



Presidential Commission
for the Study of Bioethical Issues

TRANSCRIPT

Ken Chahine, Ph.D., J.D.

Senior Vice President and General Manager
Ancestry DNA, LLC

Laura Lyman Rodriguez, Ph.D.

Director, Office of Policy, Communications, and Education
National Human Genome Research Institute

Meeting 10, Opening Remarks and Session 1
August 1, 2012
Washington, DC

DR. WAGNER: Our next session is on protection of private and public genomic databases. And if Dr. Chahine and -- Drs. Chahine and Rodriguez would come forward. Wonderful.

We'll be hearing first from Dr. Ken Chahine. He is the senior vice president of Ancestry.com, well known online resource for family history and general manager for Ancestry DNA where he leads the development and commercialization of population genetics at the website which is the same name, AncestryDNA.com.

Dr. Chahine is also professor of law at the University of Utah, held various positions in the biotechnology industry including president and CEO of Avigen.

Dr. Chahine, thank you. It's good to have you with us.

DR. CHAHINE: Thank you. So I decided given the limited time not to have any slides. I just have some brief remarks. I suspect you're going to have some questions for me so I'll keep those brief.

So with that I want to thank the Commission for the invitation to share my thoughts on genetic privacy. And the views that I'm going to be talking about today are really my views and not those necessarily of Ancestry.com or Ancestry DNA.

And as you said, Ancestry.com is the world's largest online resource for family history. With our collection of billions of digitized records our customers can research their family histories, build family trees, upload pictures and share stories which they do.

Ancestry DNA is a subsidiary of Ancestry.com and we recently launched a new service that allows customers to use genetics to trace their ethnic origins and find distant cousins. And Ancestry DNA's service preserves and analyzes genetic information,

genealogical pedigrees, historical records and other information from people all around the world in order to better understand global population genetics and create products to help our customers make discoveries.

And we hope that our research will be an invaluable tool for future generations and engage the interest of a wide range of scholars in genealogy, anthropology, evolution and medicine.

So as a provider of genetically derived information and stewards of genetic material Ancestry DNA understands its responsibilities to the customer. The entire service from collection to banking was designed with security and privacy in mind. For example, the DNA collection kit is designed to be anonymous so no name or other personal identifiable information is included in the packaging. And by using a random alphanumeric code which is linked to the personal identifiable information by the customer online the anonymity of the individual is preserved throughout the testing process and beyond. In addition, all primary genetic information is stored in a physically segregated server on multiple hard drives with limited and restricted access.

We have conducted multiple third party penetration tests to assess any weaknesses in our security and I would say in short that the system is designed to require multiple independent and internal breaches of security to be compromised. And we plan to continue to learn and make the system more secure as time goes by.

Another point I want to make is that Ancestry DNA took the position of being transparent. So our informed consent is approved by an institutional review board. It's an opt-in consent. And we also voluntarily address the often overlooked issue of

ownership by clearly stating in the terms and conditions that the customer retains ownership of their DNA and their data. Ancestry DNA retains only a license to use it consistent with the informed consent.

And I think importantly our privacy policies and procedures have not been a reaction to public demand or outcry, and I would say quite the opposite. In a survey of Ancestry customers only about 14 percent raise privacy as one of the reasons why they didn't purchase a previous version of the DNA test that we had. Instead, the primary barriers cited were cost and lack of understanding. So we have been proactive in safeguarding the customer's information and genetic material.

Let me make just one other point that we were addressing at the end of the last session that sort of vexes me and that is that even though we clearly put an emphasis on the private and sensitive nature of our customers' genetic data we should keep in perspective that our DNA is anything but secure. We leave it behind almost everywhere we go every day.

And why would anyone wanting my genetic material spend the time and effort to hack into a secure data system when they can easily follow me to the nearest coffee shop and retrieve my used cup. Okay? So I think that while doing so I think is certainly reprehensible, to the best of my knowledge it's not illegal in terms of the identification of my genetic material.

So in other words, there are realistic limits to what we can do, and I think while we should certainly take steps to mitigate these risks it's something we should just keep in mind. So, thank you. I'll stop there and answer questions.

DR. WAGNER: Tell you what. We'll move to our next speaker and then

get questions together. And that is Dr. Laura Lyman Rodriguez. She is the director of the Office of Policy, Communication and Education for the National Human Genome Research Institute at the NIH.

In her work at the National Institutes of Health she has assisted in shaping NIH policy for sharing genomic research data and has helped implement the genomic data-sharing policy across that Agency. And we're pleased to have you join us. Welcome.

DR. RODRIGUEZ: Thank you. And I'm going to break the trend and I'm going to use slides but I'll do my best not to let them slow me down. And hopefully they'll be useful in going through this.

So I am here today to talk about how NIH has worked through the changing landscape both scientifically and from an ethics perspective and a technological perspective with advancing genomics research for public good while so many questions remain unanswered and in fact answers are continually being updated.

So, the focus of what I will talk about is a model that we created through the data-sharing policy for genome-wide association studies which has now been in place for approximately 5 years. And this policy created an expectation for sharing of whole genome information in a central NIH repository. And so it was an extension of data-sharing practices that had been in place and our long-held belief at NIH but in this case it removed any type of monetary limit whereas previously data-sharing had come with a certain level of investment in a grant or an award. And this was based really on the nature of the data and the value of the data for the research community and the potential to return benefit from it.

We're guided through all of this by the fundamental principle of

maximizing public benefit from the research investment, and that investment is speaking to resources, federal taxes, but also really the investment of the participants that contribute their time and their information, very sensitive and personal information to the primary research studies which then come into the central resource.

And embedded within this principle is also several values which we really look to in each of our policy mechanisms that we put in place to try to respect the participant interest, to promote the data-sharing and to have freedom to operate so that, again, there's the maximum amount of innovation and public good that can be returned from the community resource developed through the database.

You all have the policy and so I'm going to walk through the model fairly quickly. But just to remind you there's different phases of this research and it does begin with the distributed system of the individual research studies, participants interacting with investigators where there is a relationship and that relationship is scripted in some ways at least from a paperwork perspective through the informed consent document.

And this again is something that NIH did in moving forward with the policy. While we were going to focus on de-identified information we did move beyond the regulatory boundaries and attach consent to the use of the data through this resource for the first time. And we did that by asking every institution that submits data to the Agency to tell us what the data use limitations will be based on the informed consent form that the participants signed at the time. So that data is always used in subsequent time periods based on those parameters that at least from the local institution's viewpoint would be appropriate for that study population.

In terms of privacy we also advise, because the primary institution will hold the identifiers. Again, we take everything in in a de-identified or coded way, advise investigators and institutions to consider whether a certificate of confidentiality would be appropriate as a mechanism to attempt to safeguard the confidentiality of the participants in their studies.

Once the data come into the data repository, again I've already mentioned that all data is coming in in a coded way. That was very intentional so that, again, the NIH is not holding any of the traditional identifiers.

The standard that was adopted was the 18 identifiers from the privacy rule within HIPAA. This was from both a practical standpoint and a desire to harmonize what our standards were across different regulations that were operating within trying to create this resource and accomplish the research at several different sites.

We also for the data repository -- my slide's not going forward anymore. But for the data repository we did also seek a certificate of confidentiality. And this was after a great deal of internal debate as to whether or not that was appropriate or feasible because we were taking in de-identified information. And after the deliberations within the Agency determined that again because of the volume and the nature of the data, while it was not directly identifiable there were sufficient data in there that it warranted this level of protection because of the various ways, as Ken mentioned, going back and matching the data through other sources that would not be related to our repository itself.

From a technical perspective I'm not going to get into specifics here partially because I can't talk about them very well. I just wanted to mention that of course

the security in terms of firewalls that are built and how the data is managed internally for the resource is layered based on the type of information that we have. So all data comes into the basic firewall within the National Center for Biotechnology Information and then when it first comes in it's behind another firewall that has a particularly high level of security because that is the data that we are still needing to confirm does not have any identifiers. And then after that's been confirmed it moves over to another layer. All of this is very much locked down even internally within NIH to only the dbGaP staff. At the level of data users there are also expectations for anyone that NIH provides access to that they are able to meet particular data security standards at their local institution. This is confirmed by their institutional officials as well as their local IT official so that we have every confirmation that we can reasonably get that they are in a place with the appropriate infrastructure to manage the data in a responsible way.

And then the final phase of course, moving from the repository, the point of building a repository is so that secondary investigators can go out and use the information. And so again, individual requests will come in for specific research purposes. These requests as I'll talk about in a moment are reviewed based on the data use limitations and they're actually requested by their consent group. So the database is organized to reflect consent.

It has been designed from the beginning as a two-tiered system where some information at the higher meta level is in an open, publicly accessible database. And then the genotype information, the phenotype information is all available through controlled access.

And that is the controlled access portion where a specific research project has to be proposed, that research use if the data access is approved is posted publicly so that there's transparency about who has data and what they're using it for. The agreements are cosigned by the institutions, not just the investigator and again, IT officials are named on all of those applications.

And then we have data access committees. We actually have 15 at the moment. They're organized in some cases by scientific discipline so that they have the knowledge to evaluate the proposed research. In other cases they are by project based on the size of the data and the number of access requests expected.

But the primary purpose of those access committees is to review the proposed research use with the data use limitations and then also to look at the infrastructure that's available for IT security to safeguard the confidentiality of the participants.

You all also received in advance the data use certification agreement so I am not going to walk through the particular terms and conditions of use, but not surprisingly we're very explicit within there about things that investigators that are approved for access should not do with regard to ensuring privacy and confidentiality. And again, also have several conditions that speak to only using the data for the approved research use. And again, that goes back to the level of respect and importance that we place on consent for future use of all of the data.

There are also other intersecting regulations and laws that have to do with privacy around government records which all of these data are when they come into the government database. The Freedom of Information Act is one. As I've stated all of the data

are again coded and de-identified. So again we had a policy question to ask ourselves because this then could be released under traditional circumstances.

But after looking at again the volume and nature of the data the NIH made a policy decision that they would intend to deny FOIA requests that came into this so that we could assure participants that we would not -- that any access and use of the data would be for research purposes and it would come through this controlled access mechanism with review by the data access committee.

This policy was looked at by a working group of our advisory committee to the director for the NIH director who took a look at all of our policies and protection mechanisms for this repository when it was first launched. And they recommended going beyond this policy-level decision for the FOIA request and actually trying to seek a legislative exception so that the most robust protection could be available to again ensure the public that their data would be used for research purposes only.

Similarly, compelled disclosure through subpoena authorities is another issue that we received a lot of comments on and that we had concerns about. Again, I've already mentioned the certificates of confidentiality that we put in place to try and protect that issue.

And also within the policy -- and I will finish up very quickly -- we allow for exceptions to data deposition, recognizing the fact that there will be reasons not to deposit the data where it's not appropriate in dbGaP. And some of those do have to do with places where we have very localized or identifiable populations and privacy may be a greater concern.

Our data use experience to date, the numbers are up on the slides. And the point here is just to note that the resource has been growing steadily. There are over 300 studies in the resource now. We have over 3,500 approved projects. It is contributing greatly to the scientific literature and advancing our scientific knowledge.

And our experience in terms of problems with this has been very low relatively speaking, and they have largely revolved around technical issues, some small compliance issues, but nothing malicious has come through in our ability to monitor this.

Stewardship is important. We've talked about trust before, the other speakers have and I think that's something that we also built into the system from the beginning. We have our top-level leadership at the Agency that engages in the policy and looks at how the policy is being implemented locally in terms of at the Agency as well as extramurally on a regular basis. And that has proved fundamental to our ability to deal with changing technologies, changing methodologies that alter the context of what is or is not identifiable in the data resource.

And then I'll just end I guess by projecting out to how this is all put together. I mentioned at the beginning that we're dependent upon a distributed system of research. And so there are protections in place at all of the local projects, the local institutions there's oversight. And then at the highest level there are national-level policies and regulations that we're dependent upon and which we try to use as instruments to guide what happens at the finer grain.

But all of this is happening in a context where -- around trust, where we have oversight and policy, researcher conduct and research community conduct contributing

to the generation of public trust. And public trust is fundamental to the ongoing support of these activities and to participant willingness to actually contribute to the research. And without the participant willingness to contribute to the research we will not move forward at all. And I will stop with that.

DR. WAGNER: Thank you very much. Well, actually -- okay, I'll hold mine. No, no, no, go ahead. And I've got Amy first and then John.

DR. GUTMANN: I want to say something and ask a question about coffee.

(Laughter)

DR. GUTMANN: Because we hear over and over again the true statement that you can take my coffee cup and seek, you know, get all the information out of this that I can get by, you know, asking for my whole genome to be sequenced, right? So here's, I think, and I want your reaction, here's the important point that that makes and the important points that it leaves unresolved, right?

I think the important point that it makes is that the questions about respect for persons and privacy and whole genome sequencing do not boil down to ownership, right? Because I can't own what I leave behind. It's just, it's -- right? I'm spilling it and I'm leaving --

(Laughter)

DR. GUTMANN: And I spill it all over the place, right? So, however, so I think that's a very important point to recognize. Oftentimes you cannot assert ownership over your genetic material in the sense that you leave it behind all the time.

However, the interesting fact which I want to then ask a question about is that we have all these protections about how it's used. And I think that's the point a number of people have made, Nita and Christine and others, which is we still care about the particular uses. We legally care about it and we morally care about it.

And I just want -- I'll give the simple example about why that's rational to do so. Imagine a world, and it's a world that could exist in our world, where some political group who wanted to cleanse our society of genetically inferior people organized. They're allowed politically, every legal right to do so. And they went around picking up these cups knowing, identifying whose genome sequencing it was, publicizing it widely and calling for the eradication of the people they saw as genetically inferior. That would be a nightmare. They might not violate any law depending on how they went about it, but it would be a moral nightmare. And therefore we applaud organizations like yours that put into place restrictions on how you'll use it.

And my question to you is what kinds of restrictions do you think are, you know, we should as a Commission care about saying are necessary to have the trust that you both care about a lot so that people can feel free to contribute to what will make them know more but also contribute to the public good. It's a big question but I think it's important for us to get that out because it's what this Commission really cares about.

DR. CHAHINE: Why don't I go first. Okay, fair enough. Yes, so I raised that issue. Obviously the committee has contemplated it. And I just want to make sure that no question, I mean we put a lot of emphasis on it but there's other ways to get it, right? So that loophole exists and I think a bad actor is more likely to target someone and go after that

genetic material specifically than trying to go after broad de-identified information and then try to re-identify one specific person. Okay?

So in terms of your question it seems to me that having a, you know, something that sort of protects an individual at sort of the last point in terms of publicizing or disseminating genetic information being, you know, having some consequences, legal consequences at that point seem reasonable. And I think if it's done in a way that isn't overly complicated but just simple and straightforward that says that dissemination or use of that information in any manner that harms the individual list those issues I think is reasonable. I think what I like about that is that it gives us a lot of freedom between where we are today and there.

And I think, you know, my concern is if we try to get too complicated then we are, you know, unfortunately can't see all the consequences and can't see into the future, and I think we may get ourselves in a situation that we regret later. So to me I would try to go to the extreme if you will and just create something simple to that extent.

DR. GUTMANN: Sounds like our principle of regulatory parsimony.

DR. CHAHINE: Yes.

DR. RODRIGUEZ: And I would concur with the fact that simplicity here is probably what would be most beneficial just because it's very hard to parse out exactly what would be appropriate today versus in the future.

And so I think one of the other keys that's identified is in each case whether it's the private system that Ken was talking about or a public database for research purposes there has been some level of voluntary willingness to contribute the DNA, the

genomic information to whatever cause, whatever purpose you as an individual are seeking to gain the benefit of. And so respecting that I think would be fundamental.

And then there may be some, you know, few cases that could be identified where there could be harm, discriminatory practices. We have, you know, the Genetic Nondiscrimination Act that's already in place for some situations but not all situations. And so I think it really does come down to protecting who can have access under what circumstances that largely involve individual willingness and then also general good and not harm in terms of the use.

DR. WAGNER: Actually I've got John and Nita next but I've got to follow on this one because it was actually the question I had in my mind. To what degree can we really separate data use and data access? And if we were to as a Commission recommend pretty strong controls on data access for example does that leave enough room to do good science out of data use?

For example, I can ask Google how many people are pinging a website, or I can ask someone how many people are pinging a website and that's a question about -- that's a use question as opposed to asking them I would like a list of all of your people to determine how many are pinging a website.

Can we in fact control access to genomic data, still do good science because we can be more prescriptive about the use of those data? Can those things be separated?

DR. RODRIGUEZ: I'm not sure I would go so far as to say exactly how they can be separated. In my world they're very much tied together. And I think you would

want to keep them linked so that access is only appropriate if you have a use that's consistent with what again the individual has provided for. So for research purposes --

DR. WAGNER: I'm actually asking the reverse. Is there an expectation that we must give access in order to do good science, in order to have scientific use.

DR. RODRIGUEZ: So in that regard I would say that it comes down to if the access is not provided in appropriate ways it constrains the questions that can be asked.

DR. ARRAS: I need to be educated about the certificates of confidentiality. I understand that -- I mean, I think it's admirable that you are as a matter of policy requesting this from all users. But I also understand that a very small proportion of users out there actually offer such certificates in the wild as it were.

So, you know, what's the coverage of a certificate of confidentiality? How large and fine-grained are its teeth? And what sort of protection do they offer with regard to inquiries from, say, law enforcement or Homeland Security?

DR. RODRIGUEZ: Okay. So certificates of confidentiality, I'll just start with the fact that they're not perfect and acknowledge that up front. I think it's an existing mechanism that we have and the coverage to speak to your question is by study.

At this point the way that it has been implemented under the Secretary's authority is study by study which is why we can advise individual investigators that are collecting genomic information with identifiers to consider the appropriateness of requesting a certificate but we can't automatically at this point in time issue it for genomic information.

The extent of the coverage, there are varying reports as to whether or not

they work, whether or not the courts will honor them. It's a protection against compelled disclosure too so that's very important because anyone can volunteer it. And so when the investigator is providing a certificate he's saying and promising that he won't do it under compelled circumstances except for certain reasons of public health that are spelled out. But if the individual notes to someone that they are in the study or the investigator might decide or an institution might decide. So it is not failsafe.

And again, it is another area where some stronger protection could be helpful to engender better or more public trust. And we looked at those issues and tried to think what we could do and the mechanism that we have used was the extent that we could go under existing policy in terms of keeping it by a project-by-project basis, seeking one for our project in dbGaP.

DR. ARRAS: Quick follow-up. So, what amount of extra additional protection does a certificate of confidentiality provide over and above the boilerplate confidentiality information in an informed consent form?

DR. RODRIGUEZ: So it is designed to provide a level of protection, a grant of protection from subpoenas so that it can be used in court. The concept is that it can be used in court so that an institution can say I will not provide under the protection of this grant from the authority of the Secretary of Health and Human Services the identifiers that I may possess for this study because I have been granted a certificate of confidentiality.

DR. FARAHANY: So as you can tell we've been thinking about this use versus access issue a good bit, and I think both of your comments today were really helpful in articulating that. And I want to follow up a bit on Amy's question about thinking about

the different types of uses that we might need to restrict if we were to ensure the trust of individuals. So trust, whether it's to participate in studies, trust to use commercial services.

And my intuition hearing the conversation is that restriction on access is only needed if we don't have restrictions on use, right? The reason you would restrict somebody's access to it is because you don't have adequate protections on the other end against potential nefarious uses. And if you had adequate protections for potential nefarious uses then you wouldn't need to restrict access unless there's some independent reason beyond that that we think individuals' trust would be compromised.

And so thinking about some of the areas in which you might want to have restrictions on use, discrimination. So beyond GINA likely, beyond employment and insurance for other purposes as well, including probably those that Amy mentioned. Or like a number of states have done, prohibiting the impermissible sequencing of information that isn't your own notwithstanding a separate conversation about property and who owns it. And so that would prevent the average person who picks it up on the street from being able to actually sequence the information.

Are there other ones beyond discrimination or somebody who picks up the coffee cup being unable to use it that would be needed, and do you agree with my intuition that if you had adequate protection on use that you wouldn't need to restrict the flow of access of the information?

DR. GUTMANN: I just want to say the second qualification that Nita made is very important, that someone can't sequence someone else's genome. That inherently recognizes a kind of default privacy interest in that even though it's out there just

picking it up and sending it to your lab under false pretenses is considered wrong. I mean, that's just a basis.

DR. FARAHANY: Yes. Well I mean, so there's a lot of reasons why you might have that. It could be discrimination. It could be property and there's a separate kind of view of property than the one you advance, but yes.

DR. GUTMANN: It could be. I think there are lots -- I think this is an important point. I just think we're all -- there are a lot of reasons that converge there. But please answer -- that second qualification packs in a default concern about the identification of individuals with their genetic material.

DR. FARAHANY: And about 25 states have some sort of restriction already on that which is the individual picking up the coffee cup and being able to actually sequence the information if it isn't their own.

DR. CHAHINE: So, I think the short answer is for me I think it goes a long way and maybe the answer's just yes. I do think that if we can sort of block that end, you know, nefarious use I think it goes a long way.

My thing -- what's interesting from a commercial standpoint is we -- a lot of customers get data from either us or from other services. And what I see that's a little disturbing sometimes is that they upload their data to sites that quite frankly I can't even have -- I try to research who these individuals are that are doing additional research on people's data. And I'm not even sure who they are or how they're qualified and if the data that they're getting back are even, you know, valid, right? So the point is that I think that we are moving in a direction where consumers feel comfortable, rightly or through ignorance,

uploading their data to other sites. So I do think that restricting sort of the end I think is important just to be able to take care of that issue.

But the other one is just from a research standpoint. The reality is the amounts of data that we're looking to get are -- I think it's unreasonable to think that our current research institutions or anyone really could do all of the research that could possibly be done. And there's research that are important to certain subsets of society that are not necessarily going to be of interest to researchers. And I think people are going to start taking it upon themselves to start doing research. And you're seeing that through advocacy groups and things like that.

And so I think being able to -- the more access we can give and feel comfortable I think we would certainly be moving our mission forward of extracting as much information from this genetic information as possible.

DR. HAUSER: I just had a short question for Dr. Chahine. In your privacy statement Ancestry DNA speaks about aggregating individuals for targeted advertising displays. And I just wondered how that is done in the field in general and does this include genetic or only demographic information?

DR. CHAHINE: So maybe I'm not -- when you say aggregated information, you're talking about genetic?

DR. HAUSER: Bundle individuals --

DR. CHAHINE: Information.

DR. HAUSER: -- into groups for targeted ad displays.

DR. CHAHINE: In terms of the Ancestry DNA we don't have anything on

displaying the information. What we do is we talk about aggregating the data to potentially improve the population genetics through both genetics and pedigree information. But in terms of advertising, is that what you're --

DR. HAUSER: Yes.

DR. CHAHINE: -- you're getting at? We -- on the genetics side we do not, we don't have any advertisement whatsoever in terms of that data.

DR. GRADY: Laura, I wanted to thank you both first of all. I wanted to follow up with a question that builds a little bit on what Nita was talking about and ask about the public access versus controlled access data from dbGaP. Do we have any information about how often publicly accessible data through dbGaP has been useful for scientific purposes? And do we have any idea who uses it?

DR. RODRIGUEZ: So, on the open access side we can't track who is using it. So we have a list of IP addresses but we don't know what they're looking at. We also therefore then can't track what's coming out of it.

At this point for dbGaP purposes the information that is there publicly from a data perspective are aggregate-level information on the clinical side or the phenotypic side, but all of the genomic data has been moved behind controlled access now. And so there's some information that's of sort of very superficial value to the scientific community. We look at goals around allele frequencies so that they can do some quick comparisons but it is something that we hear from the community that some of the basic descriptive information from aggregate genomic data is not available publicly that could be very helpful. And they actually have to apply for the individual-level data to get even just

aggregate, even if that's only what they have asked.

We have recent -- just this week have put out a pooled set of aggregate data so that through one request it's still behind controlled access, but it's one request so that scientists can come in and have access to 14 general research use studies so that those that are just looking for large volumes of aggregate data or to cross-check what they're saying in their study with what has been seen in many other populations, they can do that more simply and not have to go through the individual request process or gain access to data they don't need which is an important policy concern.

And so what I can also say is that for other data that's not part of dbGaP but for things like the HapMap project or 1000 Genomes that NIH also runs where there was very broad consent for future use and for open internet-accessible data release. Those, the numbers that we have on that in terms of people accessing it far exceed the numbers of people accessing dbGaP.

So we do know that disciplines beyond biology or other things will use the large volumes of data accessible through those open access portals and we are restricting by virtue of the controls we've put in place for controlled access who is coming to the data. And so that's again a pressure point that we seek to balance in trying to work through the policy.

And I guess I would add in response to Nita's question too about the access. I think having broad controls or protections in place for appropriate use would indeed be outstanding from my perspective to have and engender a great deal of public trust. But I think you would still need to contend with as a society the wide variety of individual

and cultural preferences around use and access to the data that couldn't be encompassed and just solved through some higher-level legislative protections or other things. So, those would be important, it would be important to have teeth to those bands but I don't think they will cover the waterfront of questions and issues from the ethics perspective of how the data are used.

DR. WAGNER: Thank you. And we have finally from the audience Dr. Edward Gabriele who's a senior healthcare ethicist in Navy medicine. In addition to reminding us that informed consent is in fact a process of human trust he's asking a very practical question. How confident are we given the current state of DNA science and where we imagine that it's going to go that de-identification is absolute?

DR. RODRIGUEZ: So I'm going to speak for myself right now and say that I don't think that de-identification is absolute. I think even in the 5 years that we have had dbGaP going and the policy in place it has shifted dramatically. Because with the increased power of the statistics that we make possible by large data sets it changes what is possible in terms of identifying unique patterns.

And then again with the amount of information outside of our resource that is available publicly direct-to-consumer access to technology to create sequence information, matching becomes an increasingly reasonable thing that can happen.

DR. CHAHINE: And I would absolutely agree. There's no question that it's going -- I think we should operate under the assumption that if it's hard to re-identify today it won't be tomorrow or sometime in the future and we should always keep that in mind.

DR. WAGNER: We will be taking a very brief break and trying to reconvene as close to 11 o'clock as possible. But before we do, Dr. Rodriguez, Dr. Chahine, thank you both for your contributions.

(Applause)

(Whereupon, the above-entitled matter went off the record at 10:53 a.m. and resumed at 11:06 a.m.)