Episode 105: Kris Shaffer

# KL: Katie Linder

# KS: Kris Shaffer

# KL: You’re listening to “Research in Action”: episode one hundred and five.

# [intro music]

# Segment 1:

# KL: Welcome to “Research in Action,” a weekly podcast where you can hear about topics and issues related to research in higher education from experts across a range of disciplines. I’m your host, Dr. Katie Linder, director of research at Oregon State University Ecampus. Along with every episode, we post show notes with links to resources mentioned in the episode, full transcript, and an instructor guide for incorporating the episode into your courses. Visit our website at ecampus.oregonstate.edu/podcast to find all of these resources.

On this episode, I am joined by Dr. Kris Shaffer, a data scientist with a background in computational musicology. Kris currently works as an Instructional Technology Specialist and Adjunct Instructor of Computer Science at the University of Mary Washington. He also does freelance work in web and social-media intelligence, and serves as a volunteer researcher for Data for Democracy. He is a Contributing Editor and Board member for Hybrid Pedagogy and the lead author and editor of Open Music Theory, an open-source, interactive textbook for undergraduate music theory courses.

Thanks for joining me on the show today, Kris.

**KS:** Yeah! Glad to be here.

**KL:** So, I am super excited to talk with you because you have a background in computational musicology, which I don’t really understand anything about what that is, and so—and I’m assuming our listeners may not know anything about it either—so I’m hoping by, first, just starting out by telling us: what is computational musicology?

**KS:** Yeah, yeah. So, it’s a relatively small field—I’m mean music theory on its own is already pretty small and somewhat obscure—but essentially it is, people are familiar with data science or digital humanities, it’s the musical version of that using code, software, a lot of automated functions to analyze what’s going on in and around music. It’s kind of like a big data thing, so it’s encoding musical data, whether that’s notes and rhythms and chords, which is what I tend to do. Or, metadata, like genres and dates and artists and loudness, softness, length of song, things like that. You can model any of that stuff and start to see what kind of patterns emerge. Start to define things like musical style and more specific nuanced ways by accessing all that data.

**KL**: Okay so, I must admit that when I think about technology and music like the first thing that comes to mind is like club music. I don’t always think about you know these other things you are describing here. So I would really love to dive deeper into it, because I really think that especially now that we have so much music that is techno music. You are not talking about that you are talking about that. You are talking about something that is very different here.

Um so I would love to hear what are some of the research questions that you are asking in this field and is you have a couple examples of how the projects are done. That really are good examples of this computation musicology work.

**KS**: Yeah so, it could be techno music, dance music, club music, it could be pop songs, it could be classical music. My first projects were actually looking at some post war avant-garde music. From the mid twentieth century and asking what kinds of patterns in the core progression and melodies looked similar to the patterns we would understand in traditional classical music. Like Hiden, Mozart, Beethoven that kind of things. There was the composer I was looking at, György Ligeti, Hungarian composer who people might know from the 2001 soundtrack. He claimed that he was using the same cords the same vocabulary, not the same syntax not the same progressions and rules together. And a bunch of music theorists and music psychologists just took him at his word. And I said “let’s not just take him at his word”, and so I went through pain staking through a number of pieces of his. Encoded the cords just as text data and then used some of the same kind of text corpus data tools and the digital humanities world. To see what kinds of cord progressions were there and what were not and found that there was more to the story than what he said and we can’t always take composers at their word. And that was kind of my first… into it. There are a number of publicly available free data sets that people can download that contain things like cords and melody notes for things like folk songs for classical music for pop songs cord progressions. And there is a data set called the million song data set which is based on a large number of pieces of Meta data how long is the song, what’s its loudest moment, what’s its lowest moment. They use algorithms to extract the tempo of the song and so you can do things like okay compare jazz songs to new age songs to country songs which is longer, shorter, faster, louder things like that. Then start to see if you can note patterns changing over time. And so I had a class once with a number of computer and science students together where we just dug into some of these data sets and created all the models you can create in that three week intensive maymester course. And extracted some interesting things about musical style and how we define it and how structural properties in music relate to the genre that we perceive as music listeners.

**KL**: Okay so Kris I have so many questions about this I find it really fascinating. Okay so I am curious some of our listeners might be thinking and I have this question. What is the ultimate goal of this? And as in any kind of research area we have several different things we may be looking for, but I am wondering if you can share a little bit more about it this to kind of better understand the culture of music to better understand things like creativity to be able to see like you said patterns and new areas that are being developed in music? What are some of the larger bigger questions that are really being addressed in this research?

**KS**: Yeah so, like a traditional music theorist would be asking questions of how to define a particular music style with nuanced and more specificity. And start to understand how those styles emerged overtime. Then you start to understand how music plays a role in society seeing what genre influencing another genre, what style influences another style what musicians may have studied with who and ultimately what I find really interesting what song writers and composers are saying what their music and how does that actual line up with their music. Because we often just believe what they say especially if it is something that would be kind of hard to tease out and investigate. So, that’s kind of like the traditional academic music theory humanities side of things. But there is a business side of this as well that services like Pandora and Spotify try to take listener tastes and predict what other music they might like and service that up algorithmically and then hopefully make more money. As listeners stay with their service or upgrade their service. Encourage there friends to sign up so those kind of recommendation engines we found the data select is doing research for that and the music theorist is doing research on the musically structure and style. We don’t always talk to each other so both sides are kind of missing something important and we were kind of trying to bridge that in our class research. As we saw the recommendation engine folks considering style which would be more musical structure and genre how listeners interact with these things socially in groups. Treating them as the same thing and we were able to tease out that there is a significant difference between the two. And so just because a piece has a certain compositional element in it that places into its style, but that might not actually be predictive of how a listener would appraise it positively or negatively because they are using a different set of criteria sometimes. And so that’s something that for me is really interesting because that implication can make people money and has this kind of timeless humanistic understanding of how human culture works. That aspect to it as well, it is a fun feeling. Because you can start to bridge those gaps if you are reading an institutional and commercial research and talking to people on both sides there is a lot of fun things to tease at there.

**KL**: So, I am curious Kris how you became interested in computational musicology? What lead you into this?

**KS**: Um I have been writing some kind of code since I was like six on and off. So I have always been a tech nerd of some sort whether actual coding stud or just a power user. Just interested in geeking around with these things. And so as I was following the career path and music and in grad school for music I just realized this would be a fun thing to do we had a couple faculty members in there who were really involved in the mathematical side of music theory and building these new mathematical systems. And one of them being Ian Quinn who became my advisor had worked in computational musicology. It seemed like a really great cross section of all of my interests together. It’s not something I ever got a job in. It’s what I did my research on as I kind of teaching the normal undergraduate theory curriculum when I was a faculty member. But its continued to be an interesting thing as I have even moved on in my career outside the faculty track I have still been doing work on this, keeping tabs on this and what’s going on and just having fun thinking about these things. I even gave a segment of the million song data set to my data science 101 students this semester. So a very inter disciplinary course but they can dig in and understand music genres and how that relates to some of these higher level songs.

**KL**: Well this is super interesting we are going to take a brief break when we come back. We are going to hear a little bit more from Kris about some of his open source software development. Back in a moment.

[music playing in background]

**KL:** Just as the “Research in Action” podcast is dedicated to quality, so is the producer of this podcast: Oregon State University Ecampus, the university’s online education division. As a national leader in online education that’s committed to the Quality Matter’s standards of online course design, all OSU Ecampus faculty are encouraged and supported to see QM course certification and involvement in peer review. Learn more at ecampus.oregonstate.edu/faculty/QM.

# Segment 2:

**KL**: Kris one area of your work that I am really interested in as well is your engagement with open source software development. I am wondering if you can start out by telling us about some of your open software projects and maybe a little bit about what is open software. In case people aren’t really sure what that is.

**KS**: Yeah, yeah. So, open source software is essential a software were anyone could if they wanted to (they usually don’t), but they could look at the source code that makes a software run. But most of us interact with software just by you know clicking on the app and it opens and then we use it to do stuff. And underneath that I think most people understand there is code that tells it what to do and how to interact with us as we interact with it. But most software that people have on their computers will be proprietary software, and so what’s on their computer is not the code that was written by the programmers to make it work. But its binary code zeros and ones that talks to the operating system which talks to the hardware which makes it do what it does. And so most people don’t usually have access to say the code behind (I am looking at my computer right now) the Slack app at work or Twitter or their operating system there is Mac or Windows. They don’t actual know what’s going on under the hood and so for most people that is fine, but for people who want to learn how it works who want to control their own devices more who want to learn from it and build on it. We need to be able to see that initial source code so we can learn from it tweak it fix the bugs in it patch the security holes ourselves and share that with each other.

And so there is this whole community of open source software developers that are part of this on the best of days quite wonderful community of sharing there is plenty of in fighting and discrimination and all kinds of other bad social things that happen in that environment. Certainly don’t want to discount that. At its core the idea is we are in this together at some extent. And if somebody has already solved a problem well let’s put our efforts into solving a new problem, rather than solving the same old problem again because we don’t have access to their solution. So that is kind of the ideology behind it. As an academic it has always kind of resonated with my ideas of scholarship. Scholarship should be open that its public knowledge that we are contributing to and so in fact my first peer reviewed article was for the journal Hybrid Pedagogy, and it was called open source scholarship. The idea that the work we do as scholars builds on existing work we don’t wait for patents to expire til we take the ideas of a previous scholar or contemporary scholar and advance them further or apply them to new directions. So, that idea of open source software development and learning from others code and developing new code based on that is very much aligned with how I view our work as scholars and teachers as well.

**KL**: Okay so I am wondering if you can offer a couple of examples of open source software that people might be familiar with like are there a couple popular examples. And also dig into some of the things you are working on as well.

**KS**: Yeah so, a common really popular open source application I guess would be the Firefox web browser that is kind of the one most people know. Unlike other web browsers not all of them but most the common ones its build on code that anyone can look at modify you can download it, rewrite it, compile it yourself on your own. You could distribute it, you could sell it without paying any royalties to anyone else. And yeah so what that means is that there ends up being a large community both a non-profit at Mozilla but also the developers who are contributing patches when something goes wrong who are contributing new features.

Another would be the android operating system which is built on Lynix if you run android on your phone there are a lot of proprietary apps on android. But android its self is built on an open source operating system which is why it is a little bit more flexible sometimes than say IOS. There are other reasons for that. That can also lead to more bugs as people are introducing new code and there is not one person who kind of other sees everything tightly. But it also means that there are potentially more fixes as well as more ways to modify your phone as you unlock it and start downloading other versions of android that are not coming out of the google curated universe. Its actual companies like google that have a really interesting relationship with open source. You can build a business model around open source so it is free. You add value not in the code you add value in your ability to support it some of the customizations you do on behalf of individuals. And just kind of over seeing things to make sure the highest quality fully vented code ends up in the production environment for whatever device needs to run it.

**KL**: Alright transition into some of your projects. What are some of the things you have been creating as open source software?

**KS**: Yeah so, for me most of my open source software is basically the code I have used to do my research and analysis projects. So, is you wanted to do some computational musicology there is a number of scripts and applications that I have put together that are on my github profile linked to my website. So you can download and start to play around with the same data. Bring in your own data start to tweak the code and extend it and do different things. As well as programs that I wrote as I was learning new languages and learning what they can do and I had never seen a value in hiding a code that I will never be able to sell. Might as well share it if it can help somebody else out and I have also built a word press plug in for the annotation tool hypothesis. So it allows people who use hypothesis an online annotator to collect there annotations and there service then my plug in will allow for them to display them automatically on their website. Without doing any manual labor, you highlight something on the web and save it to your hypothesis account now next time someone goes to your website they will see it on that page.

So, I have a number of those kinds of things as well as some analytics tools I have built for the demand of ones program at Mary Washington where students and faculty are given domains by the university and we can contract how that program was developed over time, who is using it what are they using it for so I have got some analytics tools there. Um so yeah as well as some of my kind of activist volunteer work for Data for democracy. Mining tweets and watching government websites for changes and things like that. Any time I write a code that is proprietary for someone who paid me to write it if it is not awful I tend to put it on github. So someone can hopefully get use of it or potential open a new collaborative relationship.

**KL**: Alright well we will link to these things in the show notes so people can check out these different projects and your profile on github. And also to a previous episode we had with your colleague Jesse where talked about domain of one’s own. So, we will link to that in the show notes as well in case people are interested.

I am curious let’s dig in a little bit more to github. This is something people are not familiar with. Can you example kind of what github is? And how you are using it as a tool for these projects?

**KS**: Yeah so github it is really interesting especially talking about open source and proprietary, because github is a for profit company that allows people to share predominantly open source code with each other. And it is built on software git which is an open source program that was developed for the team who was running the development of the lyniox operating system. So all that [KL: But there is many layers here], yeah yeah yeah many layers. But essentially github is a place where is you write code and you want to share it with the world you can put it on your website and hope somebody finds it. And if they want to collaborate with you, you can email back and forth and figure something out. But if you host it on github there is kind of a social network component to it where you can follow people and see what they are working on. You can find there code you can comment on it, you can download it and use it and then tag some issues, if you find a bug or want to request a feature or reach out to potentially collaborate. It’s a really easy place for collaborators to share the code they are working on individually and collaboratively and its built on a really robust system. Of whats called version control which is to help make sure that a bunch of people working on the same code at the same time don’t want to mess up each other’s stuff. And so it’s a place its becoming or has become now the place for coders working on survey software will host there code and find each other in order to collaborate more.

**KL**: Alright we will defiantly link to this in the show notes in case people want to check it out.

We are going to take another brief break. When we come back we are going to hear a little bit more from Kris and his recent project he is working on and his research.

Back in a moment!

[Music playing]

# Segment 3:

**KL:** So Kris your work has recently shifted into a new area. Which is disinformation and hate media. And this is a little different I would imagine from computational musicology. But maybe using some of the same tools. So I want to dig in and learn a little bit more about this, but first what lead you to research these new topics?

**KS**: Yeah so when I was winding up kind of my last big computational musicology project and thinking about what’s next. I am no longer a music faculty member I work as an instructional technologist and so the research that I do can be much more open in terms of what I am studying. And this was after the presidential election and I was thinking I wanted to do something that could be a little bit more directly applicable to more people in meaningful ways than computational musicology. Nothing against that there is value in that but I wanted to kind of get my hands dirty in a bit more of an activist role.

And I had a couple friends that directed me to the group Data for Democracy which is a group of volunteer collective of mostly data scientists and other technologists. Who work on projects related to election hacking, refugee statuses and migration within the U.S. , immigration lawsuits. The groups for D for D have worked with public health, won an award from the United Nations, worked in the first initial court case that got the first travel band from the Trump administration blocked in federal court. So there has been a lot of cool things going on there really, helpful socially important things that they have been working on. So I joined that group and tried to see where I can plug in and I have been doing a lot with twitter and the twitter API already. So, I got involved in a group that was studying the spread of disinformation and hate media on Twitter. And that’s kind of how it started and just as things come up we started to pursue them and build a work flow and kind of settle into an area where we were able to make some head way. And that’s how it started I mean using a lot of the same tools, but it really came out of moving from digital humanities to data science. More broadly than wanting to do something for social good with that.

**KL**: So I am always really interested when people make these kinds of shifts in their work, because sometimes from the outside it just looks like what a weird shift to make. But then it gets explained and it makes a lot of sense in terms of kind of why you decided to move into this direction. I am curious if you could talk a little bit about how this new area is connected to your new research area and particularly in terms of what kinds of previous methods you are able to draw on. And really leverage to shift you kind of into this new area.

**KS**: Yeah so, we talked in this segment on computational musicology about some of the ways we would model musical style and musical structure. A lot of those same kinds of techniques are applicable when you are studying these kinds of texts and of course if you have a few million tweets or a few hundred campaign speeches. You are dealing with a corpus of texts. A lot of the tools computational musicologists have borrowed from natural language processing and computational linguistics and just large scale literally corpus studies. Well those are the techniques you would use to study text. So having already cribbed them for music I can go back and go let’s use them for text and this is the text that we are looking at. So there is a lot of those kinds of techniques. And I have also found that about data science in general that my training and experience as a music theorists was really fine turned towards finding clear signals and noise. Looking at messy human unstructured data and finding the patterns that were the most significant overarching controlling patterns from which other patterns where generated. And that’s long sense before I was doing that with code I was doing that with notes and rhythms on a staff and rulers and pencils and calculators and things. Finding the geometry of these things and like we talked about before the math. The math of music so these kind of mathematical structures and emergent patterns are exactly what we are looking for when we are trying to find evidence. Of say like a bot driven um fake grass roots campaign in support of a particular candidate or idea or political entity. And we are looking for people who are hiding their tracks in their activity on online. That’s what I did at my disorientation trying to find ways where a composer is not telling the full truth about what was going on at his musical structure. And so there is a lot of other lap between the work that a digital humanists will do and the work that a web intelligence analysis would do.

So that has made it a really easy shift for me, not really easy but a natural shift and one where it is more a matter of learning some new kinds of code and a few new domains of applications that it is learning a whole new field.

**KL**: So I just have to ask do you have the term web intelligence analysis on a business card somewhere? Because that is pretty fancy sounding.

**KS**: Yeah it is um. I think it is on my resume, but not on my business card. [**KL:** Maybe it should be] I have been doing some freelance work for a number of firms, and one civil rights organization. Doing this kind of research, essential looking at whether it is social media data or who owns the website and what other websites do they own, and who provides the payment service on particular sites. So that the civil rights group may say hey that isn’t cool that you are essentially the enabler of the business model of this particular group given the policies and actions that they are pursuing. And so yeah there is a lot of, its fun. And when I say web intelligence that is a couple of companies I worked for that is there things. So yeah it is yeah.

**KL**: Okay, so I am curious Kris if you can tell us a little bit more about where you think this line of your research is going. Do you have a sense of the direction?

**KS**: Yeah so I mentioned I am starting to do some freelance work here and there for non-academic firms in this area. So, that’s defiantly in terms of myself and my professional direction something that seems to be emerging in the future as I am using these less for academic purposes and more for like you know. Helping specific companies or individuals who have an interest here. And as an academic I have always thought of these things in kind of an ethical framework. So, trying to make sure when I take on these projects I am taking on projects that I really believe in. I also find as I am teaching intro to data science a little bit more at Mary Washington, helping my students navigate both the academic and the industry space here is becoming more and more important for me. Helping show them the ropes and to find their own projects.

One area that is starting to emerge with some of the folks I have been in contact with is human trafficking research. And so I am currently working with some people at the UNC Charlotte the data science there. Who have received some funding and have a lot of students and faculty and subject matter experts from around kind of the mid-Atlantic area. Studying human trafficking and the ways that victims are recruited online and the ways that online advertisements can give off indicators of human trafficking and underage victims. So, that’s an area that is starting to become more important for me on the research side, because that is more academic non-profit territory. And that is both really exciting and really frightening to dig in and see this stuff. Neo-Nazis are bad enough and before you start getting into essentially the sex slave trade in the U.S. and seeing how big that impact has on the U.S. And how uninformed law enforcement can be, not law enforcement…legislators can be in determining the resources they can give the law enforcement and NGOs and agencies. As they are seeking to address this issue.

So that seems to be as I said I think more and more about using data science in a way that can accomplish good in the world that seems to be the direction I am heading a little bit more. And we will see as that goes forward what kind of successes and roadblocks we hit and what I end up coming back to. That seems to be the direction that the research side is taking right now.

**KL**: Well Kris I really look forward to seeing were this goes. And I want to thank you so much for coming on the show. And sharing your expertise on computational musicology and also sharing about the future directions for you research.

**KS**: Alight well thank you. It was great to be part of this, thanks for the invitation.

**KL:** Thanks also to our listeners for this week’s episode of research in action. I am Katie Linder we will be back next week with another episode.

[Music plays]

# Show notes with links to resources mentioned in the episode, a full transcript, and an instructor’s guide for incorporating the episode into your courses, can be found at the show’s website at [ecampus.oregonstate.edu/podcast](http://www.ecampus.oregonstate.edu/podcast).

# There are several ways to connect with the “Research in Action” podcast. Visit the website to post a comment about a specific episode, suggest a future guest, or ask a question that could be featured in a future episode. Email us at riapodcast@oregonstate.edu. You can also offer feedback about “Research in Action” episodes or share research-related resources by contacting the Research in Action podcast via Twitter @RIA\_podcast. Finally, you can call the Research in Action voicemail line at 541-737-1111 to ask a question or leave a comment. If you listen to the podcast via iTunes, please consider leaving us a review.

# The “Research in Action” podcast is a resource funded by Oregon State University Ecampus – ranked one of the nation’s best providers of online education with more than 50 degree programs and over 1,000 classes online. Learn more about Ecampus by visiting ecampus.oregonstate.edu. This podcast is produced by the phenomenal Ecampus Multimedia team.

#  “Research in Action” transcripts are sometimes created on a rush deadline and accuracy may vary. Please be aware that the authoritative record of the “Research in Action” podcast is the audio.

# Bonus Clip:

[Music plays]

**KL:** Kris I am really curious from hearing you talk about your experience in computational musicology. The relationship between mathematics and music. Because I think that this is one of those areas where people might make some assumptions about how there is a big relationship or no relationship. So can you talk about this a little bit?

**KS:** Right yeah and as someone who has taught a lot of undergraduate music majors. That tension between there being a lot of math in music theory and musical structure on the one hand, but also a lot of artist pursuing music that think that they are kind of outside the realm of math and science and such. Then they get surprised when they come to my class and we are talking about brain science and the math of musical structures and set theory and group theory. It ends up becoming a very interesting kind of social dynamic as we work through that. Because there is a lot of I mean, like the humanities there is something very technical about musical structure and there is something very social about musical structure. And what is important and interesting about musical structure and music theory to me is the way those two intersect and interact. And so how the structural things that composers do and performers do, have an impact on the way society coheirs together around music and around these different genres.

So when we are talking about classical music we are often talking about melodies and rhymes and cords and cord progression. And there is a lot of geometry and arithmetic that go into the defining of musical scales of cords and cord types. There is a lot of statics that go into chord progressions there is a lot of set theory and group theory. For those who know those terms or concepts are kind of straining their brain to high school and college math. Then you remember a little bit about them with brackets and numbers and commas and things. There is a lot of that involved in the way particularly Avon guard post tonal music works as well as some early music. Kind of like medieval and renounce type music. And these are all things that come to play when we ask why does the C-major cord in the key of F want to…quote end quote want to go to F.

 That kind of thing and this is very much taught and this is for the social component comes in. It is very much tied to how we learn language and colors and other kinds of things in our environment. Especially as we are at an early age and grow we learn a lot of those kind of language like things. There is a lot of that vocabulary and syntax through statics as we are exposed to things. And we start to get familiar with certain kinds of sounds and certain kinds of arrangements of sounds. Which is one of the reasons that the different cultures of the world have sometimes very different sounding music and of course there are commonalities between them and many of those styles. That relates to some of the biology of the human ear and the auditory system and what kind of limits that sets on our ability to process these things cognitively.

So all of that kind of sets the stage for kind of the math of music. What kinds of relationships can our ears and brain parse? What kinds of things can we physically produce as musicians? And what kinds of things are we exposed too early in life and therefore come to expect as we engage music in the future. And so that combination of brain science and statics and inter cultural awareness is just really interesting and exciting for me. Even if a particular musician or scholar focuses on one aspect of that you really have to deal with them all together at some point. Because that is what makes music so wonderful and interesting it’s the shared experience that we have that allows us to share the experience of music. And we define that in large part, but not exclusively in terms of the statics of how these very scientific objects are put together in music.

**KL**: Well thank you very much for sharing a little bit more about the relationship between music and math.

[Music plays]

#  “Research in Action” transcripts are sometimes created on a rush deadline and accuracy may vary. Please be aware that the authoritative record of the “Research in Action” podcast is the audio.