

Unlocking Data: Advancing Impact – English Transcript

(Note: This transcript was AI-generated using [NoteGPT](#). Transcripts may include inconsistencies or errors and some sentences have been condensed.)

00:16

Hey everyone, good morning, good afternoon, or good evening, depending on where you're tuning in from! I'm Uyi Stewart, the Chief Data and Technology Officer at data.org. I'm excited to welcome you to our webinar about unlocking data to make a positive impact. Privacy-enhancing technologies have been around for a bit, but they've mostly been in the hands of a few regulators and select companies. That's starting to change now!

01:05

In 2023, the G7 data protection and privacy authorities came together to support a plan that focuses on promoting new technologies, including Privacy-Enhancing Technologies (or PETs), that can help build trust in data sharing and protect people's privacy. At data.org, we really believe in the game-changing potential of these PETs. In this webinar, we'll dive into how these technologies can help unlock financial data for public health while keeping privacy intact and making a real impact.

01:50

We're tackling the issues around accessing data while promoting financial inclusion and building fair ecosystems. We're really excited to have three international experts on our panel today. First up is John Derrico, a VP at MasterCard, who handles data sourcing and emerging tech. Next is Salil Vadhan, a computer science and applied math professor and the faculty director of the Open DP project at Harvard University. And last but not least, we have Shubham Kumar, a research scholar.

02:36

"Hey everyone! I'm really pumped to jump into today's chat about how data can be used for social good. John, let's start with you. You're leading the charge at MasterCard when it comes to data sourcing and the tech behind it. Can you share what MasterCard's vision or strategy is regarding this? Also, how do you make sure data privacy is a priority while using financial transaction data to help with public health decisions and more?"

03:25

"Alright, so we've got some new cases, which is great! First off, thanks for having me here today. I'm really excited to share some ideas and get some conversations going. To give you

a bit of background, Mastercard is a global tech company in the payments space, and we offer solutions for people and organizations all over the world. One of the key points you brought up really connects to our mission: we aim to create a powerful and inclusive digital economy that benefits everyone, everywhere. For us, that means making transactions safe, simple, and smart."

03:59

"Now that we have this context, our leadership in privacy technology goes way back. We've been building our data profile for almost 60 years to make commerce run more smoothly. It's always been about safety, security, and trust—those are the core of what we do here. Our PETs initiative is part of our innovation efforts, supporting both privacy and transparency. We see it breaking down into three key areas."

04:27

So, regarding your point about vision, strategy, and availability: as you said, part of the vision is to seamlessly use data for innovation while ensuring privacy compliance, all in an automated and flexible way. Now, about the strategy—yeah, it's a buzzword, but it really matters. How do we make tools available that democratize access to a whole ecosystem of options instead of just relying on a few one-off solutions? And from that, how do we move forward?

05:01

"We make it easy to understand things and manage them, so we can all collaborate in a flexible way. Lastly, we need to make sure we balance privacy with usefulness, providing insights for research and specific use cases. We prioritize data privacy through our governance framework, focusing on privacy from the start and implementing the necessary controls. That's where our vision comes into play."

05:35

"Thanks a ton, John, for clearly explaining Mastercard's vision for using data to drive innovation and make access more inclusive through the PET's ecosystem. Now, Salil, besides being a professor of computer science and applied math, you're also the faculty director of the Open DP open-source project and a well-known expert in privacy-enhancing technologies."

06:15

So, I'm curious about differential privacy, which some people see as the gold standard. Why is that? Also, can you explain how it helps different fields work together, especially in

public health and financial inclusion? It's great to be on this panel with you, Uyi, John, and Shubham. I'm looking forward to discussing these questions and hearing from the audience later, too. So, yeah, differential privacy is all about enhancing privacy.

06:57

I'm really into technology, especially in the field of open differential privacy (DP), which we talked about earlier. We're working on open-source software to make differential privacy more accessible, like John mentioned. It's a key privacy-enhancing tool. Some people call differential privacy the gold standard, and that's fair, but it doesn't mean it's better than other privacy technologies; it's just one of the best options out there.

07:37

The gold standard for what it's meant to do, which I'll explain in a minute, works alongside other privacy technologies. They team up with differential privacy to tackle different issues around privacy and trust. So, differential privacy is all about figuring out how to ensure that the data you share, the stats you put out, and the machine learning models you build only accurately represent the population you're working with.

08:19

They're developed or trained on data that doesn't reflect specific individuals, alright? So, other privacy-enhancing technologies focus on how to create that model without needing different organizations to gather all their data in one spot, for example. They tackle other trust issues and how to work with sensitive data, but differential privacy is all about the final product that you end up publishing.

08:54

How do you make sure that information isn't too revealing? One way is by adding small bits of random noise to the statistical calculations and machine learning models. This helps hide details about specific people while still allowing us to see trends in the overall population. It's basically the only solid method we have to think about these kinds of statistics.

09:33

Releases that make sure all the stuff we publish together can't be used to piece together information about individuals. In the past, people have tried to do this in a more random way, and it's been proven to not work. So, why does differential privacy encourage teamwork across different fields? Well, you get the best results when the approach is

specifically designed for how the data will actually be used, and that can vary a lot depending on the situation.

10:08

It's like, you know, a public health researcher looking at data to figure out how diseases spread, or the impact of lockdowns. Or an economist checking out the links between parents' income and their kids' future earnings to get a grip on economic mobility. So, it's all about deciding where to add some noise, which stats to share, how detailed to get, and which ones to focus on for accuracy, doing all that in the best way possible.

10:51

To really make the best use of data while keeping privacy safe, we need privacy experts and differential privacy pros to team up with the domain experts who will be analyzing that data. Thanks, Salil. So, John, you explained how a private company shares data to boost innovation, right? And Salil is here to discuss how to make sure that only the right variables are released. I want to clarify this a bit.

11:36

“Let’s have a chat now with Shubham, who’s on the call. Shubham, you and your team from Triple IT Camp are one of the four finalists in our Public Health Challenge, which aimed to use transactional open-source data, differential privacy, and machine learning to help with epidemiological models for public health in Colombia, and eventually in India. My question is, can you explain how this data has been used to model pandemic behavior and spot potential hotspots?”

12:25

"It's awesome to be here! So, privacy-enhanced technologies allow us to access tons of transaction data stored by banks and payment systems, all while keeping contributors' privacy intact. This kind of transaction data gives us insights into people's real-time spending habits. We can use this information to improve epidemiological analysis during different phases of a pandemic."

13:08

Transaction behavior has changed a lot. For example, people are spending less on luxury goods and more on essential items. These changes in spending patterns have been used to model behavior during the pandemic. The unique thing about transaction data is that it can help us visualize what's happening at different ZIP code levels. The more transactions happening in an area, the greater the chance of interaction among people there.

13:54

People are more likely to spread diseases in crowded areas, so economic hotspots are closely connected to potential pandemic hotspots. This is how we've tried to model pandemic behavior and hotspots using transaction data. Thanks, Shubham. We've covered the whole process from data collection to processing and finally to the users, which in this case is SHBA. Now, I'd like to take the conversation to the next level.

14:36

Let's dive into some of the challenges and the environment for rolling out this pet technology. John, if you don't mind, I'd like to start with you. For companies that are still in the early stages of their pet journey, how can they find a balance between ethics, risk, value, and growth? Also, could you share how MasterCard has piloted the necessary governance framework and the integrated pet capabilities that have been made available?

15:21

"Data clean room? That's a complicated mix of ethics, risk, value, and growth. So, if you think about it, starting with a privacy-first approach and focusing on what problems we want to solve and why is crucial—rather than jumping straight into the tech and how-to. That's a key point. Then it's about getting the right mix of stakeholders involved based on your data and your use cases. That's the first part of it."

15:57

We kicked things off by setting up a pets governance group early on. This group, made up of people from different areas, helps with evaluating our pilot projects and figuring out how to bring them into production. They also provide the guidance we need for education, both within the company and to the outside world. The second part is that when we talk about what a governing framework can do, it considers all the different stakeholders and has specific mechanisms in place. But there are some challenges to navigate.

16:31

How do you set that up in a flexible, repeatable, data-driven way? The framework is driven by the stakeholders. For instance, we were discussing different types of pet technologies earlier. How do we view this as a pet ecosystem? This ecosystem has various capabilities that can be used for different purposes. What's really interesting is that we can see how multiple pets can come into play. For example, a data clean room might be one way to make it happen, but it might also...

17:04

There are other ways to tackle this, and we might need different technologies to support a privacy-first approach, along with the right people involved as we go through evaluations and processes. I really like the idea of a pets ecosystem! So, building on what John said, can you explain how we balance the historical trade-offs between privacy and usefulness using differential methods?

17:43

"Privacy is a top-notch technique, and I'm really glad to talk about it. It ties into what I said earlier in response to your first question. Also, just to build on what John mentioned about the clean room—that's a great example. It's basically a controlled space where we can make sure that data flows in a safe and regulated manner, using something like differential privacy."

18:25

"It's one of those technologies that provides tools for figuring out what can actually be processed in a clean room and what can be allowed to move from the clean room to the outside. A great example of how different technologies in that ecosystem work together. So, privacy versus utility, as I mentioned earlier. Differential privacy is based on one of the key truths about data privacy: you can't just exploit data."

19:00

"Basically, you can whip up any number of analyses and publish as many stats as you want, and you can train a whole bunch of machine learning models on the same data without really putting yourself at serious risk. This has been proven time and again. The US Census Bureau figured out that the way they were putting out tables in the 2010 census and before was a problem because they were publishing tons of stats all based on the same data collected from the population."

19:35

This happens in the tech industry too. Differential privacy gives us a way to think about how much privacy we lose with each new publication. It helps us balance how accurately we want to share the stats from our machine learning model against the potential privacy risks. It lets us set limits and say, "This is how much risk I'm willing to take; I don't want to expose the people in the data."

20:12

We're looking at some pretty serious privacy risks here, and that's where differential privacy comes in. It helps us tackle the tricky balance between privacy and utility in a more

organized way. Thanks, Salil, this is coming along great! So, John, you introduced us to the idea of the PET ecosystem and the whole clean room concept. And Cil, you've pointed out that from a differential privacy standpoint, you can't just use data however you want without facing some risks. I really like how you put that!

20:48

There are some real vulnerabilities here, so let me break it down again like I did before. Shuba, I want to ask you something based on your involvement in our challenge. Can you give us some concrete examples or use cases showing how financial transaction data can really help public health? I'm talking about how cross-domain decision-making can work, like understanding relationships between random variables or making predictions about the future. So, what can you tell us about transaction data?

21:33

Specifically, when we look at spending on transportation, it gives us a good idea of how people are moving around. This info is super helpful for predicting new cases. By adding some extra factors to this financial data, we can forecast future cases. Now, let's shift gears and talk about another point: financial data can show us how policies put in place during the pandemic have impacted everyday people's behavior. For example, during COVID, the shutdown of big places...

22:19

"Establishment has a bigger impact on how people are interacting compared to social distancing norms. Now, moving on to another point, social groups can be linked to spending habits, and we're still figuring out how to estimate contact patterns from that. This information is really valuable in epidemiology. Thanks, that's great! Thank you, Shuba. I'd like to welcome everyone who just joined us to start thinking about any questions you might have."

23:05

Alright, we're heading into the last round of questions for the panelists now. So go ahead and drop any questions you have in the chat. John, here's my third and final question for you: can you give us some examples of how insights from MasterCard data could promote equitable financial solutions? As I mentioned earlier, our mission is to connect and empower an inclusive digital economy that benefits everyone, everywhere. I've touched on that before, but let's dive into how we and other companies can leverage that.

23:49

We're working on tools that not only spot and measure biases in data but also help reduce them. That's what SE and others will use these data sets for—not just to make sure our models aren't biased, but also to figure out if the data itself is biased. By starting from this foundation, we can think about how our mission can help create fair financial systems by focusing on the core data.

24:25

"We all need to recognize and reduce bias, which is awesome. So, Salil, I'm turning to you. To tackle that bias, we've got to do some serious filtering, right? What challenges have you faced when applying data protection in different sectors, and how can we get past those hurdles to really make the most of privacy-enhancing technologies?"

25:13

Some of these issues are similar to the bias problems John mentioned. Others focus on how people in a specific field need to analyze the data. What's the most important signal to keep, and what's less important? Making those decisions, as we said, is key to getting the most value from privacy-enhancing technologies, especially differential privacy. We can't just make the data available for everything without risking exposure.

25:47

We're exposing ourselves to serious vulnerabilities, so the best way to tackle this challenge is by having groups like Shum's team and the other participants in the PETS challenge get creative and innovative. We need to figure out the best ways to use this data for the greater good, especially in public health. Let's understand what signals we can take advantage of and think of creative ways to gain insights from it. That's basically what I wanted to wrap up with.

26:20

"From our experience with deploying differential privacy and PETs, every time you introduce this technology to a new sector, there's a lot of unfamiliarity. The laws, regulations, and organizational policies aren't really ready to handle it, which can create some challenges. We've definitely seen this kind of friction during the initial rollouts in new areas."

26:58

"All of our organizations are trying to collaborate on the pets challenge at Harvard. We have to be really careful when handling sensitive data, and that's completely understandable. But I believe that if we keep doing this process over and over, we'll gain valuable experience. Others in the same field can look to these examples and develop best

practices, which will make things smoother in the future. Thanks, Salil. And for everyone in the audience, feel free to send in any questions you might have!"

27:29

"Hey John, I mentioned earlier that I had my last question for you, but I kind of lied, so just hang in there with me for one more. Building on what S just said, MasterCard has been a fantastic partner for the PETS work we're doing with Harvard and others. So, what's next for MasterCard when it comes to collaborating with partners to leverage MasterCard's data through PETS? Alright, here's the deal: there are a few things we need to focus on. First off, we've got to keep the innovation going."

28:10

We're taking the lead here while making sure these solutions are accessible to businesses, nonprofits, and NGOs. Next, we need to stay connected with our partners in Privacy Tech to make the most of their services. We have a lot of ways to share our insights with customers and partners, but it's crucial to build trust both ways, as that can be tricky. So, how can we keep pushing for better controls?

28:48

"That used to be, you know, based on contracts, and now it's more organized around specific pet deliverables. Like I mentioned, things in clean rooms really help make more data available. And lastly, SE brought this up before—transparency in the Privacy Tech space is super important. So, how do we keep that thought leadership going to handle privacy concerns, confidentiality, and still be transparent about the outcomes? It all comes down to that transparency."

29:23

We want to make sure people understand pets on a deeper level, not just on the surface. Sometimes it feels like there's this plausible deniability or what I call "privacy theater" instead of focusing on the real benefits. That's why I think being transparent is really helpful, along with making the solutions easy to understand. I'm actually glad I asked that because now I can circle back to Shuba, right? We can see how important transparency is in this situation.

30:03

"Implementing pets on the ground level across organizations around the world, especially in the Global South. So, my question to you, Shba, is based on the idea of localism: how will your tool provide real-time insights for local health officials, like those in Bogotá,

Colombia, or New Delhi, India, while keeping data private? And what key lessons from your project can be applied globally? Our tool uses differential privacy to protect that data."

30:54

"Contributors, about privacy, as Salil mentioned, let's define differential privacy. Basically, we're adding some calculated noise to protect privacy when we release statistics. In our case, we're focusing on merchants as our privacy unit, so we're securing their information at the merchant level. Now local health officials and the government can use their own transaction data to get real-time insights into potential hotspots."

31:41

Here's a look at how the pandemic is moving along and what stage we're in. Basically, there are three stages: the onset, the peak, and the recovery. With these insights, people can make better decisions about the policies they need to put in place in their local areas. On a global level, we've realized that this is a really powerful tool for accessing large amounts of data.

32:16

"Can't access it because of privacy issues, but this info could help us create tools, methods, and functions, including what we're working on here to assist policymakers. Thanks, Shuba. I want to keep it grounded, focusing on localism, especially regarding pets and how it applies to folks in the Global South or low- to middle-income countries. So, Salil, my question is for you: what innovations do you think will be key?"

33:00

The future of pets, especially in developing countries, is looking bright. This was a big reason we kicked off the Open DP project. It ties into what John mentioned earlier about making pet data accessible to everyone. When we started the project five years ago, the only groups using differential privacy were big, well-funded organizations with loads of expertise, mostly in First World countries like the US. For example, the US Census Bureau has a ton of statistical know-how.

33:49

"With over a century of experience in disclosure avoidance, big tech companies like Google, Apple, and Microsoft have a ton of technical know-how. Other organizations, even those in tech, trying to adopt these technologies are starting from scratch, having to rebuild everything from the ground up. It's tough to get the basics right, let alone grasp the whole education aspect of what it all means."

34:27

The POS value you can get from this tech—what does it protect and how do you use it safely? Sorry, I don't mean to interrupt, but John has an emergency and needs to leave. This is the world we're living in right now. I apologize, team, and thank you so much. I'm looking forward to the rest of this event and to making more data and capabilities available as we wrap up the rest of this challenge in the future. Thanks again, everyone—I really appreciate it!

34:57

"Hey everyone, thanks for joining. Sorry to jump in, but I just noticed John had to leave. No worries, I'll keep this brief. I haven't seen any questions from the audience yet, but I want to highlight something important. I checked the list of participants in this webinar, and there are quite a few people from the Global South. I'm sure they're wondering if this is just another tech thing from the Global North and what the real value and relevance of pets is."

35:31

In my situation, that's why I'm asking you this. So, what I'm getting at is that one part of the solution is to have reliable tools that are open source and easy for organizations to pick up, especially those that might not have experience with these technologies or the budget to hire a team to create their own custom differential privacy solution. That's the gist of it, along with a wide range of resources.

36:16

We're creating educational materials to help decision-makers see how valuable these are for guiding organizational policies and best practices for using this technology. That's part of what we're working on with the Open Differential Privacy Project. But just putting out materials isn't enough; we also need to actively engage people. There's a group that's working to promote the use of this technology.

36:52

The United Nations has a lab focused on privacy-enhancing technologies, called the UN PET Lab. We're working with people from national statistical organizations worldwide, and there are various experiments and prototypes being developed by countries in both the Global North and Global South. Hopefully, this shows that similar progress can be made in other sectors, encouraging global adoption and involvement from everyone.

37:29

"Thanks for that, Salil. There's actually a question from the audience that ties into what you just mentioned. Let me read it out and start with you, Salil, and then we'll see if Shubham has anything to add. So, this is from Colin Rice, and he's asking about how using pets, especially with cross-sector data, is affected by the new national cross-sector data privacy rules compared to just sector-specific ones."

38:09

"Specific privacy rules are really important. One of my hopes is that privacy-enhancing technologies can help us get past the current limits on data flow. Right now, these limits stop us from taking advantage of the kinds of analyses we could do by combining data from different countries around the world. One of the test prototypes from the UN pet lab I mentioned earlier was a case study on international trade data."

39:01

We ran a joint test using data from various countries. It was just a prototype, not on actual sensitive data, but more like a proof of concept to show that it can be done. So, developing the technology is just the first step. Next, we need to get regulators and decision-makers on board to set up best practices. Ideally, we'd also like to see laws and regulations updated to support the proper use of privacy-enhancing technologies.

39:31

They're supposed to offer the protections we need so that data can flow more easily for these cross-border uses. That makes a lot of sense. Thanks, Salil. Here's a question for you, Shubham, and I appreciate your patience. It says, "I'd love to hear what other ways we could use financial data to drive public health insights or outcomes, especially in smaller scale situations rather than just during a pandemic." So, there are definitely many ways to apply financial data, but we need two basic things.

40:24

First off, there needs to be a clear trend in the financial data or spending categories related to that event. If the spending category shows a pattern for that event, then you can definitely use it. The second thing is that there has to be enough data. Since we're adding noise to make it differentially private, the data set should be large enough so that the noise doesn't hide the overall trend in the population.

40:55

Sure, there are definitely a lot of use cases, but it really needs to be backed by a big population. So, do you have any thoughts on that? I know it was aimed at Shubham, but

based on your experience, especially from the DP side, what other applications have you come across? Particularly in the work we're doing together, how can financial data help us gain better insights, not just in pandemic situations but also in smaller contexts? That's a great point.

41:45

That's a great question! Let me think about it for a second. While I'm doing that, let me ask Shubham something along the same lines. How can local governments or health officials in India take advantage of the tool you've developed? How can it be applied by health ministries there? Well, local governments can use our tool to get real-time data.

42:31

"Getting a handle on hotspot detection, right? So, by using this, they can tweak their policies on lockdowns—deciding if they need to impose a lockdown in certain areas or ease up on them. Another thing is tracking how people's behavior is shifting during the pandemic and how it's changing over time. They can visualize the pandemic's progress in different areas and see where people are recovering more slowly."

43:13

"Some areas are seeing a more expensive recovery, so they can tweak their policies based on that. Thanks! We only have about a minute left, so Salil, I'll hand it back to you. But I want to throw in one more question: what's your call to action? Okay, to answer the earlier question about other uses of financial transaction data, a potential example is similar to a recent application of OPDP software."

43:54

Microsoft is trying to understand the economic gap between different regions. They looked into digital literacy and how the types of apps people use relate to the economic conditions of their area. Are folks in wealthier areas using different apps on their computers and phones compared to those in poorer regions? You can imagine exploring this further.

44:41

We're talking about digital apps and how they affect different types of buying and economic activities. It's important to see how this changes in various places, among different age groups, and across different income levels. The message here is for everyone to get involved. One example is the OP DP community—you can join our Slack or sign up for our mailing list. There are also other groups out there working to push for progress and find new ways to apply this in different areas. Let's team up to bring these ideas to new sectors!

45:19

In situations where society benefits, these technologies will progress, and we'll learn to use them better for the greater good. Thanks a lot, Salil. We're out of time, but another question just came in. I really appreciate you, Salil, and also Shubham and John, who joined us earlier, for sharing such great insights. Personally, I see now more than ever that privacy-enhancing technologies have huge potential to support financial inclusion and...

46:00

"Hey everyone, just like Salil mentioned earlier, I really encourage you all to keep exploring communities that promote collaboration across different fields. This way, we can tap into data to make a real impact on important issues. That's what we're all about at data.org. I'd love for you to sign up for our newsletter and connect with us online. Thanks so much for being here today, and a big shoutout to our panelists. Have a great afternoon!"

46:36

"Good morning and have a great day, everyone! Bye!"