EPISODE 923

[INTRODUCTION]

[00:00:00] JM: Data privacy policies have changed how software organizations need to operate. As consumer preferences have shifted in favor of strong privacy, software companies are having to examine their policies around data collection and retention. Many software companies were started in a time with different norms around data. Building a new application that is compliant with GDPR is a hard, but updating an existing application to align with GDPR is even harder.

Joshua Prismon is chief architect at FICO, a company that builds systems around credit scoring. Joshua joins the show to discuss how a large company like FICO has responded to changing consumer preferences through changes in its software architecture and engineering.

We are hiring a head of growth for Software Engineering Daily. If you like Software Engineering Daily and consider yourself competent in sales and marketing and strategy, send me an email, jeff@softwareengineeringdaily.com.

[SPONSOR MESSAGE]

[00:01:09] JM: This podcast is brought to you by PagerDuty. You've probably heard of PagerDuty. Teams trust PagerDuty to help them deliver high-quality digital experiences to their customers. With PagerDuty, teams spend less time reacting to incidents and more time building software. Over 12,000 businesses rely on PagerDuty to identify issues and opportunities in real-time and bring together the right people to fix problems faster and prevent those problems from happening again.

PagerDuty helps your company's digital operations are run more smoothly. PagerDuty helps you intelligently pinpoint issues like outages as well as capitalize on opportunities empowering teams to take the right real-time action. To see how companies like GE, Vodafone, Box and American Eagle rely on PagerDuty to continuously improve their digital operations, visit pagerduty.com.

I'm really happy to have Pager Duty as a sponsor. I first heard about them on a podcast probably more than five years ago. So it's quite satisfying to have them on Software Engineering Daily as a sponsor. I've been hearing about their product for many years, and I hope you check it out pagerduty.com.

[INTERVIEW]

[00:02:36] JM: Joshua Prismon, welcome to Software Engineering Daily.

[00:02:39] JP: Thanks. Good to be here.

[00:02:40] JM: You work at FICO. Explain what FICO does.

[00:02:44] JP: So I'd like to tell people that FICO was the big data or fast decisions or AI company that they've never heard of. Basically, you can think of FICO as a company that does decisions. What I mean by does decisions is you're familiar with the concept of a credit score.

[00:03:01] JM: Of course.

[00:03:02] JP: Yeah. So we don't do your credit score, but all of them license our technology and most of them are derived from what we did around the FICO score, which is basically a score that predicts how likely you are to go bankrupt or to not paid your credit card or to not be able to pay on your house or whatever those things are.

So that is kind of our first original business. But then on top of it, we do a lot of business around predicting fraud. So if you're in the U.S. and you swipe your credit card or you put it in, it usually go through some sort of computer system that determines whether or not a particular transaction is fraudulent.

That's usually FICO software, or if you're going and you're applying for a loan, either like a home loan, or a HELOC, a line of credit, or even your cellphone that more often than not tends to go through some piece of FICO software that makes a prediction and then also applies a set of

rules that determines whether or not you should get some line of credit or have some financial action or like an increase credit line or anything like that.

[00:04:05] JM: Do you only design the algorithms or do you also hold data on these different data lines?

[00:04:12] JP: So we actually do both. The vast majority of what we do is designing sort of decisions and some of those are algorithm based. For example, we might have some sort of scorecard that predicts how likely you are to go bankrupt. We might have a set of rules.

But then we also read software that actually combines those things, because you don't just make a decision based off of a score. You also make it based off of rules that say, for example, “If you're in this state and you’ve got this debt-to-income ratio and you've been a customer of the bank this long or things like that, that isn't just an algorithm. It's actually software, and we’ll actually run that software.

What we don't do is we’re not typically the people who offer that as an outgoing service. So processors, companies like TSYS, for example, will actually license our software and they’ll run it, or will actually offer it in our cloud, the FICO analytic cloud, in which case we run it for the customer, but it’s still the customer’s instance. So it's running as part of our software, but it's their instance and they actually get to manage the business rules and the logic and all of that behind it.

[00:05:17] JM: Does FICo serve customers outside of the United States?

[00:05:21] JP: We do. In fact a huge portion of our business is outside of the United States. We have a lot in Europe. We've got a lot in Asia. We’ve got a lot in Latin America, and that’s actually one the key challenges. When we write software, we have to write software that not only does the same business purpose if you will or the same customer journey, whatever it is, for a customer in the United States. We also have to make sure that it works for a customer who is in Thailand, for example. The rules and regulations as well as the customer behavior is very different depending on what part of the world you’re in.

[00:05:55] JM: How have companies like FICO historically been regulated?

[00:05:59] JP: So that's a really good question. My flippant answer to that is pretty impressively. Just maybe a little bit short of what say a defense contractor might do. The reality of it is that the financial services world is heavily regulated and for some very good reasons, right? If you go back throughout history, you will find out that the application of credit has societal impact. So it used to be that, for example, if you wanted to go get a loan, say, in the 30s 40s whatever, what you would end up doing is you'd end up going to a bank, and that bank may make the decision about whether or not you get that loan based off of a personal relationship with you, right? Am I playing golf with this person? Does this person look like me, right?

So starting in 1970, for example, the federal government passed the FCRA, which the Fair Credit Reporting Act, and that's a law that basically regulates the collection of consumer's credit and information and access to the credit reports. It really was there to sort of have a layer of fairness and accuracy and privacy on top of something like the credit score.

So FICO basically exists because the models and the decisions that we make are compliant with what governments actually require and are also predictive for what our customers need. So we can predict whether or not a customer is going to go bankrupt or not, but we're going to do so in a way that's consistent with the law.

[00:07:21] JM: I want to spend much of our conversation talking about regulation in terms of GDPR and how that affects software engineering. In order to move to a place where we can talk about that in the context of FICO, we should talk a little bit about your software architecture, and that's a big subject for FICO, because you've been around for 63-ish years. So I'm sure there're a lot of legacy systems as well as new systems, but maybe you could just give me a birds eye's view of the software architecture of FICO.

[00:07:56] JP: Sure. So as you pointed out, we've been around for a very long time. We generated the first predictive model for originations back in the 1950s using a mainframe and then having people actually go out and scorecards by hand. What we've done since then as we've launched a series of products in individual spaces.

So you can think of originations as being one of the problems that we solve. Fraud detection, another problem that we solve. Debt management is another problem we solve. Each of these products kind of grew in their own silo, because they each have their own business problem that they're solving, and it turns out that what the definition is of a consumer and one application is not necessarily what the definition is of a consumer in a different application. So all these applications kind of grew in parallel.

About eight years ago, we really determined that what was going to happen was that people were going to move increasingly to the cloud and we also believed that the amount of time it took to deploy our applications and also the amount of time it took to integrate our applications was too high. So about eight years ago, we made two significant bets.

The first bet was on the idea of components, that we could go out there and we could take all of our different intellectual property, all of our different algorithms, all of our different rule systems and we could actually wrap them with a user experience and wrap them with decisioning logic and then configuration what are the actual rules you're applying and we could make it so that you could provision that, almost like you're going to an app store right on your phone and you could go in and you can say, “I want a rule server, or I want a scoring server, or I want whatever else.” So that was the first big bet we made.

The second bet that we made was on the basis of the first bet, which is how do you deploy these things. So, again, about seven years ago, we made a big bet on containers. Remember, that this was before Docker. This was before a lot of what's happened since. So we were actually using LXC containers, but we wanted a real containerized architecture that we could take and deploy each of these individual applications with.

The entire goal would be that we’d have a set of services, and those services would provide not only sort of logical pieces like deployability or scaling or things like that, but also business services. So things like being able to capture decisioning services or decision outcomes, being able to capture decision inputs so that you can actually create a learning loop where you would take a particular piece of input in, you’d have a set of input out or data out and you could actually determine what decision did I make and then how could I improve on that decision.

Along those lines, going specifically to the question you asked, we also wanted a set of services that would help us with regulatory compliance. So when a lot of people think about financial services and legal compliance, what they're really thinking about is PCI, for example, which is the set of rules that govern credit card information, or their thinking about Sarbanes-Oxley, which governs a lot of how IT operates or their thinking about HIPAA, which governs a lot of medical information and medical processes.

GDPR is another one of those services that we wanted to have baked into the platform. So GDPR is a set of processes and a set of technologies that are integrated into the platform that all of our other applications at FICO are built on top of. That's really critical to us. It's not enough to be able to go out and say, “Well, we've got a CSV process,” where an engineer can push directly to production. That works really, really well on the startup. That works very, very poorly when you wore a multibillion-dollar bank in Thailand or in Singapore or in Europe or in the United States. You have to have more processes and you also more specifically have to have technologies for those processes.

[00:11:39] JM: Okay. Let's take a step back and talk about GDPR, and then we'll talk about how it has affected you. What is GDPR?

[00:11:47] JP: GDPR is probably the most important change in data privacy regulations in the last 20 years. It's coming out of the EU and it is a specific set of practices and a specific set of regulations as well as principles that you need to apply to be consistent with what the General Data Privacy Regulation, GDPR, expects for anyone who is working with a European citizen.

So GDPR puts a set of practices out there that you must comply with, and those practices include limitations to what data you can process. It includes the concept of consent. It restricts your ability to process data for Europeans to be limited from a set of basically six justifications, if you will. Six basis is what they call them. So each of those bases have a definition that says you may process data for this reason or you can't process data for this reason.

Most of GDPR is focused around one of those basis, which is the concept of consent. You can process data when the customer consents to you processing data. However, there’re actually, like a said, six of them, and a lot of times, as a financial services company, we’re actually not

dealing with it on the basis of consent. We're dealing with it because we have a contractual relationship with somebody, and technically that makes us a processor, not a controller.

So all of our basis as FICO for processing is on the basis of the legal reason or the legal mandate to actually process data for some sort of financial services activity. The example that I kind of give on this one is GDPR has a series of rights. Those rights are not absolute. They're constrained, right? You can have different justifications for processing the data and then it's restricted by those rights.

In order to process the data, you have to have a rationale and then you're restricted by a set of rights that the customer has. So that right and responsibility has to flow through everything that we do. Every piece of data that we collect, we have to understand what is the data we're collecting. What are the rights that the consumer could use that would affect that data, and then what is our basis for processing.

[SPONSOR MESSAGE]

[00:14:27] JM: When you start a business, you don't have much revenue. There isn't much accounting to manage, but as your business grows, your number of customers grows. It becomes harder to track your numbers. If you don't know your numbers, you don't know your business.

NetSuite is a cloud business system that saves you time and gets you organized. As your business grows, you need to start doing invoicing, and accounting, and customer relationship management. NetSuite is a complete business management software platform that handles sales, financing, and accounting, and orders, and HR. NetSuite gives you visibility into your business, helping you to control and grow your business.

NetSuite is offering a free guide, 7-key strategies to grow your profits at netsuite.com/sedaily. That's netsuite.com/sedaily. You can get a free guide on the 7-key strategies to grow your profits.

As your business grows, it can feel overwhelming. I know this from this experience. You have too many systems to manage. You've got spreadsheets, and accounting documents, and invoices, and many other things. That's why NetSuite brings these different business systems together. To learn how to get organized and get your free guide to 7-key strategies to grow your profits, go to netsuite.com/sedaily. That's NetSuite, N-E-T-S-U-I-T-E.com/sedaily.

[INTERVIEW CONTINUED]

[00:16:16] JM: So let's imagine I am building an application like what you build at FICO. How is GDPR going to affect my day-to-day software engineering practices?

[00:16:32] JP: Well, the first thing to know is it doesn't matter if you're in the U.S. or not. If you are collecting data on Europeans at all, you're falling under GDPR, and there are some really, really, really stringent penalties that will get applied to you in an unfavorable way if you don't comply with GDPR. So that's the first thing.

[00:16:50] JM: Have those fines actually been invoked? Have people actually lost money as a course of this? I just like to know is this like a regulation that people actually need to follow in terms of like they're actually going to get financially penalized?

[00:17:06] JP: The answer to your question is yes. GDPR, there have been fines. I believe there's been roughly about $58 million worth of fines right now. Most of that is actually a single fine that got hit against Facebook. The potential fines are much, much greater than that. The way that GDPR is written, the regulators with GDPR can actually go after a percentage of your revenue that matches the percentage of your revenue in Europe. So they can actually penalize you based off of what percentage of global users are European.

So if you're $2 billion business, theoretically they could go after you for whatever Europe's percentage of the global population is for a percentage of your overall income. So that's a very big number. The numbers that have been hit so far are much smaller, but there are a lot of complaints out there, and I think everybody's just kind of waiting to see how it gets enforced. There's been a lot of sort of back-and-forth about is this going to be a really, really big problem

or is this going to be a really small problem? Is this going to be someplace where this becomes an extension of the trade war, for example.

What ends up happening is that European regulators end up going after American business. Is this something that it's going to be a little bit closer to what previously happened with like the cookie law Europe, or with some of these other things like the Cambridge Analytica scandal where the penalties were actually much, much smaller.

I think there're still a lot of unknown there, but what we also are seeing is that there is a ton of sort of these copycat laws that are popping up across the entire world. For example, in California, you've got the CC – The California Consumer Protection Act, which is going to do basically the exact same thing for California citizens. You're also seeing bills in, for example, New Jersey, but you're also seeing them in India. You're seeing them on other places in the world too. So even if GDPR itself sort of stays at this low level of active finds, the entire need for data privacy and regulation is going to become higher and higher and it's going to become more and more of a problem.

[00:19:10] JM: It sounds like at this point, the jury still out on how much penalty is going to be invoked on people, but a company with the size and inertia of FICO, you have no choice but to start acting to some degree. It's such a big strategic question; how much resource do you invest in re-factoring your processes in response to GDPR? Because if you go after it and then it ends up being kind of a set of regulations that nobody actually pays attention to in five years, then it's kind of wasted effort. But of course if you don't act on it at all and then the fine start coming down, you might just get hammered. So how do you calibrate that from a strategic point of view?

[00:19:59] JP: I don't think ignoring it is ever a possibility. I don't think it's a possibility for two reasons. One of which is purely on the basis of good business. The reality of it is, is that we all need to be taking data privacy much more seriously than we have at times. I say that as an industry. You can come up with any sort of justification you need to look at companies. Hit to their stock valuation when they have a data incident. You can take a look at the reputational damage that happens in those circumstances. Frankly, you can also just take a look at it ethically, right? We are all human beings. Do we really want privacy leaks? Do we really want

these kind of privacy problems. Do we want to have that kind of impact on people around us? So that's the first and I think more important argument.

The second argument is that GDPR itself actually has a very clear set of mandates on things that you need to do besides the fines. For example, if you are processing data and you fall under a set of circumstances, you're mass processing data, you're doing something that is going to end up having an impact on the people that you're processing. Should it get exposed? You're dealing with data that has impacted people who might be disabled. You're dealing with data that might have some financial impact to them. There is basically a whole bunch of different qualifiers.

You are required by GDPR independently of this fine, these fines and what not, to have a corporate data privacy officer, or data protection officer. You are also required to have impact analysis on all of your data, right? What happens if this data gets out there So if you've got some set of data that's got some sort of risk associated with it, you need to have an impact assessment on that. Each of these sort of copycat bills that are coming out in places like California and New Jersey that are modeled on GDPR, have different ways that they break the different sets of pros and cons about when you have to have an, A, assessment, and when you have to have a privacy officer.

But the bottom set of trends are clear. You need to be thinking about data and you need to be building in a base set of practices into your processes and into your software that will support you when you're trying to be good citizens with other people's data.

[00:22:14] JM: Okay. Can you give an example of how it has impacted you at a lower level? Maybe from the point of view of an actual software engineer that is – I'm just imagining somebody working on like a Java microservice internally. How is that person's life affected by GDPR?

[00:22:36] JP: It can go all over the place. One of which is let's say that you're writing a microservice that basically is responsible for storing customer data. When you're storing those type of customer datas or retrieving those customer data, you need to be thinking about the fact

that your data store might have to change, because there are a set of rights that are associated with GDPR.

For example, there is a right to be forgotten inside of GDPR. That means basically that you need to either delete a record or you need to have a way of tombstoning a record and making a record impossible to read if a consumer invokes that right. You also have a right to access or a data portability argument. You need to be able to provide the consumer with some subset of the data that you have on them, again, based off of the legal definitions at their request.

So when you're building these services, you have to build these services as if what you previously thought of was your internal data may now be external data that somebody outside may have access to go and either remove parts of or hide parts of depending on what right and what processes that they're invoking.

Another place where this obviously intersects is with encryption. So I mentioned the right to be forgotten. Part of encryption is certainly making sure that private data stays private, but another common use of encryption here is encrypting things in such a way that if a customer demands that their data be removed, rather than going and purging all the data, you simply delete some key associated with that data.

So the data that is in there may still be in your database if you don't have a way to remove it, but you no longer have a way of reading it because the key itself is actually gone. So these type of principles will result in a series of different technical actions you can take. Now, the question is how do you go from the set of principles to the technical actions you can take?

There're a lot of different frameworks for actually trying to resolve that. The most common one is basically a concept called privacy by design. Privacy by design is basically the entire concept of how do you build data protection through technology design. So a lot of the different technical strategies we use will be derived from the core principles and privacy by design.

[00:24:56] JM: My sense is that you have an nuanced view of GDPR's place in the world, because I'm hearing some sense of appreciation for the fact that this is going to improve data privacy, data governance for general consumers. It sounds like you believe that this has been

needed for a while. But on the other hand, as a software engineer, I'm sure you've seen this firsthand, that it slows you down. There's a painful aspects to complying with GDPR and how that affects the software engineering process.

Do you consider yourself a fan of GDPR or do you just considers yourself a fan of the sentiment of GDPR and you're not such a fan of how it has been carried out in practice?

[00:25:46] JP: I think GDPR is absolutely the right set of principles. I think GDPR, the rollout of GDPR has been mostly panic by an awful lot of people. So when GDPR rolled out, what we actually saw was that a lot of people, most of them in the marketing space or most of them having business models that were associated with marketing panicked and actually sort of shut down their websites for anybody from Europe and basically said, “You can't access this because we can't afford to comply with GDPR, or we don't know what complying with GDPR means.”

GDPR is the gold standard right now for data privacy. Like any new regulatory burden, it's got its pros and cons, but it is the law at the end of the day. All of the laws that we’re seeing sort of pop-up are mimicked off of GDPR.

So from that point of view, I think it's incredibly valuable to have at least that first set of rules out there. I think there are still a lot of unknowns. So far, I think that there're actually been a lot of unintended consequences when it comes to GDPR. The direction that I think the market is going in in compliance is a good direction, and I think people need to take this seriously. We’re still waiting to see how all of the different impact sort of falls out. Those impacts are different.

It depends on if you are what GDPR considers a controller, which is, say, a bank dealing directly with a customer. You have one set of concerns that you need to deal with. If you are a processor, which is normally what FICO finds itself, where we are actually processing data on behalf of other people. That has a different set of requirements than, of course, if you’re a subject or a citizen, you’ve got another set of requirements on top of it. So all of these are playing out, but I think on net, GDPR is more positive than it's negative. I think over the next year, I think as these other law start coming into force, I think it will help us understand a little bit more about what we need to do.

Now, I also want to take a step back and say this entire speed to market thing versus regulation, I think that's actually been a little bit overdone. The reality of it is that we do need these kind of laws.

[00:27:59] JM: What's the speed to market versus regulation thing?

[00:28:01] JP: Oh, yeah. So just a moment ago you said as an engineer. We don't always like these type of things, because they slow us down I think was your statement.

[00:28:10] JM: Yeah.

[00:28:10] JP: I think the argument that regulations have an impact that basically slows you as an engineer down a little bit is a bit overdone. In fact, I think it's actually really important to note that for financial services and for companies like FICO, it's actually really important to us that we understand what the laws are and that we build those in as requirements to our products.

Here's one other area that I think GDPR and laws kind of can have unintended consequence. Do you remember the CAN-SPAM Act back in the late 90s, early 2000s?

[00:28:45] JM: I definitely do not remember that.

[00:28:46] JP: Okay. So I'm definitely graying very quickly, but I've been in the industry long enough to remember that there was a law that got passed called the CAN-SPAM Act, because spam email was becoming an issue, and it put a whole bunch of additional requirements on top of anyone who would basically send out email out into the outside world.

The interesting bit about that was that what ended up happening is it actually eliminated a whole tier of people who had been sending out sort of bulk emails directly from their companies. Somebody was running an SMTP server that was actually blasting people directly with email in their own data center. What ended up happening is that the regulatory cost to comply with it was that it actually kind of eliminated a whole bunch of sort of smaller players who were in this space and sort of players with larger market size could afford to spend the money to be compliance

with these laws. So it basically forced most of the direct email advertising into the mail service providers. So companies like SendGrie and whatnot.

I think you may see something very similar with GDPR and with these data privacy where the cost to actually maintain information about your customers may become more and more expensive and you may increasingly be outsourcing to people who have the ability to actually pay for that regulatory compliance and who can stay on top of it. So that is actually the tradeoff that I think may end up being more significant over time.

But in both cases, the outcome of CAN-SMAP was that the amount of spam email dropped dramatically and hopefully the outcome of GDPR in these type of laws will result in consumer's having their data expose much less often.

[00:30:28] JM: That's a very interesting perspective, because it's basically acknowledging – Well, a term I've heard called regulatory capture, where by the force of regulation, existing incumbents are able to actually become stronger. So Facebook and Google will actually become stronger as a result of GDPR, because that's the size of a company you need to be to actually comply with these things.

I was reading that article about the consequences of GDPR that I'll put in the show notes that was talking about some of the unintended consequences of GDPR, one of which was as soon as GDPR went into effect, a bunch of advertising flow went from smaller players to Google, because people were confident that Google could actually comply with GDPR even though the intuition might be that you think like, "Oh, GDPR is going to hurt a company like Google."

So the calculus that I heard from you is kind of this is going to strengthen the large players, but more importantly, it's going to benefit the consumers because they're going to force traffic to be routed through the central players who have the resources to actually comply with these regulations.

[00:31:44] JP: Yeah, and it's certainly a possibility. I think the jury is still out on that. But certainly I think that there was – Particularly, to give credit where it's due, I think there was an ad tech analysis that clicks and Ghostery earlier on, and I think they found out that Google basically

was up, but it was actually a pretty negligible percentage. It was up like about 1%, if I remember correctly. It's also worth noting that it's not necessarily just on size. It's also on ability to comply. So I think that we've also seen sort of a significant decrease in Facebook at the same time.

I don't think you can necessarily say it's always about size, but I think you can say that it basically makes it so that people who can comply and choose to differentiate on that compliance have an advantage over people who can't comply or choose not to differentiate on that compliance.

[00:32:30] JM: Do you think this is actually something that is analogous to spam? Spam seems to me to be a little bit more clear-cut. What is spam and what is not spam seems a little bit more clear-cut than what is privacy and what is not privacy? The points of granularity for spam are like if I get a newsletter from a company that I use on a regular basis, like if Amazon sends me a newsletter. Is that spam? I mean, it's not something I really want. I don't really want to read it. They're just trying to get me to buy stuff. It's spam, but it's from a company I trust. So there's some kind of like granularity there.

The world of privacy is so much more granular. I wonder if this kind of regulation is even really possible to adhere to. Why not an alternative model where just when – I don't know, when something really bad happens, the company just gets hammered with a big fine in a post hoc way. I mean, I have no understanding of the law. So maybe that's reflective of my poor understanding of the law, but I guess I'm just wondering.

So I think about like when I go to websites, I now see the little notification at the bottom that says something like, "Your data is being collected. Please click this checkbox, and like here's a big terms of conditions," and you just ignore it basically. It has no actual impact on my life. It's just like this little added annoyance. I just wonder, is this actually going to practically positively impact people in the world?

[00:34:04] JP: Yeah, and I think it's a great question. I think there're two points. I think the first point I'm just going to make generally on laws and how we apply them. Again, this is sort of dangerous territory for a software architect. But I think it's worth thinking of as an architect,

because I think the ethics are important here. Then the second point I'll make specifically about GDPR.

The reality of it is, is that we do have laws and those laws, and those laws are trying to come to some sort of improvement or some sort of change in the equilibrium. Sometimes they're good. Sometimes they're bad. Sometimes they have unintended consequences, etc. But for the most part, I think it's really important to understand that data privacy, while it's being acted on by these laws, is basically a fundamental requirement.

The Indian version of GDPR that is coming out all started with the Indian Supreme Court more or less saying that data privacy is a fundamental human right. So it makes sense that when something is that important, that we capture in the form of laws to try and make sure that we've got a framework for making sure that things are done right. Then as architects and software engineers, it's important that our software reflects that, because some have made the argument that software is intrinsically immoral. It's not necessarily natively good or natively bad. It still basically does a task at the end of the day, and that task being protecting the privacy of others is a really good thing.

Specifically on GDPR, I think one of the really great/infuriating things about GDPR is that it's actually written pretty open-ended. There are a lot of things that are pretty expansive and t that it's not completely clear how far they actually apply. The interesting thing is, is that I think that's a good mechanism for actually dealing with some of the uncertainty that you've just brought up.

We didn't know about things like Cambridge Analytica before the last couple of years obviously, and the fine that was associated with Cambridge Analytica was relatively small, because legally that was the only framework they had and the only amount of money that they could go after.

So something like GDPR or the California Bill or all these other mechanisms, something that's really important about them is they raise the stakes. They basically say, "No. You don't own data. Consumers own data. You have the right to process it in certain circumstances. You have the need to interact with customers and allow customers to actually express their will via a series of rights even when you're holding that data."

That is just kind of a fundamental sea change in how we should be thinking about data. So that to me is the most important part of GDPR. The rest of it in terms of the specific of the law, the definition, I'm not a lawyer and I'm not a data scientists and I don't play either of them on TV. So all I can say is from an architecture point of view, these are sort of the basic capabilities that we build into our platform to try and support these requirements.

[00:37:07] JM: Yeah. I mean, I can respect the idea of wanting to codify data protection norms into laws even though there is ambiguity. Because by doing so, you signal to the industry that this is a change in societal norms. If we trust that our governments are reflecting the desires of the citizens, we are embodying those into a loose set of societal norms that are getting codified into laws and it's going to be imperfect. It's going to be slow to change, but I suppose what other option do we have?

[00:37:50] JP: Yeah. As you note, it gets into ethically murky areas really quickly, because there definitely are governments out there that do not necessarily act in the best interests of their people. That's always a challenge. But I think software engineering, just like almost every other form of engineering, is engineering where ethics apply. These are things that I think you as a software engineer or as a software architect or as a manager or whoever else do have to think about in terms of what you're doing, and you need to factor it in to your architecture just as much as you would architect in Kubernetes, for example.

[00:38:26] JM: So you are a software architect, right? That's your role at FICO. It sounds like you almost are – You have enough appreciation for the regulation. Maybe you can actually enjoy this added constraint on your software architecture. I mean, you're no longer thinking about solely the distributed systems capabilities or the persistence qualities of your data. You have to think about this regulatory component. Is that interesting to you as a software architect? How does it change your life as a software architect?

[00:39:00] JP: It's actually fascinating to me as a software architect. So one of the interesting things that we do with our software architecture, I mentioned earlier that we are running on top of the containerized architecture. Originally, we're using LXC for that. Now we're using Kubernetes for that. We're supporting Lambda cases like that.

The majority of the world I think out there, when you take a look at sort of the best practices for deploying, people are taking a look at CICV processes where people can do kind of this Facebook-ish flow where a developer makes a change and then tests that change and then pushes that change out into production and says – Has some sort of canary build where they say, “Okay. I want 5% of my traffic to go to this particular service, whatever it is, and then I can watch it for a little bit. Then as soon as I'm satisfied that it's working or it’s not having unintended consequences, I’ll roll it out to the entire population.”

From a financial services point of view, that type of approach just doesn't work, because there are all sorts of regulations around scores and around business processes. For example, you're not allowed to use the ethnicity of somebody in making a credit decision. Why would somebody want to do that? Well, because, frankly, we live in America. You can make an argument that certain ethnic groups are better credit risks than other ethnic groups, right?

So we have these set of laws. We need a set of business practices, right? So things like the FCRA, to actually control the legal ramification of technology. Then what you do is you say, “Okay. Given that I can't use this, I need a mechanism for somebody to be able to check and make sure that I'm not going to roll out a decision that is illegal, right? I'm going to have some way for somebody else to come in and actually review that decision, say, “Yes. This is compliant,” or “No. This isn’t compliant,” and then approve it and then I’m going to have another person come and say, “Okay, this is ready for deployment, or it's not ready for deployment, and here's the scenario that I’ll deploy it in.” Because another thing that we have to do under Sarbanes-Oxley is we have to have a segregation between developers and production so that nobody can go in and do something that they shouldn’t be able to do either for fraud reasons or for money laundering reasons or for whatever else.

All of those sort of legal conditions become software architecture. For example, when we rollout a decision service, we create an instance of it that's in a design time state. You’ve got your data scientists or your business user who goes in and sets up the rules. Then they promote it into a staging environment where we create another container with the exact same configuration as the first container. Then another user comes in and they review it and they say, “Yes, this is legally compliant, or not meets our business goals or whatever else,” and then they submit it for

approval. Then the third user can come in and then promote it to production if you need that kind of process.

All those steps are completely optimized. If you had everybody in a room and everybody was looking at it, you could rollout that process in a matter of minutes, because it's all completely automated behind-the-scenes. But what we're actually enabling is regulatory compliance, but also frankly just plain old good deployment practices for any of these decisioning services.

[SPONSOR MESSAGE]

[00:42:19] JM: As a programmer, you thinking object. With MongoDB, so does your database. MongoDB is the most popular document-based database built for modern application developers and the cloud era. Millions of developers use MongoDB to power the world's most innovative products and services; from cryptocurrency, to online gaming, IoT and more.

Try MongoDB today with Atlas, the global cloud database service that runs on AWS, Azure and Google Cloud. Configure, deploy and connect to your database in just a few minutes. Check it out at mongodb.com/atlas. That's mongodb.com/atlas.

Thank you to MongoDB be for being a sponsor of Software Engineering Daily.

[INTERVIEW CONTINUED]

[00:43:15] JM: How do you think that GDPR impacts the world of startups?

[00:43:21] JP: It depends again on what those startups are actually doing, right So GDPR breaks things down into two really big categories if you think of it. You can think of controllers, the first category, as people who actually own the relationship with the customer. For example, a bank and their relationship with a customer. You can think of a second category as processors, right? So I'm, say, a targeted marketing provider. I'm going tell you what offer you should give a particular customer based off of the basket and analysis of that customer. So those are processors at the end of the day.

What you have to do is different depending on what your relationship is with your customer or with your customer's customers. Specifically, what do you have to do to make sure that you always are keeping the customers key to private, and there're two ways to do this. One is that you can build out all these processes. But the second is you can do everything within your power to minimize the data that you're taking in.

So the way I like to describe this strategy is I think everybody's familiar with the concept of technical debt, right? Every line of code that you have basically is a liability, but it generates some revenue associated with it. So to get what revenue you need to get to run your business, you have to run your software against it and each line of software is a liability, because you have to maintain that software. You have to make sure that line of code is working properly, etc.

Think of customer data exactly the same way. Every record that you have is a liability. There's some value to customer data, right? Data is the new oil, right? But it's also a liability. The more data you have, the more you fall under these different types of data protection regulations. Because of that, the higher the cost of maintaining that particular line of code is.

Another way to kind of tackle and to minimize your impact as a startup is just don't collect the data unless you really have a need for it. Then if you do have a need for it, understand exactly what your need is for it and can you strip it of anything that might have bad consequences for you. If the answer to that is no, is what are the practices that you're building into your business? What are the processes? Do you have a privacy impact? Do you have a data protection officer? Also, what are you building into your software, right? Is all of your data always being encrypted? Do you have standard ways of getting out data that you can expose it to the outside world? Etc. So it's kind of gradual depending on what you're doing, and there are strategies that you can use both minimizing your data and then of course privacy by design to try and restrict the impact to you as well.

[00:45:54] JM: Give a few more tactical suggestions to people who are listening. Maybe they work at a company that's large or small. How should they prepare for GDPR? How should they change their internal strategy, their org structure, their software architecture?

[00:46:11] JP: So that's actually a much bigger question than I'm probably qualified answer. That's actually probably something that you'd want to pull in a DPO, a data privacy officer, to really talk about that, because there's actually a set of legal requirements around that.

So from a purely business-oriented point of view, that's where I would point at. From a technical point of view, I would actually say treat every single record that has anything to do with a person going through your system as if it were your Social Security number, your credit card number, your name, your children's name, etc.

If you start thinking that way and then you start thinking about the steps that you take to protect your own identity and you start applying those mechanisms to your customer, I think you're going to get into a much better place very, very quickly. It's really easy to look at a database table and do a select* by name count and say, "Oh! I've got 25 million records in my database." If all of those are basically some bit of information about you or some bit of information about your wife, your kids, your family or friends, etc., I think you're going to take a very different view of it.

So start with shifting your thinking about how you view data. Then on top of that, take a look at what you have to do to make sure that your data sources are private. What do you have to do to make sure that you have ways of exporting that data for data portability? What processes and what code do you have in place to make sure that people are not accessing what they shouldn't access? But then also understand that building in those mechanisms also makes you more vulnerable.

For example, we've already seen attacks where people have basically filed a data subject request, basically, give me all of the information you know about me for somebody else as an attempt to get into their identity. So think through those type of permutations. Think through those flows and then say, "In software, what processes? What code? What systems do I have in place to protect myself from that?"

[00:48:04] JM: What are your predictions for how data policy will evolve in the next five years?

[00:48:09] JP: Oh, man! If I could do that, I'd be running my own consultancy somewhere inside of Washington, D.C., right? There's my optimist side and then there's a very pessimistic side. My optimist side is that people take data privacy seriously that they get better at self-regulating and controlling their own data processes that we don't have any more data breaches, that those data breaches aren't catastrophic. That we don't have anything more like some of the stuff that's happened with Cambridge Analytica, that we don't have sort of these targeted nation states going after data sources. That's kind of my optimistic world. Unfortunately, I don't think we live in that world.

I don't think my pessimistic view is correct either though. I don't think that it will continue to be businesses as usual. I was talking many years ago with the corporate marketing officer for a very, very, very large beverage and foods company. One of things that they said then, which actually kind of stuck with me was you as a company never own your brand, right? You never own how customers view your company or what your company stands for out in the outside world. It's always in your consumer's hands.

I think we have gotten to the point where we're thinking in the same way about data. You don't own the data. Your customers own the data. You have legal rights to use that data. You have legal rights to manipulate that data. You have legal rights to act on that data either via consent or any of the other basis for processing data. But at the end of the day, it's still the customer's data and we have to think about it in that context.

[00:49:45] JM: Well, Joshua it's been really fun talking to you. Do you have any final parting thoughts on data privacy and your work?

[00:49:53] JP: This is one of the reasons I really love being in the enterprise software business. It's because I do think that there are a lot of business models out there that are really designed to profit basically when you're leveraging customer's data and you're extracting every single value out of it for as little overhead in terms of regulatory compliance or whatnot as possible.

Thankfully, I don't think that's the world that we live in and I don't think that's the world that financial services live in, right? We are there to do good things for our customers. Therefore, we

can afford to spend the time and the money and the attention on protecting their data, and I think that that's something that personally I actually really love about my job.

[00:50:35] JM: Okay. Josh, thanks for coming on the show. It's been really fun talking to you.

[00:50:38] JP: Thank you.

[END OF INTERVIEW]

[00:50:48] JM: Cruise is a San Francisco based company building a fully electric, self-driving car service. Building self-driving cars is complex involving problems up and down the stack. From hardware to software, from navigation to computer vision. We are at the beginning of the self-driving car industry and Cruise is a leading company in the space.

Join the team at Cruise by going to getcruise.com/careers. That's G-E-T-C-R-U-I-S-E.com/careers. Cruise is a place where you can build on your existing skills while developing new skills and experiences that are pioneering the future of industry. There are opportunities for backend engineers, frontend developers, machine learning programmers and many more positions.

At Cruise you will be surrounded by talented, driven engineers all while helping make cities safer and cleaner. Apply to work at Cruise by going to getcruise.com/careers. That's getcruise.com/careers.

Thank you to Cruise for being a sponsor of Software Engineering Daily.

[END]