in the context for me, this was the stalking of my mother where the person stopped me so that he could intimidate her through threatening her children.

And this is a pretty normal manifestation of stalking. This experience had nothing to do with cyber stalking for me. Because it was in the '90s and I'm from Canada, and we didn't have the Internet back then. I didn't have an email address. The person who was doing this stalking had probably never touched a computer. However, I got Internet access the next year. And so, my entire dynamic with technology has been shaped around what it means to have privacy invaded and threatening. And threatened.

So, I wanted to design this for my students that it's going to touch on these issues. And the first thing I start testing is whether or not -- what kinds of information you can see about humans given that humans do document themselves as species and organisms on the iNaturalist website and there's no rule against it and there doesn't necessarily need to be.

So, in contrast to the slide before, where there was a picture globe and a dense representation of observations as seen through, hear when I type in the species human, the page comes up blank. That is by

design. It's by design for user privacy so that the only species that is protected from any kind of data mining are humans -- is humans. There are other kinds of protections that are put in place from a more ecological framework. And this has to do with part of the difference between design taking place between data scientists and between ecologists as opposed to people who are more used to seeing humans and human-centered issues. Like the sociology of stalking or like the representations of human in -- humans in natural history.

This is a slide to indicate that there's a fairly simple user error that means that humans do actually occasionally show up in this search. And I want to just give a sense of the degree of information that is shared when this user error happens. And it's typically not user error by the person whose uploaded the image. The error happens via someone else. And so, through a user error that I artificially constructed in order to test and be certain that this is how it worked, I made it possible for one of my human entries, which is part of my body, and there will be no representations of other people's data in this process. Everything I used was my own. Even if I studied other people's in order to understand it.

So, this was a picture of my hand. And what I really want you to see on this teeny tiny screen, or at least now as I explain it while it's remitted as evidence, is there there's a latitude and longitude for where this picture was taken because my geo privacy settings are in general open. And there's an accuracy of 65 meters. The other thing that I want you to notice on this slide, is that I posted this at the moment that I took it. And it manes that I have told someone my exact location with a small margin of error.

And while these are all things that are controllable, they're only controllable if the people using these tools you said that. And there's no reason to believe that the everyday person using this does understand that. Other animals have distinct kinds of protections. So, like I was saying, there's a system in place to protect endangered species from poaching. Which is that when one is -- does a search for black rhinoceroses, and this is another way that they represent data. So, you type that in, see an image of the globe and you see where the observations are.

And in that context, they give you a latitude and longitude for the endangered species with a margin of errors for 185 kilometers. There's a different protection for those that are considered to be threatened that are not given to people. The user did not obscure that data. But cataloging it as a black rhinoceros in the environment, it became automatically protected in a way that humans are automatically protected in very different ways.

So, when you're thinking about this kind of material and you're using it and designing it because you want to bring it into a class room, one finds one's self in a real ethical pickle. I don't know if this slide is gonna work. It seems very still. There's a really great picture of a pickle.

Someone finds themselves in a real ethical pickle when you're thinking about, okay. So, I think this -- a real believer in it. And I would like my students to think with it because the more I see the tension between the way humans are protected and the way animals are protected by this program, the more interesting it becomes for a historian who is interested in natural history and the production of scientific knowledge for its implications for humans.

But I'm coming from a position of authority with my students. And anti-oppressive pedagogy requires constantly thinking about that authority and the potential for coercion, especially in digital infrastructure. Whenever we're engaging with technology, I am in a classroom. I attempt very hard to privilege open source materials precisely because it immediately brings me into a conversation with students about the ethics of the technology that we'll be using.

So, I have very complicated feelings about not the ethics of the project, but the ethics of some of the implications of the project. So, what do you do once you've sorted these things out as a teacher? You can ditch the project. You can decide that the stakes are too high to do it. You can decide that your students'

lives are already all over TikTok, doesn't matter. I don't recommend that decision. Or you can decide radically rework the assignment. Which is ultimately what I tried to do.

And the rest of the talk is gonna focus on what that was like. So, instead of doing an assignment where my students would download the app on to their -- into their personal devices or sort of at the more protected category create an anonymous-style account on a computer-based web browser that would be separate from their own personal devices. Instead of having them do that kind of work, I decided to offer a kind of a assignment that would require them to think about these issues without imperative to participate.

So, students were invited at the end of the assignment to make a decision about whether or not they wanted to participate. And none of them did. I don't necessarily think this is because they were scared off the project. It in no small part has to do with the fact that students don't want to do extra work if they're not doing extra credit. And I can't provide extra credit for something I perceive to be potentially dangerous.

One of the other option ways to do this that I had considered was having them sign a waiver. That's really common. It's basically the equivalent of asking someone to sign an end user license agreement. Students would read it. So, something where they had to show me they understood this material was really important. There were different sets of questions. A set of kind of mechanical questions based on the premise that in a history class you should know something about the thing that you're talking about before you talk about it.

And this is actually one of the most basic things I wanted my students to learn. Do research before you make a claim. So, I wanted them to know what the project's intended purposes are. Digging data for environmental sciences and get the everyday person interested in science. What are the rules of the platform? And I had very specific questions about that. I asked how it was made. So, do a little digging into not just what its intentions are now, but how did the project evolve. And then I wanted to run a series of questions that would touch on the kinds of privacy issues I considered. I took the same model with this in the class room as I did with my representation of the hand here. Which is that I decided the only person's privacy I could ask them to invade was mine and to give them very specific questions about that.

So, one of those questions -- so, this required a lot of vulnerability on my end. And this was ultimately an ethical tradeoff that I needed with the student. Which was deciding that if I had considered them being that vulnerable and wanted to think with this project, and have already made myself vulnerable by participating in the project, I may have -- I may as well present myself with that information.

Present them with my information. Most of them gave text-based answers. You're training students as researchers, they're not great yet. So, the most significant part of the assignment was not things that they found about me, but were things that I was able to take up with them. So, lots of students notice that I'm incredibly geographically unstable. But they didn't understand the degree of it until I showed them charts like this. And by degree of it, I don't mean the geographic instability. I mean, how visible that was on the platform.

Then I ran an exercise while taking up the assignment with them where I shed them a tight chunk of a map from my being present in Pasadena where I collected all of my observations over a period of time.

>> Four minutes to go.

Dr. Kit Heintzman: 4 minutes?

>> Yeah.

Dr. Kit Heintzman: Great. When you ask your students to look at a map like this and see what they learn about you, they say things, like, you take a lot of pictures of dandelions. Which is precisely what people using this think they're doing. And then I show them a map that contains all of the same information

related to these images which is publicly available. Which is that the time and date at which all of these were taken. And so, while they didn't have the energy to dig deep enough to find these kinds of things, and they did find other really cool things. There was a sense of shock when they learned that not only could they learn things from the way that I write in the profile, or from the pictures that I've taken. But even though I primed this on concerns about privacy and the importance of thinking about in that, it wasn't until showed them this kind of information. I'm walking in a similar environment over a long period of time that they sort of got a sense of what that means in terms of privacy and in terms of moments where someone is vulnerable to someone deeply paying attention to that privacy.

I also crowdsourced from my students moments where I invaded other people's privacy. In addition to asking them to identify things about me that had nothing to do with natural history, I asked them to identify at least one moment in my profile where an observation image captured something about another living person. So, in what way have I potentially given information about someone else that they did not expect some rando to be publishing on the Internet? And my favorite answer who comes from a student who says this picture of a wild turkey includes a car that has clearly been in the driveway long enough for the snow to accumulate and melt. Providing evidence for the person to whom the license plate was registered was living at the house that was shown because it does give the exact address of where I am with precision in December of 2019.

>> One minute to go.

Dr. Kit Heintzman: As I took this, when you're taking photos for this kind of a project, I don't notice everything in the background. So, I had to then go in after the student pointed this out to me and redact it. The two trade-offs that my students located around doing this kind of work was that the pros of the service like iNaturalist is that you have more information. And we live in a world where we fundamentally believe that more information begets more information begets better information. The cons is that you can't control how that information gets used. And you might not want people to have more of certain information.

And I'll leave you -- before questions -- a few thoughts on brainstormed solutions. Most are more practical than the ones I put up. But they are relevant to the audience. One is good science requires input from humanists and social scientists who are trained to be attentive to an entire different sets. Privacy statements are basically useless for the everyday user. A privacy statement tells me something is happening. I may not believe it. But doesn't give me anything in the way of an education about how data works. Part of what this means is that you have -- as possible when users lack computer, Internet and data literacy. And I think that that is something that coders and software engineers and humanists and ecologists can and actually should collaborate on. Thank you.

>> Thank you, Dr. Heintzman. Any questions? Okay. I will read a question from Britt. I have a question, if privacy statements are useless, what we can do instead? Also, what do we do about apathy towards your own privacy?

Dr. Kit Heintzman: Right. Two fantastic questions. I'll go in the reverse order. See apathy with privacy is one of the things I found most. I think of generational, but also acknowledge that I have personal experiences that make this different for me in terms of my relationship with technology. So, any time I introduce iNaturalist, I present a few options. There's a low-tech opt out option. There's a way to design with students an entirely different way of doing this same assignment that doesn't require any online activity. And one of the things I find fascinating and I've never had a student take me up on it. And I do frame it in the context of privacy whenever I introduce the software, what I've been able to discern about the privacy. And where the gaps in knowledge are and where they can attempt to find them. They are not computer scientists. And even pointing them to the resources is not particularly useful. They experience that.

They also are incredibly indifferent. And this was the thing when I first started working with students on this work, I found really surprising and I didn't manage perfectly the first time. Because the first time all the students were like, cool. Okay. Whereas after that first ultrasound, I've spent more time trueing to explain the stakes of that. And in this iteration, none of my students opted in to use it. So, some of that has to do with balancing apathy they may experience as well as encouraging them to be less apathetic. But also in this context, not using the software was less work than using it. At least in terms of the assignment's objectives. It probably would have been easier to snap photos and say something brief.

The other thing about privacy statements being used is I fundamentally think we need statements that go beyond questions of how does your data get used by data brokers? How does the information get stripped from you as a part of participating in any kind of online space? And focus on how the social aspect of participating in these environments is a privacy issue as well. Privacy issue isn't just the stuff behind the screen, inside the computer. It's also the reception on the screen. And I think we need to consider it if we care about privacy. We need to consider that essential to our conversations with the general public.

And I think iNaturalist has a pretty good chance to do that.

>> Okay. I think that's all the time we have for questions. Thank you, Doctor. We will move on with Emmanuele Bassi after one minute. Thank you. So, next on this -- on this track we have Emmanuele Bassi with archaeology of accessibility.