Scott Hanselman: This is ACM Bytecast, a podcast series from the Association for Computing Machinery, the world's largest education and scientific computing society. We talk to researchers, practitioners, and innovators who are at the intersection of computing research and practice. They share their experiences, the lessons they've learned, and their own visions for the future of computing. I'm your host today, Scott Hanselman. Hi, I'm Scott Hanselman. This is another episode of Hanselminutes in association with the ACM Bytecast. Today, I'm chatting with Yaw Anokwa. He's the founder and CEO of ODK. How are you, sir?

Yaw Anokwa: I'm doing very well. Thanks for having me, Scott.

Scott Hanselman: So ODK stands for what?

Yaw Anokwa: Originally standard for Open Data Kit in the sense that it's an open source data collection toolkit. These days we just call it ODK in the same way that people don't call IBM International Business Machines anymore. Right? It's just become an acronym that means data collection.

Scott Hanselman: Okay. And on its face, it is powerful offline forms. And it's interesting because people pick their taglines and it matters. Those taglines matter. Every word matters. You didn't say powerful forms to collect data. You didn't say data collection toolkit, you said powerful offline forms. Why is that important?

Yaw Anokwa: Well, because ODK's value add it's primary differentiators, the fact that the forms themselves work entirely offline and they're not sort of simple forms like you see in Survey Monkey or Google Forms. They're really forms that end up looking like mini applications. So the forms can collect obviously text and numbers, but they can collect GPS coordinates, they can collect pictures, they can scan barcodes, they can have complex logic associated with them. So there really are powerful forms and they work entirely offline.

Scott Hanselman: Where they always powerful forms. Because I understand that you started this as a software engineering intern at Google. Did you start with text boxes over data and then when did you get into complicated data formats?

Yaw Anokwa: We started complicated. So the story there was that I was a grad student at the University of Washington in Seattle and my advisor Gaitano had sabbatical at Google. And so we went there to build a powerful offline form tool because it was just right around the time that Android was starting. And so as researchers we thought here's an awesome opportunity to build something on an entirely open source data collection, the operating system, mobile operating system, and do something that you couldn't do on a basic J2ME phone or a Nokia phone at the time. So we went in wanting to build powerful offline forms.

Scott Hanselman: And offline meaning it's going to put it on local storage of the device that you're on. And then when it docks or when it gets connection or when does it sync up and is it a bidirectional sync or a one way sync?
Yaw Anokwa: Yeah, it's a bidirectional sync. And so the typical use case is that you have a data collector, or in our industry we call it an enumerator, somebody who's out in the field, usually remote place with no internet connection and they're maybe doing a household survey, they're going house to house and they're filling out forms. And when that mobile device sees either a cell connection or a wifi connection, it sends all the forms that have been finished up to the server. And if there are any updates or new forms, those get downloaded automatically in the background. And we have users who are maybe in the Amazon rainforest doing community based force monitoring. They're out there for months at a time collecting data. And whenever there's a connection, the data goes up and new forms come down.

Scott Hanselman: And I'm imagining simplistically people walking from door to door doing censuses, but that's humans talking to humans and helping them fill out forms. But as you said, you're doing surveys, I'm assuming that you're not just talking to humans, you're talking to devices.

Yaw Anokwa: It's primarily humans in the sense that the typical use case is a household survey, but it varies. So for example, the surveys don't have to be about people. So we have examples of users who are documenting the health of forests or farms or users who are monitoring animals. The common case though is there's a person who is interrogating asking questions about a thing and collecting data about that and then sending that to a place where the data can be acted on.

Scott Hanselman: Help me understand the software architecture here. There's a collector and there's a server, but are you a giant world database? Are you like a power BI?

Yaw Anokwa: No, not so much. So we provide software that people can install either locally on their own machine or we provide a managed cloud hosted service. And each customer or each user, they set up their own machines or have their own account to collect the data just for them. So the data doesn't flow into one big database, it flows into their database, it's their data and they can do with that data, whatever they want to do with.

Scott Hanselman: That's pretty cool. I mean, the idea that this is for the people, by the people, there's nothing underneath it. There's nothing hidden or sneaky in it. It's power to the people in the literal sense embodied by software.

Yaw Anokwa: Yeah, that's exactly right. There are some platforms where essentially it's free, you're collecting the data, but behind the scenes the data is being used and sold and marketed. In this case, ODK's and open source projects, it's targeted exclusively towards social impact organizations, folks who work in public health, folks who work in climate monitoring, in election monitoring, and it's their data and they use that data for critical decision making. And because a lot of the data is sensitive, sometimes it's government data, sometimes it's personal data. We want to make sure that it's clearly belongs to the people who are collecting it. And there's nothing sneaky or hidden about it.
Scott Hanselman: In the early 2000s, you had your own consulting company, you did some work at Intel working on some custom hardware, then it looks like you started doing some volunteer work as a software engineer. Help me understand where you went from the classical working for a big company doing stuff to public goods.

Yaw Anokwa: Yeah, I suppose it happened, I was a grad student at the University of Washington and I saw a talk by Neal Lesh. I remember this very day. And he was talking about how he was sort of a wandering do-gooder, he was nomadic, he was wandering around East Africa and he gave a talk about how he was fixing computers in hospitals. And I thought that was so compelling that I talked to my advisor and I said, ... At this time, I think I was building NFC hardware for Intel as an intern. And I thought, I just don't think that that's the best use of my time. So I volunteered. I actually went to Partners in Health in Rwanda. They were setting a medical record system there, and I volunteered there for six months. It was mind blowing. I really saw that with a little bit of technology you could change the lives of so many people.

And so what I did there was I set up a medical record system and I saw that that system, electronic medical records could increase the quality and the scale of HIV and TB care. And so having seen that, I thought as a computer scientist, well, can we make this more abstract? Can we make this more usable for folks who don't have to set up a medical record system? And so the general problem of taking a paper form, which is ubiquitous in social impact, digitizing it and giving that data back to the people who are collecting it, that's what started me on that journey. And so yeah, I gave up on the ubiquitous computing and the NFC stuff and I thought, well, here is something that is in many ways so simple but can make a huge difference to folks all over the world.

Scott Hanselman: Now I recognize that you're not done yet, but I'll use the word dedicate your life. Did you know that this was going to be it? Did you find your thing at this point?

Yaw Anokwa: At the time, I have to admit, I remember when Gaitano, my advisor at that time called me and I was working on some SMS based project in Tanzania. I was sitting on some beautiful beach having a beer and Gaitano called me on my cell and said, "Hey, I'm going to Google to work on this data collection project." And I thought, "I'm okay. I'm comfortable here. I'm exploring new research ideas. I don't want to go back to the states to go work in Seattle in some office." And so I told Gaitano no, but Gaitano was very persistent. He called me and my buddy Carl Hartung was also a grad student, called me and said, "Come on, you got to come back and stop messing about." So I returned to Seattle to Fremont and I started working on it. I didn't think it was going to work. I'll be the first to admit it.

I thought, I mean, giving at that time $600 smartphones to rural farmers in Uganda just didn't seem super compelling. But I got into it. We deployed the first versions of the software and it was just amazing to see how people took to
it and how valuable they found it. And that was about 15 years ago actually. And you were asking, is it my life's work? I must say it's become my life's work because I've seen the impact it has on people. And if you let me sort of expand on that. You are a software developer, a lot of people on this talk might be software developers as well and listening to this podcast, it's extremely hard to build software as you know, it's hard to build software, it's hard to build software that is useful in any way. It's hard to build software that works.

It's hard to build software that has an impact and it's hard to do all of that and make it open source and have this sort of positive impact. And so over time, I've come to realize that what we've done is really incredible. It's rare and as one of the founders of the project, it's sort of become my responsibility to protect this sort of unique thing that exists. And so it's a responsibility that I take seriously. The software's having such impact on the world that we have to keep it going. And at the end of the day, I enjoy it. It's just a fun project to work on to hear from users about the impact that they're having. And we just have to keep it going. So yeah, these days it's my life work. I get up every morning thinking about data collection. I go to sleep thinking about data collection. It's on a bad life.

Scott Hanselman: That's definitely not a bad life. I currently work at Microsoft, so we're making a lot of big software, but it's a bajillion people. And one of the things that I've found is that the engineers and the program managers that don't talk to customers, that don't spend time with them, they have an [inaudible 00:09:38], a malaise of just time to make the donuts and they just kind ... But if you sit with someone whose life you changed, it fundamentally changed. It sounds like you got that experience early, you talk to customers and you can't stop talking to customers.

Yaw Anokwa: Yeah, I always do it. Even to this day, I do a lot of the customer demos for people who are signing up for our cloud hosted service. And so pretty much every morning starting from 5:00 to 6:00 AM, I am in calls with customers showing them the software. So I do that a lot. I'm ODK from the beginning because this open source project has always had a dynamic and active community. So we have about, I think at this point, maybe about 14 or 15,000 people on our community forum. That started as a mailing list. So I talk to those folks all the time. We also have a technical advisory board with people from the community that I talk to every month or so. So I love talking to customers. It's how we know that we are on the right path and doing the correct kind of work. So there's no [inaudible 00:10:37] there.

Scott Hanselman: Yeah, you're addicted to it in the best way. You have to talk to the customer otherwise-

Yaw Anokwa: Yeah. I mean, otherwise how do I know what we are supposed to build next? And we have to do it also because the context in which we work in is pretty unique. So if you're at a Microsoft or Google and you're building software for
businesses or folks in the West, then it's very easy because you're functionally building software for people like you. But we are not building software for people like us. We're building software that works in very specific environments under certain constraints. And so the only way that we can build that software and make it relevant is by having an open lines of communication to the folks who are using it. And so that is key to our success.

Scott Hanselman: And that customer empathy seems to come out in your forms. I've looked at a number of different ODK forms and for some of them, particularly ones that are in developing countries, I noticed that, I don't know how to say this, the buttons were kind of big. It feel kind of chunky. And I was trying to understand why are these buttons so big? And then I started to apply some empathy to it and I think I know why and wouldn't you mind telling me.

Yaw Anokwa: Yeah, so there's lots of sort of UI affordances, user interface affordances in ODK, and those are things that have been learned over the years. And so the first time we deployed ODK was in Uganda in 2008. We were working with some rural farmers. They were in huts without a lot of electricity and a lot of them had callus fingers because look at your fingers, Scott, you're a programmer. Your fingers are soft and squishy. No calluses. If you look at a farmer's fingers, they're callused.

And so the way touchscreen works is through this capacity touch where electricity sort of passing through your fingers, a small charge and if you have calluses, that small charge doesn't work so well. And so what we do with ODK is that the buttons are big because our customers, our users, they need a big touch surface. Our buttons are big because most of them don't have glasses or corrected vision. And so they're all these sorts of affordances that ODK may look chunky to you because you're living a very comfortable life. But I assure you that millions of users that we have use ODK because of that chunkiness. It's because these big chunks are our affordances that enable them to collect the data that they need.

Scott Hanselman: I appreciate that. And again, no disrespect intended by the word chunky, just trying to express the sense of a, it's a swollen UI and it made me think about my father who would work with his hands his entire life. And I watch him struggle with his iPhones because he's got big hands that look like banana bunches that are covered in calluses and he jabs at the screen. But as soon as I just pumped up all the fonts on his phone, everything got better. All the hit points are better.

Yaw Anokwa: Exactly.

Scott Hanselman: And here's the thing though, my lack of empathy initially was like, "Well, I don't understand why you're having a problem with this. It isn't that hard. It works for me and my dainty fingers work fine. What is the problem with you?" You're building that empathy all the way through because you've got the collect application and then you've got ODK Central. Right?
Scott Hanselman: Now is ODK Central a database or is it a server that catches data and where does the data end up?

Yaw Anokwa: So Collect is the mobile app. So is in the play store, the Android play store and Central is the server. And so the server runs on somebody's infrastructure either on our manage hosted solution or on premise or locally on somebody's server. And it hosts the blank forms and it also hosts the data that is being sent. So the idea is that maybe a project manager puts the forms that they want the data collection to fill out on the central server, the phone links up to the server, downloads it, and then the submissions go back to the server. And once it's on the server, we make it really easy for folks to sort of use that data, be it downloading it as a CSV or Excel file where they can do their reports and natural your programmers, it has a beautiful rest API so people can integrate it with other systems including tools like Power BI and R and Python to sort of pull that data into their data analysis work streams.

Scott Hanselman: You say that you can do that OnPrem, so you could have that private and self-hosted with a little support, but you also of course offer your own reliable cloud version of this as well.


Scott Hanselman: This is a public good, but if people who are listening right now, if they have a text boxes over data problem, you've solved that.

Yaw Anokwa: I think so. And a lot of our users think so. Yeah, it's a nice solution for folks. We didn't always used to have the cloud hosted option. And I want to talk a little bit about that because as computer science researchers or as open source enthusiasts, we always thought like, "I mean, obviously people want to run their own servers," so we don't want to be in the business of running servers. We'll just give people the software and freedom, et cetera, et cetera, et cetera. It took us a long time to reach the point where a lot of our users don't want to run their own servers. It is a lot of work. And so we came to a point about two years ago where we wanted to solve two problems. The first problem was that it's really hard to build an open source project and make it financially sustainable when you're giving away the software and applying for grants that come randomly.

And so it was hard to of thoughtfully evolve the software. So that was the first problem. And then the second problem is what I just mentioned is that people don't want to run servers. Running server is hard, email delivery's hard. There are all these sort of hard problems that you don't want to be taking on as a nonprofit or government. So we launch our cloud hosted solution for that reasons that it makes people ... very easy to get an account and just collect the data you need. And it gives us this sort of predictable revenue that we can use...
to hire people and sort of thoughtfully evolve the software as opposed to randomly evolve the software based on what grant we get at what particular time.

Scott Hanselman: And that now is self-sustaining. And you're, like you said, you're talking to customers every day and they're signing up every day.

Yaw Anokwa: Yeah, it's been really amazing to see. I suppose it took us 14, 15 years to get there. But yeah, it's a very popular service, cloud service. It's enabling us to of ship more software quickly.

Scott Hanselman: ACM Bytecast is available on Apple Podcast, Google Podcast, Podbean, Spotify, Stitcher, and TuneIn. If you're enjoying this episode, please do subscribe and leave us a review on your favorite platform.

When did you know though that it worked? You just said that 14, 15 years. So I'm getting this sense that you're an overnight success in just 15 years. So anyone can do it, just grind alone in a room for 15 years.

Yaw Anokwa: I think that's the key to it. We always talk about there's lots of things that we could be doing better, but there's something to be said about showing up every day and just trying to get better every day. And so when we started in 2008, I knew it was working because I got an email maybe a few months in after we deployed the software from a group in Kenya and they said, "We have deployed ODK on 3000 phones in rural Kenya and the project is over. We've collected so much data." They were distributing water filtration devices. I'm addicted to email.

Instead of doing my graduate research, I'm just responding to support posts because that's more fun. And I responded immediately. I was like, "Did you mean 30 or 300?" And they're like, "No, no, no, no, 3000." And that's when I was really blown away because at that time I didn't know we had that many users. And to me, it was mind blowing that somebody would take just a random software from the internet, put it on a bunch of devices, deploy it to the middle of nowhere on 3000 phones and have success with it. And so I knew that we were on the right track then because somebody had found it useful, had success with it, and we were not involved at all. That's when I knew. But yeah, that was in 2009 or something like that. And then we've been grinding ever since.

Scott Hanselman: From a technical perspective, were you in any way surprised that that production load test worked? They could have said we put it on 3000, but it only worked on two.

Yaw Anokwa: I was very surprised because I have to be honest about that. As grad students at the time, your job isn't to write production quality code. Your job is to write the minimum amount of code needed to test an idea, evaluate it, and write your paper. And so anything that looks like production code means that you've spent
too much time on the engineering side and not enough time on the research side. So obviously, we try to make it as robust as possible. But yeah, I was surprised we've made a lot more progress since then. We've been very fortunate.

Scott Hanselman: The first Android phone came out just 14 years ago. Like Android has taken over the planet and particularly the African continent, you can't go anywhere without bumping into an Android phone on the continent. But that means that in summer of 2008, there wasn't an Android phone on.

Yaw Anokwa: That's correct, yeah.

Scott Hanselman: Or was there?

Yaw Anokwa: There wasn't. And in fact, I'm sure the statute of limitations have passed by now, but myself and Carl Hartung, who's my co-founder in ODK, we were again interns at Google and pretty sure we brought the first Android phones to the African continent. Now we didn't tell anybody we were importing these new devices, but we loaded up 20 phones into our bags, flew to Uganda for our project and brought the Motorola, I think they were G1s, HCC G1s, little flip phones.

Scott Hanselman: Yeah. HCC G1 was October, 2008.

Yaw Anokwa: Yeah, we brought those phones, 20 of those phones to the continent. Now Android is everywhere throughout the world. ODK runs on, see last I check maybe like 22,000 different Android devices. It runs on TVs. We actually have maybe about five or six users who use ODK on TVs, which it doesn't really make sense, how does that even work? But the analytics show that people are using it on TV. So yeah, it's crazy.

Scott Hanselman: Do you ever dig into that? I'm imagining someone coming into a health clinic and they're sitting in the room with a smart TV on a remote, filling out a form waiting to get seen. Do you know what they're using it for?

Yaw Anokwa: We anonymize all this stuff, so we know that people are using it on TVs, but we don't know who and why. My guess is that it's really nice on a big TV when you're doing a training, so you have a team of 500 people who are going out to collect data and you have an Android TV. Well, you might as well and solo indicate there and then people can see the interface and so they must be using it for training. But I can't imagine somebody in the rainforest carry around a 27 inch TV on their back and using that for data collection.

Scott Hanselman: Now this is an open source. It has a wonderful API. You pointed out that it has a rich web API. Does it have plugins and the ability to talk to other systems or does it talk to other systems via its open data formats?
Yaw Anokwa: Yeah, it's a great question. So ODK, it is open source and has lots of APIs. I would say the way that most people, most sort of developers interact with ODK is two ways. The first is by forking it in some ways. So ODK is kind of the first and most well known of data collection platform in the global health and development space. And so there are a lot of derivatives of ODK that are out there. And so chances are if you see somebody collecting data in the field somewhere, either they're using ODK or they're using an old clone of ODK or they're using ODK's formats in some way. So that's sort of very typical. The other way that people interact with it is with our API. So we have specs and APIs that a lot of platforms implement. And that's been kind of nice because we have the sort of standard representation for our form that people can implement their own ODK compatible services and there they're a bunch of those.

Scott Hanselman: And you do have a relationship. I understand that XLS form is a standard for building forms in Excel. And is it true that ODK can speak Excel forms?

Yaw Anokwa: It's more than that. So XLS form is a format that was developed and pioneered by a company called Ona. Great group of folks over there. And now the ODK team maintains it. So we maintain and evolve the XLS form standard. We maintain an evolve platform called Enketo, which is a web renderer for that form standard and a lot of the core library's under the hood.

Scott Hanselman: And these are standards, published standards that we can go and learn about and use.

Yaw Anokwa: Yeah, they are published standards. They're not like WC3 standards in that way, but they're sort of the defacto standard for data collection, particularly offline. So yeah, both the forms, the underlying libraries and the APIs, we maintain all of those.

Scott Hanselman: I'm just kind of thinking about the timeframe here. I think the cloud just started when you started and Android phones just started when you started. You really had a good idea at the right time. I mean, how good is that?

Yaw Anokwa: Yeah, we feel very lucky. We're at the right place at the right time with the right team. It just sort of worked out really nicely. Very fortunate.

Scott Hanselman: I do have a weird relationship with the word luck, so I want to gently push back and I think you'll probably agree with me that luck is opportunity plus being prepared and you were absolutely prepared and an opportunity presented itself. So you didn't get lucky. You made lucky.

Yaw Anokwa: I understand that point, for sure. As I think my wife says it, you get dealt a set of cards and you play them well. But I also never want to discount sometimes the randomness of it. So maybe I was lucky in 2008 and then I've been working at it for a long time since.
Scott Hanselman: Yeah, that is the thing, it's hard work. There's luck that you see in a press release or a single Instagram post, but then there's also the five years of grinding-

Yaw Anokwa: For sure.

Scott Hanselman: ... that makes the big difference as well. Now I understand that ODK is being used in the public health sector. It's being used for large scale disease surveillance. Largely, it’s humanitarian, it's the public good. Are there commercial uses? Are people using it just in companies to do late stage capitalism or do you discourage that?

Yaw Anokwa: We don't discourage anybody. We believe in sort of freedom in the sense that you can use it for whatever. I know of militaries that use it. I know of banks that use it. I know of a seat manufacturer for an electronic car company whose name starts with a T that uses it. What could it be?

Scott Hanselman: Can't win them all.

Yaw Anokwa: In many ways, we don't want to prevent anybody from collecting the data they need, but our focus always is on the social impact sector. So those are the folks who determine what features we build. And if banks or other companies want to use it for capitalism, that's fine. It's their prerogative.

Scott Hanselman: I hope this isn't too personal of a question, but you could turn on the marketing machine and hire a bunch of sales people and turn this into a money printing machine. But you at every turn have not done that. I don't even know if you have a marketing department.

Yaw Anokwa: I am the marketing department. We don't have-

Scott Hanselman: You've chosen though not to turn this into a machine.

Yaw Anokwa: I guess, maybe I'm simple minded. I'm not at a business mind or whatnot. I just don't get the point of all of that. There are plenty of data collection platforms that have great marketing. Our job, the team's job, our focus is just to build the stuff that folks find useful, relevant to the work and allows them to do that work. We generate enough money to do that and no more. And we have a lot of fun doing it. And I think our marketing machine really is our community, the people who use it in the field. And sometimes when I get on calls with customers, I ask them always like, "How did you find out about ODK?" And often the answer is, "We use this other platform and our data collectors refuse to use it," because they get paid per submission and if the submissions get lost, they don't get paid.

And so they really trust ODK and they refuse to use something else. And I think that trust comes from years of us doing the right thing when the going got tough and not using it as a money printing machine. We really, genuinely,
everybody works with the team. We believe in producing a public good, making it as widely available as possible and helping folks who are helping others do their work. And if I want to do something else, yeah, I could just get a job somewhere else, but what's the point of that? There's plenty people who are doing that.

Scott Hanselman: You have enough and you sleep better at night and you are affecting the lives of millions and millions of people, which is-

Yaw Anokwa: It's pretty good. I just don't understand why more people don't do it. I have to be honest with you. It just seems like to me, it's such a joy and a privilege. I don't know why people aren't stopping what they're doing now and doing some more stuff, but yeah, different strokes for different folks, I guess.

Scott Hanselman: I love that you say that your marketing team is your community. I know that you are a big fan of Discourse, not Discord with a hard D, but Discourse.

Yaw Anokwa: That's correct.

Scott Hanselman: The forum. And you run your forums on discourse as well. So when you're out there deploying open source software, you're also recommending other open source projects for deployment.

Yaw Anokwa: Yeah, for sure. I am not sort of religious about open source. I think it prevents locking in these kinds of things, but to me, the very idea that other folks are out there building software and essentially giving it away and giving folks the power to do what they want with the software is just a nice thing to do. So I love Discourse as a platform. We use it a lot in ODK and whenever I get a chance to sort of advocate for it, I'm the first one to do that because it ultimately, it's been the heart of our project because it enables us to connect to our community and get input from the community and that's essential to building great software.

Scott Hanselman: Very cool. Well, folks can check out ODK at getodk.org.

Yaw Anokwa: That's correct.

Scott Hanselman: You can learn all about it. And I think the number one thing, as a member of your community, as the open source community, as the technical community, as a computer scientist myself, it's about awareness. So I'm happy to be a member of your community and your marketing department now.

Yaw Anokwa: Perfect.

Scott Hanselman: And if folks are listening right now-

Yaw Anokwa: You're hired.
Scott Hanselman: Yes. If folks are listening and you have a problem and you think, "Wow, ODK has solved that problem," and you can check it out at getodk.org. Thank you so much, Yaw Anokwa, for chatting with me today.

Yaw Anokwa: Thanks for having me, Scott. Yeah, if there's any way I can help anybody listening to the podcast, my email, I always like to share my email. My email is YAnokwa, first initial, last name at getodk.org. You can send me any question you'd like about open source, about data collection and I commit to answering it. So send those emails in.

Scott Hanselman: Absolutely. Brilliant. This has been another episode of Hanselminutes in association with the ACM Bytecast. If you're listening to it at Hanselminutes, there's other shows that you can listen to. And if you're listening to it at the ACM, be sure to explore the back catalog of the ACM Bytecast and we'll see you again next week.

ACM Bytecast is a production of the Association for Computing Machineries Practitioner Board. To learn more about ACM and its activities, visit acm.org. For more information about this and other episodes, please do visit our website at a learning.acm.org/bytecast. That's B-Y-T-E-C-A-S-T. Learning.acm.org/bytecast.