Privacy Tech, Data, and Nuclear Waste - Dan Bogdanov, Head of R&D at Cybernetica -E53

Eric: This is Eric Hess with The Encrypted Economy, and on this week we had Dan Bogdanov returning to the podcast. Dan is the head of R and D at Cybernetica, Professor at the University of Tartu, and the co-inventor of Sharemind, a secure privacy enhancing technology platform. So we started to move in this episode into privacy enhancing technologies and we cover homomorphic encryption multi-party computation, zero knowledge proof,

functional encryption at PGA or programmable integrated circuits, federated learning, neural networks, and digital identity. We also talk about Sharemind, Cybernetica's secure computing data analysis platform. I lit off the second episode by asking them how privacy enhancing technologies could change the ability of businesses to outsource functions that previously they would not have due to their highly proprietary.

If you haven't listened to my earlier episode with Dan Estonia: The Rise of a Digital Nation, I encourage you to do and of course, if you enjoy this episode, I encourage you to share it with others. And with that, I bring you my second episode with Dan Bogdanov on privacy tech, data, and nuclear waste.

Welcome to The Encrypted Economy, a weekly podcast featuring discussions exploring the business laws, regulation, security, and technologies relating to digital assets and data. I am Eric Hess, founder of Hess Legal Counsel. I've spent decades representing regulated exchanges, broker dealers, investment advisors, and all matter of FinTech companies for all things, touching electronic trading with a focus on new and developing technologies.

Dan: Some of these technologies are very suitable for outsourcing, meaning like put stuff, put workloads on the cloud. Homomorphic is very good for that because assuming that you want to use a single cloud provider, then you just stuff everything onto your cloud of choice homomorphic, fully encrypted, and then they run all the workflows using homomorphic encryption and all is good.

Multi-party computation requires multiple non colluding parties, meaning that putting everything on a single cloud is counterproductive. You're making the cloud a trusted third party. And that means that there are outsourcing technologies which are technologies, which are very good for outsourcing like homomorphic and encryption.

And then I would argue that trusted execution environments are also pretty good for outsourcing. And then there are these partnership technologies where you have distributed control. Whereas, if you put stuff on the cloud, then you assume that the cloud has certain controller, they can run stuff on your behalf.

And these consensus requiring technologists, like MPC, people need to have hands-on for things to run, meaning that your partners are symmetrically. Dependent on you and you're dependent on them, unless you all agree that you will run this, then you can't run this. And that's a whole separate type of technologies.

It's if you try to do partnership things with homomorphic encryption, then you run into a situation where somebody is still more king than the other. You can emulate it on trusted execution by using specific architectures like remote at the station. And so on. We do that with share mind where you just, before any calculation can start, then everybody remotely attests and get received approved.

But this is the algorithm were. It's middle ground. There is still one party who is running everything, but they can't draw anything without everybody else saying yes, I fully agree to what's being done in the future. May be homomorphic. Encryption will also gain similar capabilities. Functional encryption is something that has been designed for that purpose, but they're not fully integrated yet.

Eric: Are you talking about ZKP, Starks, Starks.

Dan: No. The, the things you mentioned in that list was mostly all zero knowledge proofs. And we're also working quite a bit on zero knowledge proofs. We are performing research in this dark zero knowledge program. That's currently ongoing where we are focusing, especially on e-government.

And private public interactions and using zero knowledge in interactions between governments and citizens and companies. So for example, currently we are building a system that would allow you to pull in your health records from healthcare providers digitally signed, and then prove in zero knowledge to.

I don't know, think of it. Like maybe the military recruiter that your health records correspond to a certain standard. They won't know exactly what's in your health records, but they know that this proof was put together from digitally signed health reports from the. Ballad health care providers. And the whole sort of a flow of trust is preserved by using zero knowledge proofs.

And that is again, an interesting model we see that supports a distributed. Data architecture for a government or a large corporation that you're pulling data in from multiple sources of data about yourself. And then you're making the proof about this data instead of just passing this data on for someone to read in the clear, you're actually constructing a statement, which says that I had this data.

It was signed. I ran this code on it, calculating the results of this algorithm that was standardized. Maybe you provided me with that algorithm and here are the results of that algorithm. The other, the verifying party looks at the results and says that, okay.

Understood. Yes, I indeed. This is irrefutable evidence that you have calculated this result from this data and because it was digitally signed, and I also have a zero-knowledge proof of the signatures.

Then I have to assume that based on information from these sources. This is true. And again, this is going to create interesting new transactions between participants in a digital society that we hope will be there in future versions of the Estonian government or other governments in the world.

Eric: And when you were talking about functional encryption, what were you talking about specifically?

Dan: Yeah. Functional encryption is specific cryptographic construct that you can use. Again, that you can calculate a certain function. You can give somebody data and then they can run a certain function on it.

And there are limitations on what they can calculate. It's basically limited to that. That's a very over simplified version of it. But the thing is that mathematics is giving us a lot of primitive. Fully homomorphic encryption, multi-party computation all of these things and converting them into real world platforms that you can just go and use on your device is something that takes hard work.

And one thing that I am very thankful for is that the crypto asset and cryptocurrency community has given us a lot of very good research on MPC, threshold, encryption, and also blockchain technologies and hashing, things like that. This has been something that pushes the world forward. Then these days if you're a cryptographer and you are doing zero knowledge proofs or.

Cryptographic research, then there are a number of startups. In the FinTech or payments domains that are willing to hire you and you can go in and you can do research you things. I would say, obviously it's assumed that at the end of the day, you will be building a service. You will be building something, but people understand that the technology isn't there yet.

And for some interesting reasons, Venture capital is willing to put money into things which I would argue is still in the horror and the stage.

Eric: And I think that's going to continue. Uh, you know, particularly with some of the AML KYC verification that is now becoming, you know, FATF and governments are requiring more AML upfront, even impacts decentralized exchanges.

And so in a decentralized exchange, basically it's very library. So how do you ensure that these checks are being performed while still facilitating the transaction in a lightweight way

with a, a fully decentralized exchange? That's basically just. Without people running around.

So I think the for cryptographers, the field is only growing because this issue of how to actually include, leverage personal information, maybe secure multi-party, maybe through zero knowledge proofs. Whitelisting and checking identities, but in a way that doesn't slow down the transaction in a way that doesn't allow a Dex, which is lightweight to now have to take on both the obligations of a compliance department.

And now that you've taken in this data, now you also have privacy concerns. So you've taken it. And they're like, great. I've done AML. No, no. Now you have all these privacy regulations that you have to comply with because you took this data. And so it's, you it's poison, right? Like it's the last thing you want is not one regulatory regime.

You've just inherited two. And probably there's a few more that, that come from that. You know, now you're just going back to a fully centralized organization. So depression. To find lightweight solutions to facilitate. This has just increased and Krypton central is central to that, figuring out ways to, to both remain on the public blockchain, but yet privately sharing this information in a way that's trusted, but it doesn't slow down the transaction.

So exciting times for photographers,

Dan: the issue is going to be pretty much key because a lot of the constructs. Pretty heavy when it comes to math and processing needs. And obviously all the chip makers are going around then the opening champagnes because every time chip makers, what's the incentive for chip makers to make stuff faster.

Not everybody is playing computer games on top of their new fancy. Graphics card. So in order to just run a, even to run this call, I don't need a very big investment in equipment that routes running on top of a Mac book air, which is not the fattest laptop in town. So the question is how do you keep selling cool new, powerful chips?

So you need to add new features and security is indeed one of them. So building all kinds of cryptography and security technologies, privacy technologies into heart. Is going to be an interesting direction and we already see how Intel has gone this way with SGX AMD, with ACV. And this is being embraced by cloud providers, Microsoft with a lot of its con confidential computing work.

Also IBM cloud provides SGX boxes and there's a, it's a developing field for sure, also in hardware. And now the question is with software. If it's source or you have the source code, you can easily go and audit a hardware question is always going to be, can you trust the hardware? Because nobody's capable of going down into a five nanometer process and then reverse engineering.

It not okay. Not no. But it's very expensive and not fully accessible to everyone. So trusted hardware is going to be interesting as well. Whether a chip that promises it is doing accelerated fully homomorphic encryption is doing that, or it's also on the side not doing it. So now we're doing something else trying to analyze it in some way.

That's an interesting question. Hardware makers will need to answer in the future because putting these technologies from papers for research papers, fundamental research to the field is a significant undertaking. So the math we need. For building our uh, X road and share mind and these smart ID systems.

It was very, we had to develop the math a little bit and massage it into protocols, but when we had to go and build real world systems around it, with all the economic considerations deployment architectures fraud detection, obviously as well, just as you mentioned, is going to be injured. Regarding the fraud and the regulation things I do, whoever came up with the thing, that data is not the new oil, it's the new nuclear waste.

You want to get rid of it at some point, if you used it, then at some point after that, you need to get rid of it somehow because it becomes toxic. So very much agreed with what you said there as well. I spent

Eric: a while since I've considered the hardware integration used to work a lot with wall street firms and developing, trading algorithms and years ago there were people were building chips with integrated algorithms, built into the checks.

Part of the problem of course, is you got to be pretty sure you want to stick with that algorithm. Cause it's hard to change. He used to go a number of years ago, a pen system for modularizing components of the chips so that you could actually modify the hardware on a modular basis. There's this whole research paper I read.

And if I even tried to do it justice, I read it. I understand that conceptually, but I could never explain it. Uh, but it was really about, I just remember it was like a magnetic polarities and that's the way you would modularize the ship, be able to change the magnetic polarities to you know, changed, it wasn't like.

It didn't give you a whiteboard to start again, but you could basically take existing code and through different magnetic components change you know, modularize that functions.

Dan: It would be interesting one day if maybe, or programming. Comes down to neural networks, just setting the parameters for the neural network, but the chip is just a big neural network and you just set the weights and set the layers.

And that might be something where this works, but if you're doing hardware, then obviously there's. Whereas the trade-off of how quickly you want to go onto market. Then how many units you want to build? Because you can get SPGs programmed pretty quickly,

but they're not as performant as a six. And they six might not be as performant as fully cooked circuits on transistors and so on.

So there is the economic aspect to everything cryptographic. And this is one thing that I currently also see. A lot of cryptography gets a lifeline, although economically I'm not sure it will survive. And that's something where venture capital also is creating an interesting situation and where the quantitative easing and the availability of money on the global market is creating a situation where it's not sure that all the technologies we currently have, and all the coins and tokens systems are going to be sustainable in the long run.

But we are learning so much about them for now. And this might inspire something that we will actually build a need in the future. So I'm very excited about all of that work. So trying to keep our eyes open at all times. So we've

Eric: talked a bit about a few things we haven't talked about, share mines your project.

You've noted it, but let's see. Let's dig into it a little bit. What are the technical aspects of share my, that differentiates it from another end to end encryption?

Dan: So think of it like this secure channels on the internet between you and your online bank, they ensure that you cannot, but nobody can listen into your discussion with the bank.

If you will send them a transaction that you would execute it, then you will put it on the child. And the bank will receive it. And it's, there's a pretty good level of trust that nobody has changed the transaction, and nobody has eavesdropped on it. Share mine tries to do the same thing with analytics.

You put the data owners, put the encrypted data in on one end. And then on the other end, there is an analyst who gets the results out. And the whole process from beginning to the end is under increased. That's the simplest way. That's what we want to do. We want to do end to end encryption for data processing and analytics and a share mind is considered as an application server, which runs your business, logic, or queries with the data.

On their protection, meaning that even if you are the host of a share mind server, then you do not have access to the data. It's even worse. You can't do everything you want with that data. Because as I mentioned previously, either using hardware at the station or multi-party computation, the consensus protocols, you are restricted from doing anything you want with that.

And once you've gone through the processing as directed by the algorithms that have been deployed on your version of, on your instance of share mind, then you disclose the results to someone who has the keys or the ability to reconstruct the result. So that is the

core idea. And what we believe is to be useful for is all kinds of data-driven services that want to operate.

In a way where they don't have to be trusted by all their users. So imagine that you want to start something new. You want to pull together something that nobody has done before. You want to get all the health records about all the people you want to run an AI model on it. You may be wanting to find people who are at risk of.

Disease, but the model doesn't work for you in that way that you can go to every person and say, Hey, this is my algorithm. Please run my algorithm. It's not like the common model where you have a check. It's a self-check. I go in, put my data in, and then I get a result out whether I have a risk or not.

If you're a government these days, then you would like. All of this knowledge is algorithms to be ex run on everybody's data. So you could intervene before diseases get more complicated before they get more expensive to treat, for example, and for that, you suddenly get the need that you need to pull together and integrate and link databases, which not everybody might be comfortable.

Doing, especially if you're doing like population scale things and share mine, there is a solution for that use case among other things. Then there's also other kinds of analytical projects. For example, one thing we did in Estonia with a local tax data and education data was we set up a share mine instance where information was put by the tax office on.

Students who had enrolled in it curricula and then non it curricula over a number of years. And at the same time into the same share mine instance, data came from the ministry of education about their study studies at the universe. The idea was that, can we understand what's the sort of a return of investment on it?

Studies, do they do the, it students actually graduate in time? What do they do with this? So and are they in some way different from the non 90 students maybe who went to study sociology or languages or biology, physics, or something like that? And what we achieved with. We were able to have two agencies with a distributed silos, encrypted data on site, put it into a shared environment where only the statistics that had been agreed to would be run.

And the statistical output was provided, which was the report on students, graduation, behaviors and earnings and things. And at no point, was anybody able to look at the confidential tax records or the somewhat less confidential education records? So it's a sort of dream of, linking data together.

And analyzing it and learning the insights without having to worry too much about privacy. And also last year 2020, where there was a landmark case from the court of justice of the

European union, which is like. And let's say it's the highest court there is in Europe. And it was about the data transfers between the EU and us.

It was yes, it again, how to do with max Scherms and his work on upholding the GDPR actually. With, uh, big American tech companies. So the decision which pretty much said that the European American data transfer regime is not sufficient for actual protection. So you have to figure out something else.

If you want to continue transferring data from Europe to the us. And that created a mayhem in health research and in various domains, which had relied on previous legal regimes for. Working across the Atlantic ocean between American and European partnerships. And obviously very quickly after that the agencies self-data protection in Europe and all the other people started calling.

Up with solutions that actually multi-party computation has been proposed as a technology for cross border collaboration by one of the agencies, I believe it was a European data protection board or the European data protection supervisor. One of our statements from November 2020 actually suggested that multi-party computation would be such a way.

So again, this is something that's pushing the technology adoption for. Because again, we've tested MPC. We've tested, share mine drowning in Europe and us data centers like across the Atlantic, MPC being done, this worked, and it didn't even cost an arm and a leg. So that was nice as well. We certainly see that share mine.

We'll have a part in the future of building systems for AML anti-money laundering information sharing. We are working on that with some partners on cybersecurity, threat intelligence cybernetic are also has an R and D collaboration with the air force research laboratory in the United States on building a cybersecurity threat intelligence exchange network between.

Estonia and the U S that is that is a very interesting project. And consider that if we pull that off, then we there's little things we can pull off because getting defense organizations to exchange data across borders, even among our lives, as we are both NATO members, it's still a challenge. So might be that we are able to really improve partnerships of nations and countries using technology.

Yes. Yes, exactly. Create blueprints, create a blueprint. And that's something where the goal is with share mind in. I've also been interested in it. They're doing a health research, could be improved using this sort of technology, but what we also see, and that's actually now about MPC on the broad market.

So sharing. It's an NPC system designed for analytics. It has a prebuilt library for running progressions, certain machine learning algorithms and so on. You can put it into mobile

apps, web apps, a lot of that, there's, it's a pretty good system now. And who else is doing MPC these days? Facebook hasn't their own MPC framework.

You can find it. I think also open source. Google has very joined and computers. Apple has where a MPC system, which they plan to roll out this fall in iOS 15, that's the, somewhat controversial, child sexual assault material countering system, which was supposed to scan iCloud photos. And then.

I compare the hashes to a known list of legitimate material using secure computing, like MPC, like techniques. So MPC technologies are being embraced by large companies. These days, I, we haven't even talked about federated learning, which is deployed in every Android phone. So I would say that the maturity level of privacy technologies is quickly growing.

Not every application will need share mine. Some of them will go for a set intersection. Some will go for homomorphic, some will. It's something which we haven't even heard about yet the paper's coming next month, maybe. But the point is that going for technologies that allow you to trust your service provider less is I believe the next step in the maturity model.

It's not fully against the cloud providers as well. I'm totally certain that a cloud provider, the big cloud provider providers of today are able to provide a more secure server experience than something running in your broom posit or in some of the local small scale data centers, which don't have all the same levels of control because of the economies of scale.

And at the same time. So cloud used to be the thing that you scared people with. Oh, it's not secure, but that's not the point. Really a cloud can also embrace these technologies and they can offer them for you. And many have, as we mentioned before, with the. Microsoft and IBM clouds and I believe AWS heading in that direction, Ali clouds as well.

So I do firmly believe that in time we will be able to have end to end secure analytics workflows, and we've shared mind. We are trying to create one technical. Which is battle-tested, it's been deployed on multiple clouds across the ocean by governments and companies. And this is our contribution to that ongoing.

And

Eric: you mentioned federated learning, do SMPC applications these days or frameworks have that federated learning integration with component tool?

Dan: No not really. Federated learning is still a totally different idea. Whereas NBC technologies assume that the data is put into the protocol.

All the data is put into the protocol. You have a specific application, then federated learning is a soul. It's a form of machine learning, which allows you to keep a lot of the data on premises at different sites. And then you do certain amount of local processing, and then you go. To a central note, and then you do a certain level of aggregation there, you put together the next layer, next version of a shared model, and then you go back and then you continue processing on site and then you again, come together in the center and then you put together a shared model and you go back into do this iteratively so that you don't put all the data, all the eggs in the same basket.

Okay. So I believe Android phones use it for keyboard prediction, machine learning for keyboard prediction, and that's actually rolled out on billions of devices. So if you're looking for a good example of a privacy technique being widely deployed and that's that, and. The other technique, the other a great example is we contact tracing exposure notification system that Android and Google rolled out for COVID-19, which also allows you to do processing of data without having anywhere a central super database of who has COVID and who hasn't.

So it's a really elegant design parts of it come from MPFL university in Switzerland, but it has been modified. Apple and Google engineers. And it was also a great show of collaboration and standardization because apple and Google phones are able to do this intraoperatively, meaning that apple phones can alert Google phones and vice versa, which is not always the case for, um, I don’t know things like, I message wasn't instantly available on Android phones and so on.

So that is a, that was a great show off again. Partnership. Yeah. And

Eric: in what would be an example where an organization might use an SMPC framework together with a federated learning framework?

Dan: You can actually combine them brother. Interesting ways. NPC frameworks can be the. All of the federating learning frameworks, and that allows you to do some additional privacy tricks on this federated learning model.

So there is this thing called differential privacy, a technique which allows you to add statistical noise to the outputs of a query in order for a potential attack. Not to be able to understand what the input data was because sometimes by looking at the outputs, you can have a bit of an idea what the inputs, but differential privacy makes that a lot harder.

So federating learning, federated learning techniques, the common ones. They do still leak the model and sometimes machine learning models, leak information about the input data so that you combine federated learning with MPC. Then you might be able to use a differential privacy like techniques to throw in some noise and make sure that a machine learning model is you're building.

They don't have clearly encoded input patterns in them. So for example, if you overfill. A machine learning model. It might actually contain one of your input data records as a sort of an example. Because machine learning tries to cover every input, it hasn't in the worst cases, it just says, okay, so 99% of the data I can fit with this.

And then the 1% of the data, which I was forced to include, but it doesn't fit into this nice big model. I just keep it here as a, this was an exception and I'm just keeping it here. I will say yes to this, but that also means that it's encoded in the machine learning model one input record. And that might be leakage if it's encoded in that model.

And if that was private data, then. Yeah, you might land the in hot water by having such a model. So combining federated learning with multi-party computation allows you to add potential extra features in privacy features to your machine learning models.

Eric: So as you're building out, share mind, uh, I guess you have a fully functional product, but where do you take the product?

From here, like w what do you see as the future for share mine?

Dan: The future for share mine is that we need integrations. We need to make it easy to use, because today we provide SDK. We provide some ready-made tools already. So you can get a statistical environment up and running very quickly, but we understand full well that potentially users of share mind might want to contain some of their existing environments.

They might want to keep on using the interfaces they currently have. So integrating, share mind into business intelligence system. Whereas systems statistical systems some way which you use to collaborate, maybe there should be a share mine version of, I don’t know, doodle or a survey system. We actually have built a service system on top of share mine, which can run on the cloud as a publicly available commodity service.

It's doable and that's, I think where we need to go, we need to build solutions. We've got the technology. Now we need to go and start targeting individual. Business lines, individual solutions built on top of it. And then we will get to like really wide scale roller. And

Eric: I, we had turn roll off from duality technologies on the podcast a while back.

And he, he, his comment was, he goes, my mission is to make homomorphic encryption boring. And what he really meant was like ubiquitous. And he, where he said, I don't want people to be talking about whether it's homomorphic encrypted or secure multi-party or not in a few years, it'll just be something that, of course it was.

Why wouldn't we do it. Yes. So exactly that, that definitely stuck with

Dan: me. It's a good friend. And we, uh, when he was developing homomorphic encryption years ago, when I was doing multi-party computation in the same DARPA program, so we've actually, taken the, how a lot of these discussions together.

And I'm very happy for all the success that he's been having. And there's many others with whom we've been doing this because. A lot of people working on secure computing believes that this is actually how information's are supposed to be. And I am one of them. The principle behind there is pretty, pretty simple that, okay, this is a bit of this might get a little bit philosophical, but consider this computers.

We have data protection regulations because we designed computers wrong in the past. Computers were designed not to protect data, but to copy data. And that was logical because why would you want the computer? Because what do you do with it alone? The power of computers really blew up when you had them networked.

When you could copy data from one to. So then they really flourished the whole idea of computers by getting data exchange and ease of copying. But the problem is that computers make copying data so easy, but there is no control. If I send you a file, then you can put it on the internet. You can put it on whatever, dark web BitTorrent, wherever you can print it out in a million copies and throw it around from a helicopter and computer made it possible to make copies.

What if we had built computers so that copying was somehow better controlled, but still possible. This is what I think we're doing good, fully homomorphic secure multi-party computation. We're building the next generation of computers, which gives you better control over how data is. And maybe we won't need privacy regulation in the future.

That's the sort of we'll need it for some other things like ethics. We still need to figure out the ethics so that you wouldn't build totally in ethical things. But a computer built on these technologies might actually solve a lot of data protection. We might be able to throw away three fourths of the GDPR, if that becomes ubiquitous.

So fully agree with Kurt here, we're sharing the same dream. I would like the world to be like that as well. So even once we built these computers of the future, which are programmable with rules of what they can disclose when humans will still be needed to figure out what the disclosure rules are. Because again, the guy in the utilities company will need to know which address they need to send the water.

In my house and so on and the electricity. So there's a, there's multiple layers of security needed. And this is going to be one that I'm hoping. Crack open a lot of the exchanges and join processing and analytics in a way that we don't have to ask so much of the privacy questions anymore. The ethics issues still remain.

And I'm adamant on repeating that. One of the, I mentioned the AI and D governance project. We did a machine learning model on houses and a fire emergency. Can we predict which houses catch fire using a machine learning model. Then we could go and try to do interventions, maybe go in and ensure that these people have their fire alarms, that they have batteries in these fire alarms and so on.

And the question is when can. Go on some people's ground, personal like homes and ask them to, we have a, we are afraid that your house will catch fire soon. This sounds like a racket. It sounds like. Exactly. So, we have reason to believe that your, a building might be in danger of. Which is are you going to light it on fire?

Is this a Fahrenheit 4 51? Or what is it now? So there's a sort of this whole, even if you have great machine learning models, can you really do interventions? You might not be able to do interventions. And then the question is. I do get the economic effect. Obviously a person can go, and I would like to do a self-check for my building.

What's my fire risk score. But how many people do all pertain these days? Not many. You have to do campaigns. You have to get them. I don't know, catch them on Facebook or something. Then you got to get them to consent and then they do all of this. So the question, a big one. And I don't know how to solve this is, and possibly it can't be solved.

Maybe it shouldn't be solved is exactly how you do interventions on this machine learning thing on a grand scale. Because again, I do believe that in certain place, This should be done in order to help people catch diseases that they might be in danger of contracting in a few years or something they have in their genomes will tell that they will catch cancer in five years.

It’s just hasn't come out yet. We'd like to see. Proactively help you with that.

Eric: That's intriguing. So certainly from a health perspective, I think if you come to somebody and you say we have information that suggests that you might have these health issues, if you think about the channel, possibly people might be more inclined to believe you.

If you can support the fact that you have credibility when it comes to things like their housing or their security. I think it gets a little harder.

Dan: Another thing cybernetic recently did with the government is we help design a consent service for secondary use of data. The government has a lot of data on people and it's pretty good quality like on average.

So the question is, could other services use the data that's that the government has on me and maybe the government built consent service, which is I don't know. Open ID connect

kind of a login thing. And that I go and start users start up service and the startup say, okay, before I give you your I don't know report, I would like to access your tax records.

And I will now forward you to the government of consensus where you will give consent. And then I will receive the text records and I will give you your crazy. It looks nice. The question is exactly that does consent solve everything and yes. Consent is nice and it is a very powerful tool under the GDPR.

If you have a consent, you do, you can move data across borders and all of that. Obviously the giver can also revoke consent, but that does happen rather, really I have to say, but the question is exactly that you can't do the big things based on consent. You can't do big research. You can't push a human condition for.

Just based on consent, assuming that you want to do a big health study and you get ask people to opt-in for it, you will get a certain biased sample of people who are like forward thinking, possibly younger, possibly more educated, something like that. But you will need a stratified sample. You need people who, men and women old and young, various social backgrounds.

So you may have to throw away. Data that people who consented. So consent doesn't solve everything. And the question about health face in a way, some people would also say that nobody else knows what's good for my health. Let's just look at the COVID-19 vaccination situation right now. Is that a good example?

There are people coming and telling you, okay, there's still a chance you can catch the disease if you're vaccinated, but it's so much smaller and you will be able to protect your loved ones, but people are still doing their own research. So the point is exactly that. In some of these services will not make people happy, even if they will be arguably good for them.

And then the question is exactly that, should we continue to stick to these uh, consent opt in models? But the question is that research should be able to continue to be. One of the things I didn't mention about this fire emergency model is that it was relatively well explainable. The first interesting factor about houses with catch fire are that people live in them.

So yes, if you are living in a house, then that house is much more liable to catch fire. Which is very explainable. It's really explainable AI. But yes, we knew that before and there are some other things I'll be, obviously there is some non-trivial things like when it's raining, or the temperatures drop, and people start heating their houses and then accidents happen more probably.

And that's already relevant, but you don't really need to go and do targeted interventions based on people living in a house and it's getting cold. I in the fall and people are putting

on their stoves. Is, this is you know, the whole, the target group, you should be targeting. You can just go door by door and say that we are targeting everybody who has a, where you're living in and there's a wooden stuff.

Or we know that your heating system is like that. So we're going to everybody. You're not targeting individually. And that also creates interesting options because if you paid attention to what Google is doing with their advertising targeting recently, then their NFL Lucy system basically does a similar trick.

They say that you can target only a group of people, but you can't target individually anymore. And that is creating the power shift in advertising. Instead of being able to target the individual, you can only target the group. It looks good for privacy actually, but also it has a different side, which is that by building these new advertising systems, all the big companies like apple, Google, others are creating, they're a walled gardens of advertising.

And because advertising used to be a big open market, but now we have safari brave Google, all of these browsers, which have a lot of tracking preventions. It seems to me that the one losing out on most of this is just Facebook, but we'll see how that pans out. But again, it comes to the same idea of interventions we used to be targeted individually, and that was a very targeted intervention that turned out.

Not very legal, according to certain regulations on earth. So now let's see if we can blanket target people and let's see how this slide will change the whole advertising economy. Interesting times ahead, privacy technology is being deployed. Facebook is hiring hundreds of privacy engineers. I'm being told hundreds.

Even like I've lost some of my people to the big companies in the U S so that uh, I know that this is a growing topic and people, if you're listening to this consider researching privacy technologies, because privacy engineers are hopped on the job market.

Eric: Yeah, I wonder on the on the blanket approach, defining the group.

If you have a lot more data points for defining the group, you can effectively define a group with the same individual characteristics as the individual you're targeting. You know, it's less of a blanket. It might be more a swatch.

Dan: Precisely. And that's the thing there has been research on for example, on Facebook, also the filtering goals. If you want to target somebody on Facebook, then the filtering controls as they were available, allows you to go pretty granular, basically identify individual people. If you knew what very unique combination of socio-economical attribute was.

So the same problem exists here. The blanket might be a band-aid.

Eric: Yeah. And, and honestly the blanket might actually make the targeting more intrusive and more effective because from a data perspective if you're able to collect. You're, you are basically you're, you're running a probabilities, right?

So it's actually more efficient to run probabilities with a whole host of characteristics from a group. And particularly if you can identify that the group X tends to act in a certain way that targeting can be much more effective. You know, in many ways, it's almost like it's almost like just doing the analysis at a different

Dan: yes.

And there's a hole currently. Obviously we have there's environmental ism right now. And there's this term called green washing, which is claiming that something is environmentally. Great. If in fact it might not be, for example, an electric car, a Tesla running in Estonia versus running in Norway where most of the electricity is a hydro.

It's a totally different thing when it comes to environmental racism. So there's green washing, washing privacy washing.

Eric: I was going to get to that. That was Dan's. I just taking ideas,

Dan: leading you into this very well. The point is that there's quite a lot of privacy washing around where folks are saying that in order to improve their privacy, we're deploying this new system. But the question is what's the actually. What can you actually derive after you've built the system? What are the capabilities?

What is the new monitoring or surveillance systems that are built as a byproduct? And this is not always immediately understood in the privacy enhancing technologies course that I teach at the university of Texas. The first lecture we always start with, we take a new innovation that they either take some of mine that I found from newspapers, or they find one on their own and they analyze it with regard to the explicit and implicit.

Features. So one thing is that you'll read the press release. This new system will allow us to blah, blah, blah. It's totally anonymous. And then I asked them to analyze them with what are the possible additional use cases or the impacts that the press releases are meeting. So what would you be able to do with you have built this kind of infrastructure and that's what?

The, in the first lecture, I try to get the students to think in these terms. And then they are taking the technologies in the follow-up lectures in a lot of different way. They start thinking about, okay, if I deploy this when I will be able to reduce data leakage in as this kind of a system. So that's a sort of uh, alignment lecture and they take indeed privacy, washing and greenwashing, I think are very similar privacy and environmentalist.

Yeah, you can fight for them using similar arguments, just, use less data, use less electricity.

Eric: Shifting gears a little bit to digital identity, they will just wrap it up on there. You know, what are we evolving to in terms of digital identity and how do you look at the solutions that you're developing or hope to have to

Dan: develop around it?

Very interesting. That's a great question and very relevant right now, given that the apple right now is iOS 15, has the ability to have driver's licenses in certain wallets, maybe also ID cards. The whole wallet based digital identity is a really one potential future solution. In Europe.

There is a new regulation for digital identity coming. The current draft also speaks a lot about wallets. Uh, the situation is. The Estonian model I explained before is about, I prove to the government that I am this person with this personal code. And thus I shall get access to certain services. The wallet approach instead is a sort of a.

Digital version of a paper world. You have a wallet in your pocket. You take out the credential, you take out a risk paper saying that you are receiving certain pensions. For example, that the wallet idea goes and implements this exact pattern in digital world. And now the question is how we will implement it in technical.

We'll be using the secure elements in Androids and iPhones. Will we be using some clever kinds of cryptography, like threshold cryptography, like we're doing with smart ID split key is how cybernetic calls the product will we be using nothing? Will we just have the data on our phones for COVID? And that might actually happen because, uh, COVID-19 immunity certificates, which tell you how many times you've been vaccinated Europe in a flash of collaboration managed actually to standardize those.

And you can basically go with an Estonian certificate to UK or France, and they can validate. And it works, actually, it works wonderful, but it does. But the point is also that the system is such that these can be revoked. And the other interesting aspect is these can be cloned really easily. So obviously if you have a pension certificate credential, then I shouldn't be able to clone it and claim it's mine.

So one of the interesting challenges about the whole. Wallet, a revolution that's coming both in the U S and Europe, and possibly also in other territories is going to be, how are we doing enrollment and onboarding? How trustwoarthy will this be? What will be the level of assurance that the document being shown is actually mine?

I don't know how it is in the U S but in well in Estonia and some places where I go and they actually check my immunity certificate these days, they don't ask me for my document,

meaning that the name I have on that certificate could be. Obviously you can differentiate between like boys names and girl's names, but the point still is how do you check that a credential actually belongs to the person presenting it?

And that's going to be interesting and there with that, again, the Estonian model works, but it has some other, let's say some people would say features. It's quite online. The Estonian ID system means that the query is made to a revocation list. For example, meaning that there is an online component, you can't do all the identity proofs in a digital ID in an offline fashion.

In this wallet system, you might be able to. And now there is a question of whether we will deploy more cool cryptography, like attribute based credentials or zero knowledge to prove that I am older than 18. So

Eric: this has been on the longest podcast taping that we've had, but yet it still feels like it's a, it, the time went by really quickly.

Yeah. It was so great to have you on the podcast. And just to get your perspective on so many.

Dan: Thank you very much, Eric, for having me really I, I love the back and forth. I loved how you came up with privacy washing in last sub second, always good to have these discussions where we understand each other.

Thank you so much.