Terry Sachlos: Welcome back everyone. Do we have all the panelists? Waiting for one. Everyone has their ... everyone's keeping hydrated...

Speaker 2: Caffeinated is more.

Terry Sachlos: and caffeinated. Okay we're going to kick start the next panel. We're going to try to get back on track on time so we can get everyone out of here in an orderly fashion. So maybe we'll probably have a limited amount of time for questions at the end. My name is Terry Sachlos. I am an assistant professor at the Lassonde School of Engineering. I first of all want to wish you a good afternoon. Thank you for attending, it was an amazing a plenary, I have to say. I was amazed and slightly scared with the potential of the technology but some of the things that we saw were actually on the health applications in which that that's actually the subject of this panel.

I'm actually very excited to open up this panel which is titled Big Data, Health and Science. Which focuses on the applications of AI. And we certainly have a stellar cast here to tackle this very topic. Health is something that I'm actually very passionate about. My research interests are within health research and I'm very excited to see how AI is going to be having an impact not only in advancing health science and research, but also in patient lives. And I think that it's always important to remember that there is a massive potential here that this technology will impact all of us, at the individual level. Whether it is directly or through a family member, there is a chance that we may actually be personally touched by this type of work. So at the heart of all of this is actually big data. And big data is not something new for health research.

We only need to look at the link between smoking and cancer. As a reminder it was not the biochemists or the molecular biologists who we're looking at tumor samples that discovered that connection. But it was actually the epidemiologists who do a population based analysis based on surveys in finding a relationship between smokers and non-smokers. This was back in 1950. Sir Richard Doll, identified this link and this was at a time when we had manual surveys. This was still paper. So imagine the potential now of the tremendous amount of computing power that we actually take for granted and the advanced algorithms that are being designed and we saw some of those. Imagine the potential of how that can actually touch us as individuals.

And so throughout the day we've also been hearing about the potential barriers and challenges and this also stands true for the health space. So today's panel will try to dissect and understand some of these challenges within health applications and hopefully provide some insights and potential solutions. Similar to previous panels, each speaker has about 15 minutes, at which point we'll have questions at the end. That being said, I want to introduce very distinguished speakers. I want to start by introducing the trailblazing and very talented professor, Pina D'Agostino who holds many titles including a professor at Osgoode Hall Law School. The founder and director of IP Osgoode. And also the conference organizer. Thank you for organizing this fantastic conference. Thank you. Can we give her a round of applause?

We will then move on to the equally distinguished professor James Elder, who is a professor at the Lassonde School of Engineering. And is also across the points to the

Faculty of Health, at York University. James actually develops these mystical AI algorithms. I'm looking forward to that. We'll then hear from Victor Garcia who has a very long and illustrious career in technology. He's currently a managing director of ABC Live Corporation and faculty at Schulich Business School. And he's also spending a all a time as the chief technology officer at Hewlett Packard Canada. So he's seen lots of things over his career. And then finally, we will hear from a rising star of Ian Stedman who is a fellow at SickKids. And Ian Brings a very unique perspective on combining law, ethics and philosophy. So without further delay, I want to open up to Professor D'Agostino.

Pina D'Agostino: Alright Thank you, Professor Sachlos let me just find my slides. All right. So hello everyone. So this is a deep dive into the past for me. It's I guess going back to about 2004 when I was finishing up my doctorate and Professor David Vaver to start here with us today was my supervisor and he was actually the PI on this grant that we got to study this eHealth network. So we kind of took it on tour over 10 years ago. And I have to say that I got really frustrated and I know with many of my co researchers on the team did as well because we were sort of getting a lot of role dices and yawns in the audience and it wasn't fun. So I said, you know what, the heck with this need to research something else. So that was it. Now, and that was looking at data before it was called big data, but it was very much big data.

So now we call it big data. It's sexy. We're talking about AI. Everybody's on the bandwagon. So I saw, you know what, let me jump back on, given that we did a lot of this work over 10 years ago, and I think it's more relevant now that people seem to be awake and aware of the issues and ready to receive the information. And a lot of it has already been in some ways discussed today, some of the key notes pointed out. I'll give you first some brief context then look at the eHealth network that we studied and then towards solutions. And 'towards' being the big word here because we're not there yet, we'll get there. Data has come up again and again, of course it's the fuel for AI. And our morning panel did a fabulous job in unpacking that. And even though not apparent, and even though we actually don't really like it and we think it doesn't make sense, IP is often implicated in these scenarios.

And so among the IP copyright and it seems to be the most applicable, and we've talked a little bit about how that is. And in the realm of copyright we like to talk about categories of works. So in order for copyright to have a hook on the data, it has to be a protectable work. So how does it find a home in the copyright act? Well you could look at different categories of works and their literary works, compilations, artistic works, photographs came up quite a bit. So those are all types of potential copyright. So really the standards are not very high for copyright protection. The bar is quite low, anything really that is fixed anything more than a line drawn with the aid of a ruler was what one case had said. So, and I like to use my Italian heritage. You don't have to be Leonardo DaVinci to get copyright protection.

So there are many complexities. And one of these really what this goes to is the core. When we talk about governance, and this is maybe some of my misgivings early on was, we always talked about privacy, cybersecurity. Really the security, both the physical and intellectual of the assets, but never really about the ownership rights. So that's what we looked at. And just to put it out there, when you look at these highly intelligible grid computing environments you have that are a jurisdictional, you have many challenges to the framework animated very much by the law is. So is the law even adequate to even think about these things? And there are many complexities that arise. In terms of the ownership as to who is the actual owner and authorship has come up in terms of who, if it's a human or not joint authors.

And then also the contractual agreements in place that also needs to be looked at. And then term of protection. There's lots and lots of questions. We noted that privacy, there's so many other rights, labor laws that can apply, the ethics of it all. So there's all kinds of overlapping and conflicting rights. What we cared about in our project was really to look at the elephant in the room and which is not so much an elephant maybe right now but we are seeing it and you have many of these grid networks, highly computerized intelligible environments that are springing up. And we need to be mindful of the framework in place. So who owns what? And just as some examples when we were presenting this, I looked back to the beginnings of eHealth Ontario, Canada Health Infoway, and there were so many reports and studies on these and we were pouring millions of dollars into this and there was no real clear idea in terms of the ownership rights.

So what we did, we started back in 2004 was to study one of these health initiatives back in the UK. And this was called e diamond. Have you heard of e diamond? Okay, well I'm not surprised and you will, Roger has of course. You'll see that in the end you'll know why we haven't heard from e diamond because e diamond is no more really. And e diamond was funded to the tune of 4.1 million pounds and it was really meant to help save lives. So e diamond stands for, it's not about actual diamonds, but it's diagnostic mammogram, national database project funded by the UK government to really help diagnose, treat and eradicate breast cancer. So it was really the goal was to amass large amounts of data from across the UK in order to get this done and to do this housed at the University of Oxford.

So Oxford, the diamond was there and that was the key player in making this all happen. So received a lot of attention. Then the prime minister, Tony Blair even said in 2002 that an individual hospital will not have super computing facilities, but through the grid it could buy the time it needs. So the surgeon in the operating room will be able to pull up a high res mammogram to identify exactly where the tumor can be found. This is in line very much with professor Lee, right? Like a lot of that real time data. And this is what they wanted to see happen, at least in the breast cancer and treatment space. So we called ourselves something. So our group was called IMAGE intellectual property rights and medical databases for a grid environment. And there's an ESRC number. So that's the equivalent for shirt in Europe.

So it's the economic social sciences and Research Council. And we were funded and it was a multidisciplinary research team, which I hope that something that you could come away with after this conference is that we can't crack this on our own, in our own little silos. It's not just the law that's going to do this, understand the solutions, not just the computer scientists or the health fund data scientists. We have to come together. So we did that. We were four lawyers, two computer scientists, and a social scientist and it was really an ethnographic study looking back. This is now 2005 when we started this, looking back to the two year pilot project, which was e diamond and tried to figure out how did it work and how could this be scaled eventually to make it really scale across the world and the potential could be endless.

This is just a diagram of the data flow. On the ends you see in blue, those are the hospitals that would collect the data. In dark green are the universities, which would then actually put it in an intelligible form, which would then be uploaded onto e diamond. And light green is of course come up IBM, which was the main tech company involved. And Miranda was a startup from the University of Oxford, also involved in the software space. So that's how you see the data flow all going to e diamond. Now, some challenges and this is just kind of a slice of patient female walks into the hospital and has her mammogram done and a query artistic work depending on, there are many factors, but potentially there's copyright first layer of copyright protection. Then there are many different other players that can be involved. The radiologists actually reading the scan, putting in notes, potentially another work.

Is there a joint authorship there between the tech and the radiologist? Then it gets encrypted on a CD gets uploaded and it goes into the federated database. And along the way, there were other people involved in that process. Now at the grid stage, when it goes into that database, imagine with AI, the goal was to have many other e diamonds out there and populating this grid structure. So just something that hasn't really been mentioned yet. And this is sort of one of the key doctrines, cornerstones of copyright is the distinction between an independent contractor and employee. And I know Doctor Malik, you had my book up there and was actually a good pass. I liked that one, but that's where I really uncovered the distinction between the two. Because if you're an independent contractor, then for the purposes of copyright law, then you're meant to retain the rights to your work.

Unless of course there's an agreement stating otherwise and that's the big power of contract. And then on the other hand, if you're an employee then that those rights are the during course of employment, doctrine applies and your employer actually owns the copyright. When you think about everything that we're doing and Maya spoke about collaboration, you have to think about each of our roles. Are we employees, are we independent contractors? What is our relationship when we think about the big realm of copyright and of course the contractual relationships because contract in this case, Trump's. You have with those arrows, all different contractual relationships. Now how many of you think that they actually had contracts in place? Let me see a show of hands. All right, that's pretty good. Okay. Well one and the answer is partly right because we see that between Ardmillan, the hospital and the University of Edinburgh, there was actually some nice contracts.

And to the tune, the fact that when we look back to the study, we looked at all the emails, the agreements, they took about four or five months to negotiate this. And there were IP clauses and you know, for an IP lawyer you'd start licking your lips when you see this. The first, and this was the first iteration of the IP clause was that all IP arising from performance of the work by Ardmillan. So that's the hospital for the purposes of the e diamond project shall belong to Oxford University. How do you think that went? Well, it didn't go well because that was Oxford mainly negotiating that and the hospital doesn't like that. And so they came up with this one.

Where all IP now arising from the performance of the work by the hospital for the purposes of the e diamond Project shall remain vested in Ardmillan in the hospital and Edinburgh shall procure that Ardmillan shall grant a license to Edinburgh. So to the university and the university shall grant a paid up license to Oxford so that each of the universities shall have, the right to use the mammograms, including any adaptation thereof and other data for the purpose of the e diamond project, which for the avoidance of doubt shall include a right to make the same available to any other e diamond project participant for such purposes only, that's the clause that that stood. Intuitively this makes sense to a lot of the parties involved because it was really use rights between the universities and ultimately the e diamond itself had use rights of the data, but ultimately the ownership rights would rest with the hospital. So that's sort of, well I circled that and you see that relationship was somewhat clear, although going back, we did have a question mark.

There was an independent contractor involved. So not everything was perfectly ironed out because of all the parties involved. And this is what you will see happen in these relationships. 'Cause there are many different players involved. Now I'll just quickly go through these, but you see in all the other cases, no written agreements. So no contemplation, no thinking, no discussion of any intellectual property rights any ownership rights, nothing of the sort. And so that's the case for the next quadrant. So the other relationships are pretty unclear. We have question marks there because of course then if there's no contract stating otherwise and there's an implied license. And this is the same with the guys and Casey L and you see question marks there because of the other parties involved, employees, nurses.

I remember one case it said, it actually had said that an agreement would follow and it never followed. And this is what happens, you get excited, you start doing the work and you know, the law is seen there as a nuisance. You have to deal with it, you don't deal with it. And this comes up again and again in the many events that we tend to organize that IPO has good actually the last one and a women in IP in entrepreneurship where we looked at startups and how startups really are really gang ho to get going. They don't think about the legal aspect. Until hopefully not, it's too late. So in the case of the software companies involved, those are clear arrows because there was a collaboration agreement which delineated all the use clauses there. So in that Oxford had a perpetual nonexclusive license to use the software for non commercial purposes for e diamond.

Really when we think about in some who owns what, what can we say? Well, likely the IP resides within the hospital still. And it's clear in the case of Ardmillan. In the other cases we have to say that it's implied. And there it's still uncertain because anything that is implied is not as good as a written contract. In the case where there's other parties involved in, I can guarantee on the most part and all these collaborations there are going to be independent contractors involved, there is ambiguity. So that needs to be thought about beforehand because the results could be perverse. In one of the cases that we looked at, one of the nurses she was hired, she was a retired nurse and she had done most of the work on the grid and it could be, we thought that she had some stake in this, but there was no agreement saying anything.

But she was the independent contractor. For the purposes of the actual e diamond project. So this project which was meant to revolutionize breast cancer treatment and diagnosis of it and really help save lives. In the end the data is stultified because the project and if you look at the clauses is it only gave itself rights for the e diamond project only and there were no reuse rights because of then the other dimensions of issues that you need to think about, which was the ethics, clearance issues that could be problematic because the patients walking in to that first scan, that MRI only agreed for e diamond and not for all other uses. So when you think also we actually also shopped this around to some patients groups and there they also queried whether they had a stake at all in any of the data or the works coming from the grid.

Something that, you know, we thought about. When we look towards solutions, and this has come up a little bit, we have to think of good governance and you need a fundamental vision, policy. When we think about legislation re generous came up, maybe amending the Copyright Act, something that actually, the workshops in the UK really enjoyed was to have a carve out exception to have something that was also more encompassing, not just for non commercial but also commercial. Because the reality is, and I think it was Dave who talked about it, that it's an illusion to think that research is just purely non-commercial, especially in these spaces. Also you have to look at the contractual relationships in place guidelines and practices. If institutions have those many universities have IP policies in place, these owner centric and vendor centric.

And then the actual technology, whether that could be seen as an enabler and of course education. Back then there was very little awareness of the ownership and IP implications but I think we're actually there now. We're at least I don't think we're rolling our eyes anymore when we hear IP and data and all these important issues. Some things that we had thought about was looking at different models that you could use when you're looking at health data and you're putting it on a grid structure and when it's being accessed in real time where there should be a hospital centered approach, patient centered, an independent organization may be coming in or a hybrid solution or maybe others. We thought about all these things. This is back in 2004. And then of course, and this is something that comes up a lot in copyright, it's the notion of balance.

You want to incent innovation and you want to be able to reward that creativity, not have a chilling effect, but at the same time you have to be mindful of also the patient's interest with the bigger goals are and all the differing goals. There are so many different stakeholders involved. Some other governance issues that we had looked at that come up is really the tension between what is public, what is private, the ethical dimensions, the bias in the data and that's come up safety and security issues and also interoperability in terms of the actual technology being used. Whether there's just one provider model or you can have multiple providers involved to democratize the process and what problems can come with that. Just some final thoughts data governance is absolutely critical. And if we don't think about this beforehand, then you end up fighting.

And that's not what we want because the risks are way too high. And the possibilities way too amazing to overlook. We want to make sure that we get this right and have as many different experts and also the public at the table expressing themselves and really thinking about what it needs in terms of the next stuff frontier now that we have AI here at our doorstep. So I'll leave with that and I look forward to my other co panelists hearing from them. So thank you.

Terry Sachlos: Joining me thank you Professor D'Agostino. Very good. Thank you, James?

James Elder: Great. I'm James Elder From York University and I'm joint appointed to the Faculty of Health and the Lassonde School of Engineering and a member of the Center for visual research at York University. And I want to talk about what I'll call visual AI, which is AI that involves visual observation of human activity and the balance between the right to privacy and the public good.

My lab is called the Human and Computer Vision Lab at York. And our primary research mandate is to build more sophisticated machine vision systems through a better understanding of human visual processing. So that is why the joint appointment between the faculties of health and Lassonde School of Engineering is so critical to the research effort. And we're very interested in developing AI systems that work well, that play well with humans. And so that requires also an understanding of human perception and cognition. And I think this will come up. In terms of the application domain, what I want to focus on today is what I would call the goals of wellbeing and safety and security for humans. So that's in a sense health application, but it's a specific form of health application.

Okay. The big question I want to post today is good or evil. We have billions, literally billions of these kinds of cameras now floating around the planet. And they are rapidly being connected to machines like the one shown on the right here. There's a lot of opportunity but there's some risk associated with that. And I think to answer that question, we really need, there's no single answer. So we need to really think about the application domain. So let me walk through a couple of examples, a few examples of application domains. One of the things we work on in my Lab at York is search and rescue everyone's search and rescue. So this is a collaboration with the National Research Council and the Department of Defense in Canada. If an aircraft goes down on Canadian territory then typically aircraft are dispatched to search for that.

And believe it or not, still the state of the art is, is searched by human eye. Okay. So there's actually a human operator looking at his bubble window from this aircraft for this downed aircraft. Now D and D has an acquisition program for new search and rescue, uh, aircraft that have more advanced imaging systems. So there's an opportunity to have machine intelligence, which will provide assistance, target detection for search and rescue. Good or evil. So most of the video is of unoccupied territory. The risk to privacy is quite low. The upside is very high. And in particular to find survivors, the likelihood of survival decreases very quickly, as time elapses. Improving the efficiency of search and rescue is really important.

So I'd argue that for this use case, there is a balance pretty much weighted toward the benefit as opposed to risk. But if we start moving into populated areas, urban areas for example things get a little bit trickier. So another example project is on road weather analytics. So here we're talking about cameras that aren't on an aircraft but they're on vehicles and what we want to do is analyze the road surface and be able to in real time, automatically update the vehicle on the road condition is it snowy, rainy, dry and so forth. And that's actually important for road clearing activities to dispatch service vehicles efficiently. But it's also obviously applicable to your own vehicles computer and updating it's techniques for managing road conditions. Here it's kind of intermediate ground, we're interested in the road surface that's kind of uncontroversial and the benefit is high, but of course there can be pedestrians and license plates in the field of view. So it gets a little trickier.

What we are primarily focused on these days is more generally what I'll call urban mobility. So that is trying to understand how people are getting around in a complex urban environment. And there's a lot of reasons to be interested in this problem. So we all, I guess have to commute and the GTA. So, we're familiar with the challenges. So the cost to the economy of the congestion and the GTA is very high. There's a big factor in terms of contributing to carbon footprint. And a third issue that's becoming increasingly urgent is issues of accessibility. In Ontario we have a very good accessibility plan, but we're nowhere near on track to meeting those goals. And that is going to become more urgent as we all get older.

And the demographics shift. The more and more of us have some kind of mobility challenge. Okay. So lots of reasons, lots of potential, reasons to study these problems. And we think there's a lot of opportunity for what I'm calling visual AI to contribute here in terms of operations. So understanding what's going on in real time, but also in terms of planning, statistical analysis. An example of the former is actually an example of both would be Toronto's vision zero goal, which is to try to eliminate pedestrian fatalities and to understand the dynamics of that, you really have to understand how different kinds of traffic are interacting, for example at an intersection.

That's the motivation, but there are lots of challenges. And till the straight that I'm showing some examples of what I call raw urban videos. So we have some highway traffic on the left. We have some pedestrian traffic in the middle and we have a building entryway on the right. The idea is can we use this kind of video data to contribute toward solving some of these challenges? So one problem is that the raw data itself is not that useful. It's huge in volume in terms of pits. And so typically, even though we have these billions of cameras all over the place, the data is discarded usually pretty rapidly after being collected with little analysis. So that has to be managed. The second challenge is that the way that we deliver the data to end users is typically very primitive.

A good example is to walk into any security office in a major institution. So probably there's something like this at the library here or at the university campus. So typically you'll see an operator like this sitting in a chair and faced with dozens and dozens of video monitors, each of which has multiplexed into different views showing different views of the environment. And then typically these separate views are time multiplex, so they'll switch between cameras over time. So this is a completely cognitively impenetrable user interface. If something challenging or complex happens, it's virtually impossible for the operator to figure out exactly what's going on. So it's very important to get beyond this very primitive representation. And thirdly, and maybe most importantly for today's conference is that the raw video is potentially quite invasive in terms of privacy.

People maybe in the field of view here and may not want to be recorded. So just to summarize, here's a table showing various application domains. I haven't mentioned all of these. These are some of the things we're working on and are interested in. And some of these I think connect well with. Professor Lee's fantastic keynote talk, for example, longterm care facilities. But then we have the potential risk to privacy on the right. So how can we address these kind of opposing factors?

We have this project called ISSUM, which stands for intelligence systems for sustainable urban mobility. Which is we're leading at York University. And the goal is to build what we call integrated intelligence systems. That will go from sensing through modeling and visualization simulation and then application to various problem domains. And hopefully we'll address some of the things I mentioned earlier in terms of congestion, emissions, accessibility both in real time operations and in terms of planning. And this is a pretty major five year project with substantial funding from the province and from our partners and the institutions involved. It's five faculty at York University and University of Waterloo. And Really a nice chain of expertise from way from sensing to application domain.

And many interesting Ontario companies, innovative companies and public institutions. Okay. So what's the basic idea? So I would say that the heart of kernel idea for ISSUM is shown on the slide. So we have this massive investment in video infrastructure and in urban environments. And that's great because the data are dynamic. They reveal the dynamics of the city, which is what we want to know. But they're problematic because they're two dimensional and the city environment is three dimensional. And that's where we run into this problem. Of the security office where we have all these disparate two dimensional views trying to represent something complex in 3D. But fortunately on the other hand, we have in 2019 very good access to and the ability to create accurate three dimensional geo databases. Now these databases unfortunately are static. You can sort of see where I'm going here is to try to combine these two into one thing, which let's say we'll call a realtime system for urban awareness, which we'll be both three dimensional and dynamic.

So this will be a much more powerful a framework in which to represent this information both to a human operator who must understand what's going on, but also to analytics machine vision and machine learning analytics. That's kind of the core idea of ISSUM. And here's a sort of primitive form we can say early ISSUM just to give you an example. These are highway traffic data which you can see are being coerced into a three dimensional form. Okay. So we are able to understand the pose of the camera relative to the roadway. And were able to, for each of the cards detected, model it as a simple in this case is simple cuboid and then represent it however we want in terms of a plan view, whatever. This is kind of an early idea.

ISSUM is really trying to move that to the next level. And to give you an example, if this loads, let's see. There we go. Okay. So this is the York Keele campus, many of you have been there. This is a 3D model that was created in conjunction with ESRI Canada and my colleague in York Gund Hassan. And you can see we have lots of building models, but we also have dynamics. We have vehicle traffic, we have pedestrians. I want to emphasize that in this case the vehicle traffic and the pedestrians are not real. They're sampled from a model. But the ambition is now to connect this kind of dynamic, three dimensional model with real data coming in real time. And as a test case, we're working on this building coming up here, which is called the York colonnade. So it's just a breezeway that goes along the common there. And we have two cameras at either ends of this breezeway to give us realtime video of the pedestrian traffic.

So my post Doc Peel Claudio is working on this problem of connecting up all these thoughts. It's quite challenging from a systems engineering perspective. We've got computer vision, machine learning, photogrammetry and geomatics and computer graphics and simulation and crowd modeling and so forth all coming together. But at the end of the day we get something like this. On the left, what you're seeing is the raw video I talked about earlier, but now with the detected and tracked pedestrians in that space now because we have algorithms again to automatically understand the pose of the camera. When we detect those people in the 2D image, we can back project them to the 3D scene and Geo locate them on the ground plan. And that allows us to situate each of those people in our three dimensional model and represent them as avatars.

We end up with the result on the right, which is now in a three dimensional model. This is a video of that. But if I had the actual model here, I could turn it around for you. You could see from different points of view, you could even assume the identity of one of those pedestrians. You could click on their head and become them for the period of time that they're within that field of view. That's an example of our vision for ISSUM. Of course there's lots of value added deliverables that can be entailed by that in terms of statistics that can be used for these various problem domains. But what I want to focus on is really that idea of converting raw data to a three dimensional model. And so just to summarize there are many we think, many opportunities for what we call visual AI. AI using visual observation of human activity to help us in our daily lives in particular in these dense urban environments that most of us live in.

But we have this concern about privacy. And so to kind of achieve the first goal without incurring the cost of the second barrier. I'm introducing this idea of three dimensional virtualization and in particular, I'd like to advocate for doing that near the edge. So that means relatively early in the processing of the video data and the motivations are actually partially technical. So as I've tried to argue we need to do this anyway in order to really understand what's going on, we need to understand it in three dimensions. We need to have a three dimensional model. And this is necessary for both human and machine interpretation, but it also has the great advantage of anonymizing the individuals that are in the scene and therefore protecting their privacy.

And maybe most importantly, that gives us the potential opportunity to share these data much more widely. In other words, to democratize this information, which currently is really in the hands of just this very small number of organizations. And that's important partly to have access as private citizens, but again, to foster innovation in

Canada, I think that would be a great idea. Okay. Thanks very much.

Terry Sachlos: Thank you James, that's very interesting. Victor our next speaker.

Victor Garcia: Great. So we'll spend a few minutes talking about how big data and it's connected technologies, machine learning, artificial intelligence are transforming health sciences. We have been seeing for several years now how entire industries, whether it's retail, aviation, banking, agriculture, manufacturing, are rewiring themselves to be able to provide more effective services, increase profits, serve their customers, better, change the game. All of them have one thing in common. All of them have been learning how to use the data, information that is coming to them from all kinds of sources to help them improve the way they do business. And the same is the case in the health sciences sector. There are many applications of data, machine learning, artificial intelligence that are being used in an attempt to improve effectiveness, efficiencies, get better results, reduce costs, something that the health science sector has seen before.

The transformation of healthcare has happened for many times. And the difference here is that when often they have been asked to do more with less, often less money. Right now we have the opportunity to do more with more, more data, more and better tools and that is exactly what is happening. At the heart of this is something we know as big data and this as you probably know, refers to huge amounts of data coming at you from multiple sources, different volumes at different velocities with different types of variety and value and veracity and data science is something that we have used to marvel and understand and create and use and share data and a set of tools that we use to analyze often using statistical methods to find patterns meaning and get insights to tell us what to do with the knowledge we get from data.

We should be careful about the marketing aspect of big data because we know that there are some big issues with data. Everybody will tell you that that is the thing that we should all be into. The fact is we are having problems with the inability for big data to represent the real world without true biases, huge problem. We are having difficulties with integrating data that comes from multiple sources, sometimes from different countries. Sometimes different formats, we're having problems actually proving that getting all these data and paying huge amounts of money to transmit this data is actually resulting in innovation. So these are things that we have to be careful as we tried to understand what is the real capability of big data to change the way we do things. Despite these issues, we are seeing huge applications in the area of health sciences and huge applications that have to do with how we do things, how we become more efficient, how we use resources better.

Again, one thing that we have to take into account is that the use of big data in health science is very different than how we use big data in other sectors, retail, manufacturing aviation et cetera. Data used whether for research or diagnoses or other areas of healthcare, have features that are very unique to healthcare are very distinct, not only distinct to data in other cases, but even distinct to how we used to get data in the healthcare sector. Therefore, some care has to be applied to how we really use big data in healthcare. Nevertheless, we have huge applications and in fact the interest is so big that a whole new industry has been created. We call it healthcare analytics. This is a term that has emerged over the past few years to describe the activities used by a number of companies that focus on using data from four main areas, clinical data, insurance claims, cost data, pharmaceutical and R and D data and patient behavior and sentiment analysis data and said there are some huge players among them.

IBM, SAS, Optune, Truben, Cerner, McKesson, Microsoft, Ge Healthcare, and a number of places that are kind of selecting one or more of these four areas too heavily play in the healthcare space. So why use big data? What are the potential benefits? So let's briefly look at forder areas, where we put really a group, the majority of use of big data is in the healthcare space. One is increasing process efficiencies. So using information and using insights to help us use what we know as value streams better, use resources better, assigned people better, reduce issues fraud and essentially trying to improve the effectiveness of providing health care. Big Data is also helping in terms of facilitating practitioners work by making the work of diagnosis better, supporting the practice of telemedicine, integrating electronic medical records with electronic healthcare records, digital imaging labs, et cetera. We are using big data to speed up and improve the results we're getting from research and development an area that we'll talk briefly about.

And using things like genetic data, DNA analysis and trial results data to improve the way we produce results from research. And finally, we are using big data to improve the experience that patients have in dealing with the healthcare sector by making patients aware of what's going on by making them part of the process of creating and using data, by using an instant alerting system to resolve have problems by helping prevent opiod use and for the most part, leveraging some of the new realities of big data on them social media and social networks to use patients to create and use their own data.

Let's look at one example. And that's the example of improving the value stream process efficiencies. The value stream is when you have a series of linked activities in a process that you need to accomplish [inaudible]. For example, when you perform a surgery, you don't just look after the operating room that had a whole bunch of connected resources and people and parts of the hospital that have to come into an integration effort to ensure that nothing's happened. And this is a classic problem with any industry that has to deal with assigning resources to use in shifts. Too many resources, too many people, you lose money, not enough resources, issues of quality or in the case of healthcare, potential fatalities. The example we have here from Paris, these are four hospitals that using an open source engine use data from 15 years of expertise and history, treating patients together with external databases, whether public holidays, flew patterns, tourist volumes, sporting events, education institution schedules to start predicting how many patients will become into any particular time.

And using these data in make the process of assigning resources to those terms more efficient, highly efficient. A closer example is Humber River hospital. I actually serve in the advisory board at this hospital and this was the first time that data and AI power system was created as a command center. Not just to look after the physiology of the patient, but the physiology of the entire hospital from beginning to end. For the moment a patient comes in to the moment the patient comes out and every aspect controlled by information coming from a variety of systems, the Meditech EMR system, the Med Works bed management system, the Ascom communication system and this information is collected and displayed tiles into high quality HD displays. To assist practitioners and different staff members in terms of what is going on, real time information.

For example, the emergency department tracks visits by patients all the way from door to doc, keeps track of multiple visits, can detect if a particular person is coming back every so often, that could lead to potential issues of opioid abuse. It can spot patients with abnormal readings. For example, an abnormal cardiac enzyme in a non-cardiac patient could be representative of a problem that has to be looked after. So all of these data has been analyzed and people that are been dispatched to provide a better service. The system has been so effective that in the first few months of operation, it gain a 20% efficiency and by the next time that it was measured, it had achieved a 40% efficiency, which is equal to about 40 new beds being created. In fact, one of the reasons that we looked at the possibility of using the system was that the hospital was right out of beds.

We either had to go and get more money to build more capacity or find a way to use technology to become more effective and more efficient and this is where it was developed. Another example in using big data to assist practitioners who are relying more and more on data analytics to make better diagnosis and improve treatment. Essentially what data is providing them is the ability to use and get more value from the reporting phase. Essentially what happened, the predictive phase, what might happen and the prescriptive phase. Why did it happen?Wwhy it is likely to happen again. So we had examples such as IBM Watson, which is heavily being used by combining huge amounts of data, genetic data, similar data from other patients and providing assistance to physicians in terms of improving the way they diagnosed. One of the reasons why this is needed is because misdiagnosed patients is a huge problem. Study produced by the Mayo Clinic in the US show that over 20% of patients are misdiagnosed.

So physicians typically have to rely on either taking time off second opinions or a number of techniques to ensure that their diagnosis is correct. By using the technology, we are essentially automating the process and becoming much more efficient than humans could be in analyze data, genetic data, historic data, and automating the process of providing answers back to the physician. So not replacing the physician, but making his or her work more effective and more efficient. Big Data is really helping the area of R and D it's almost as big deal as providing a better picture. And R and D depends largely on pictures. If we look at the history of Galileo with a primitive telescope, he used the telescope to investigate the cosmos, uh, and he in fact maybe made a drawing of the moon than many, many, many years later after we have used different telescopes or satellites, we found that it was like really quite accurate.

He had a pretty good idea. He took that data that the telescope gave him to produce a picture. Now the fact that we over time got more and more data, better pictures. This then replace the fact that when Neil Armstrong was about to learn the lunar module, he saw his own eyes that he could not because there were rocks that nobody has seen. So now a better picture, allowed him to make a better decision. So the same is being done right now by getting data from a microscope, which is essentially a telescope backwards, and producing data to help us understand what are the results, what our trial tests, data information giving us. How can we make the process faster? How can we rely on information that is more methodical? How can we do things faster and cheaper? And one example of that is what happened with the Ebola virus problems when an automated system found that in a fraction of the time, that it will take in fact, in a few hours they were able to find a potential treatment that in fact worked and was applied to solve the problem.

Now, most medical treatments that we are familiar with are designed to treat the average patient. Often using a one size fits all model. For the most part a biological chemical centric process that includes diagnosis, treatment, observation, cure, over simplification. But that is typically what happens. The area of precision medicine that takes advantage of big data, machine learning and AI is helping us improve the way we do this by customizing, by taking information and customizing the way we treat patients and is having tremendous effects in terms of using genetic data, allowing us to actually, change genes right at the source by injecting information, changing information, and essentially modifying the potential source of problems before they happen. Let me just quickly advance because a lot of time here. So I just wanted to finish this by talking to you about some of the issues that exist.

In fact, without going far we can look at our eHealth Ontario electronic health records system that would hardly be considered anything, but probably the best example of the use of big data to help the healthcare sector. The combination of data for multiple sources for multiple stakeholders to allow us to make better decisions. Yet it's system that has failed after spending 17 years, $8 billion, the latest the auditor reports have said the system has not worked, hasn't been deployed. If not finished, nobody's using it to the extent that it should be used, why not? And this points to one of the many reasons why these things are often very, very, very complex. And it has to do with the complexity, difficulty, technical issues of collecting, sharing, securing as well as in some cases privatizing or governing data, particularly where we have to do to deal with highly sensitive information, patient data, location data, data that is not different than the financial data we like to protect.

So when we look at the fact that with big data comes a big responsibility, we have to look at the issues that come with it. Integration issues, data sharing issues, profiteering issues, which we know in terms of using data collected and reselling that in some cases anonymized to trade, privacy issues, cyber security issues, socioeconomic implications and of course, legal and ethical issues. I'll finish there because I've run out of time. I would try with Aisha.

Terry Sachlos: Thank you victor. Thank you very much. And our last speaker, Ian?

Ian Stedman: I'm going to pick up on a lot of the same things that we've been talking about and then kind of bring it into a very specific case, kind of like you were doing with Humber River. But what I want to do is talk about the integration of clinical decision support tools. So the kind of thing that was being designed in Professor Degas Dino's presentation. So the idea that we're gathering data to help us understand breast cancer diagnosis and treatment. You're designing a tool that can help with clinical diagnosis and clinical treatment. I want to talk about the integration of those tools into the electronic health record, which again picks up on the end of Mr. Garcia's presentation. What I want to cover though is recent developments at a local hospital as well, SickKids.

So early this year is SickKids announced the first ever chair and biomedical informatics and artificial intelligence. Dr Anna Goldenberg here. She was already a CFR chair, vector associate head of Health CRC Canada Research Chair. So incredibly accomplished. And then shortly after they announced her as the first ever chair of AI and bioinformatics, they announced that there would be a new president. And the new president was currently the chief pediatrician, Dr. Ronald Cohen. And he has without a doubt, a pretty bold vision for the hospital and what they're going to do. He would, as I've highlighted here, accelerating improvements in disease treatment and prevention through the precision health physician child health informed by cutting edge technologies such as AI genomics, advanced imaging. So this is kind of what we're talking about at Humber, but now brought to life out of children's hospital just down the road. On March eight, after they announced his appointment, he actually flew off to Hong Kong to announce a $25 million donation, a joint donation to the University of Toronto, and to SickKids to fund joint center for healthy children.

The center will focus in large part on leveraging these tools and the integration of these tools in order to do what they're going to call precision prevention at SickKids. This is what I want to get into today to do something like this. So it's a big enough idea that you'd probably need a task force. They have one, they had a task force and it predated these appointments, these announcements. So the task force here has four circles that in the middle have the word ethics. So let's go through that a little bit. The first thing you need if you want to integrate artificial intelligence into your clinical care, into your electronic health record system is ideas. You need solutions you need, you need tools, that could actually work. And then if you're going to do this in a public health care setting, you probably need somewhere to store those tools and the data that they're using.

So at SickKids, they have the center for Computational Medicine and high performance, HPC for health, high performance computing for health. It's a big cloud based infrastructure there. So you're going to have to design these tools, house them somewhere and then figure out how to translate them back into clinical care. The bottom circle basically meaning integrating them in a seamless way into the electronic health record. That all sounds nice, but you can't do any of this if you don't have a vision for how you're going to do it and you can't really have a vision for how you're going to do it if you don't have some sense of what the implications are, the governance and policy and ethics implications. So before you really get anything rolling, you've got to do a lot of social, legal, ethical background work to really put these pieces into operation.

I look at this chart and I think ethics in the middle isn't really appropriate. Ethics would have to be woven throughout every decision, every kind of process as you design something like this, and we'll get back to that in a little bit, but this is no inexpensive or small undertaking. To do something like this as Mr. Garcia pointed out, you'd ultimately want to design a system that allows all of these different things, all of the different data, patient reported outcomes, everything to be integrated in a seamless way so that it can be instantly basically fed back at the point of care into a tool that helps with treatment and diagnosis. I'll give you a quick example because I think it's important to talk about patients a little bit and have a real example. This is 19 out of 82 pages of a health record for a young man who visit his doctor about 180 times.

This is just for a family doctor and the first 18 years of life, so this is just an 19 out of the 82 total. This is written in the stereotypical doctor's handwriting. You can't really read it, so it's okay to put it up here and I have consent. The patient went undiagnosed for 32 years, was subsequently diagnosed at the age of 32 with a one in a million rare genetic disease and the genetic disease had seven different symptoms. If you could read this, you would find out that these seven different symptoms repeat over and over and over again in these 180 visits to the GP throughout the first 18 years of life. Right now if you're a doctor and you want it to diagnose this disease, you'd have to know something about rare auto inflammatory diseases, which means you'd have to know something about this chart.

You don't have to know this chart, but that doctor would have to know this chart. In this chart there are 26 different rare auto inflammatory diseases and on the bottom left for you, there is 11 different systems that are implicated variously throughout these different diseases. So you could have someone come in with different manifestations of whatever's wrong with them, different symptoms in different systems, and you'd have to try to figure out how to get to an answer for them. I mean moving forward this are the records again. If you have a program like Fino tips, you look it up and these records were digital, the program would have picked up the diagnosis by the sixth visit.

There's a pretty powerful idea there in that here's a real example of the possibility of giving in the case of SickKids a child, but a person 31 years of their life back because in this case there was also a treatment and the treatment made all the symptoms go away and drastically changed quality of life. This all matters, but it's also very complicated. I want to get back to some of the things that we were talking about with the challenges of designing a system like this. I don't want to focus on the actual tools. I want to focus on the challenge of the system design,

Where the data will come from and how it will be collected. Of course, in a clinical setting, in a hospital setting, every single person who you're collecting data from is generally speaking a vulnerable data subject. They're either a patient or a parent or someone who's there for a reason that relates to their health, where they're relying on someone else for treatment and care. How you collect it is going to be pretty important. You also have to ask questions about what kind of data you collect because there's going to be low quality data and there is gonna be high quality data. Imagine you were trying to figure out that diagnosis earlier and you were collecting data about one system without looking at how it related to another system. So for example, that disease, that diagnosis it implicated ophthalmology, dermatology, so the skin, but also internal medicine. If you're asking that patient the wrong questions, you're not getting a complete enough dataset.

Understanding if you're going to build a tool like this, how do you get complete accurate and clean data that can actually be used by the tools that are integrated into the system is going to be pretty important. Let's talk about data management for a minute. If you're going to do this, you're going to want to keep that data somewhere. In healthcare, what's unique about it versus the the private sector setting we've talked about enterprise setting is that you're going to be working in a publicly funded institution. There is expectations of how you handle these things is going to be heightened and it's going to be a lot more, I think, transparent the way that the world currently works, there is going to be a lot more scrutiny on it.

How long can you keep patient data? What we have rules about that. But given what you want to use it for now is diagnosis and treatment and care. That's going to be an in determinant kind of idea. You want to be able to use those tools in your system forever. So what does that mean for data? How do you keep that data and can you.? Those are big questions that you'd have to kind of address for free to start building. Now in a healthcare system, you might also want to use data that someone else has gathered or that someone else has ownership over or stewardship over. If you're a hospital designing a system like this, you would first have to figure out how do you do that? What policies does that other institution or provider need to have in place before you can cooperate with them?

This comes back to Professor D'Agostino's work on those relationships. Just to build a tool to integrate into the system are completely complicated. But if you're going to build the system to integrate all the individual tools, you have layers upon layers upon layers of different considerations, moral, ethical, private, like everything is going to be in there but many, many different times. So you need a lot of hands on deck to try to figure this stuff out. In this case like you would want to make sure that any institution you do work with has standards that are at least if not as good but greater than yours.

You also want to look at data access. So in a healthcare institution that's also a research institution, you have a lot of different players who are going to want to have a stake in this data. Not only do you have investigators or the researchers it's the clinicians, the patients and the families. The institution is going to use this data to try to figure out when do we admit someone, when do we refer someone? How do we get people in and out of our ER? They're going to want to do administrative kind of work with it. And then you have third parties. So there's always this question. This morning Carol brought up the Toronto Star articles about what can you do with data? Can you sell it? These are questions that although may seem obvious to us in this context, we'll have to be addressed upfront before you design something like this in a public space.

And if you want this to work efficiently, you most certainly have to pool the data. You most certainly have to give these data scientists who are trying to find patterns, access to large amounts of data that maybe had gathered in a disparate way. So in the healthcare space and in this science research space, a lot of the people who are working in these research help hospitals also have appointments at universities and historically have been trained to have control over their data. The currency of academia is to publish and so you don't give your data up. That's going to have to change in many ways if we're going to take all of these useful chunks of data that had been gathered in silos and put them together to understand what they're telling us, we have to somehow move past in these settings, the academic model of publishing.

How do you incentivize the sharing of that data? That'll be interesting. Privacy, security and other legal considerations. I don't want to touch too much on this except to say we've talked about it a lot. One of the interesting things that I think will come up is when you're putting together a clinical trials. I sit on the research ethics board at SickKids, generally speaking, a researcher comes and they have an idea of why a clinical trial might work. Here's a drug we want to try it on this population evidence suggests it should work. We need your approval. This is what we're looking for. This is our inputs and outputs that we're expecting. With data, it's totally different. Sometimes you're looking at data just for the sake of looking at it. You're running an algorithm through it to see what patterns it can uncover, what trends it can uncover, what predictions it can make.

You don't go to the research ethics board up front and give them your lists of 2 million variables that you're going to look for. That's not how it works, but you still want access to that data. So it's going to really matter how that data is collected. I think what you're going to see going forward is that clinical trials change how we ask patients for consent. The type of consent we ask them for when you're conducting a clinical trial is going to be different. Secondary use is going to be almost as important as the primary purpose of the clinical trial to say we don't know, the seven systems are implicated. You know how you're going to target one through a clinical trial, but you also know that you know nothing about the six other systems and how they relate. So you're going to want secondary use just for the sake of exploring. And I think that's going to be a huge difference in the healthcare setting is that that's going to be much more integral upfront and in most if not every clinical trial.

And I think as we understand better how to anonymize and de identify data, the types of research that people will consent to and be interested in conducting will change as well. I wonder if you'll have a lot less of a focus on the actual output of the clinical trial and just the focus on the purpose. We don't expect that we're going to get the answer, we expect that we're going to get the data through this kind of interaction with patients. And so you might need consent to do all those sorts of things at an institutional level. There's going to be issues with getting consent for old data, use of old data, so data that's already been accumulated and that you don't have secondary use consent from the participants from. I think that's going to be a big issue.

One of the biggest ones that I think you're going to run into is just educating the workforce. So I had as professor D'Agostino mentioned the baby on Monday at Sunny brook and we were told to show up two hours before, it was a scheduled C section, two hours before and it wasn't an airplane. We don't have to be there for going through security checks. We answered the same questions to different nurses sitting in front of the electronic health record for two hours. We must have killed seven people that she didn't smoke during pregnancy. There's a ton of data collection that's going to be important. There's a ton of opportunity to create greater efficiency in data collection. But if the workforce doesn't understand how important it is to collect that data cleanly and smoothly and once for the patient experience, then it's not going to be seamlessly integrated. So a big part of building these types of systems, if they're going to be at the point of clinical care, is going to be educating those people as to the value of these.

We're collecting this data and we don't know why we're collecting it, but one day we might know what it tells us. So take it seriously. Every single one of you, that's a big lesson. That's a big undertaking. Lastly, harmonization. Data comes from different places. You're going to want to put it together. You're going to have tools designed to accept data a certain way. That's a huge undertaking. But you also want to make sure that you understand how to prepare it, to clean it, to make sure there's not a lot of noise in your data and to anonymize it. And then when you design these tools to validate them. Now data validation is a whole conference in itself. I just want to put it on the table, but there's going to be a ton of different issues that come up with the integration of data in the electronic health record or these clinical decision support tools. And they're going to have to be addressed in advance in this setting. And I think that's uniquely different about healthcare is because it's publicly funded, there's a lot less forgiveness from the public looking in at it.

We'll have to get all of these questions right. I think I have zero seconds left. So let me just close out by saying, I think the takeaway here is again, what's been said, there is a lot of opportunity for collaboration. It's necessary we have these tools, we have computer scientists that are just itching to create these clinical decision support tools and be allowed to use them. But the socio legal conversations are behind. So if you do anything today, meet someone new who has an interest in something similar to you but maybe isn't the exact same as you and start making connections because the biggest thing we can do is start creating these collaborations to get these conversations going. So the social can catch up with a technical so we can really get something like this into operation. Thank you.

Terry Sachlos: We're running out of time. We only have time for one question. So any burning question? This gentleman had his hand up first. Is there a mic? It's running, mic on foot.

Audience member: My background is the public sector. I used to be very involved in politics and I think I'm very interested about this secondary data collection? So do you think it is absolutely necessary does government have to legislate, sort of like the framework or consent the idea. Cause nowadays it's only the primary collection. Like do you think there is a need for that? Like sort of restricting or limited like a use without consent?

Ian Stedman: Consent is consent. From my perspective, sitting on a research ethics board, I think the most important thing is that how we think about consent has to be much more robust. How we talk about these things. You have to allow people to consent to things. But what's happening now is that people don't understand the things and they're harder to explain. It's hard to explain to someone that I want to collect your data for a secondary use. Can you consent to that? And they say, what does that mean? And you say, I don't know yet, but we might know in the future if these tools that are being developed can be used. I'm not so sure I understand how the government would get in there and stay ahead of it to be honest. But I think what you have to do is create an expectation. And we do this through the Tri Council policy statement in ethics. An expectation that when you have a setting like this where you have vulnerable populations of patients, et Cetera, that you're taking consent seriously robust informed consent before you collect data. Yeah.

Terry Sachlos: So I think we're running out of time. Does anyone else want to add anything else? I think in the interest of time, we will call a wrap. Hopefully you can hang around to talk with the speakers if they are available, please join me in thanking our panelists. Thank you very much very interesting.

You were mentioning in your stuff.