Policy Perspectives from Members of the Bioethics Community
Henry T. Greely, J.D.

DR. WILLARD: Our final session will involve three bioethicists, and we're fortunate to have some real experts in the field to discuss a variety of perspectives on key policy issues involved with contemplating a large population study.

Could someone close the far door? That's a subtle hint to people that they should be inside and not outside.

In addition, we've asked the panelists to address or identify specific mechanisms or processes to address the issues that they raise. I don't know if you have time in your comments to address that. If not, we surely will follow up with questions specifically on that.

So our first speaker will be Henry Greely, who is the Dean and Kate Edelman Johnson Professor of Law at Stanford University. He specializes in the legal and social issues arising from advances in the biological sciences and in health law and policy.

Mr. Greely, thank you for being here.

MR. GREELY: Well, thank you. I'm happy to be here. I appreciate the invitation. Despite the PowerPoint, I am not my friend Pilar Ossorio, and in fact I'm not using PowerPoint at all. I'm not sure whether this is an eccentric affectation or laziness or a desire to actually have you watch me and not watch the PowerPoint slide. I like to think that it's because as a lawyer I very rarely have data, and if you don't have data, there's much less power to using PowerPoint.

What I'd like to do is talk about two big ethical issues in large population resources. Before doing that, I'd like to say that I think the overall issues about whether this project should go forward are quite fascinating, and I'd like to pick up on something that one of the members of the last panel said. I think having a big goal, having an audacious idea, is a really important thing. I think it has significant externalities not just in terms of public relations. It's not just Bono and people not going to sleep, but it inspires researchers, it inspires students, it inspires people to go into the science. I think moon shots have externalities that are sometimes overlooked. I say that without having any opinion on whether this particular moon shot has a scientific value to justify its financial cost. I'm not competent to answer that and I just don't know the answer to it.

I think I am competent, though, to say that if this does go forward, it will face a host of ethical, legal, social and, perhaps most difficult and used in a broad sense, political problems, and I say that as a battle-scarred veteran of about a decade of the Human Genome Diversity Project, which met many of these same difficulties and ultimately failed to surmount them. So I do think that if one decides to go forward with this, a careful study of similar past projects, successful ones and failed ones, will be very useful in letting you know not just what methods may or may not solve some of these problems but what problems you're going to hit, because the one thing that was overwhelmingly clear to us in the HGDP was that there were far more land mines in that project than we had any idea about going into it. We discovered a few of them, to the cost of many body parts. There will be more that a project like this will hit. But looking at the land mines that have been exploded in the past or that have been diffused in the past will be very helpful if this goes forward.

Now, I was asked to talk about or to specify three particular ethical issues that I thought were especially important. The three that came to mind were, first, issues of control of the uses of

these materials and data; secondly, issues of the return of information to the participants in the research; and third, issues of confidentiality. The issues of return of information to the participants in the research my colleague Pilar is going to speak about next, so I won't say anything more about that other than to say it's really important, and listen very carefully to what she has to say because I think this may be one of the most dangerous of the land mines a project like this will face. Instead I'll focus entirely on the issues of control and the issues of confidentiality, starting with control.

By control, what I'm talking about is the research participants' ability to control how the data and the personal materials, the personal biological materials that person has given to the project end up getting used. I want to start this discussion with a story about some litigation that's currently in progress in the state courts in Arizona involving members of and the Nation of the Havasupai, a federally-recognized Native American tribal government, nation, that lives in the lower Grand Canyon.

It started in 1989. Researchers from Arizona State University started a genetic research project with the Havasupai aimed, according to the allegations of the complaint, solely at a study of non-insulin dependent diabetes mellitus, an issue of great interest to many Native American groups, and certainly to the Havasupai as well. The facts that I'm going to tell you about are not yet proven facts. They're allegations in a complaint, and as a former litigator I know exactly how much suspicion one should view unproven allegations with. In this case it's a little bit modified because most of the allegations of the complaint are taken from a report written at the pay and at the request of the defendant, Arizona State, who had an independent investigation of the situation done, and that entire report is attached to the complaint.

So in 1989 the diabetes study started, bloods were taken, family histories were taken, clinical information was taken from the Havasupai, and only over a decade later did they learn that the researchers involved were not just studying diabetes among them but studying schizophrenia and also studying issues of historical origin. The Havasupai were outraged. They were particularly outraged since they had been specifically reassured that only diabetes research was going to be done and had specifically made that a condition of their initial approval of the research.

They further were outraged when they discovered that the samples weren't just sitting at Arizona State with the researcher who had come down to Supai Village and talked to them and met them and who they knew, but the samples were distributed all over the country and all over the world to researchers they had never met, had no relationship with, and had no appropriately or misplaced sense of trust in. The result has been litigation, bad feelings, and I suspect a very long time before the Havasupai are willing to participate in genetics research again no matter how the litigation comes out.

This is a particularly powerful example, I think, of the fact that people's interests in the research that is done with their material and their data are not limited to things that affect their physical health, their personal economic well-being, their insurance status, but people sometimes care about what you do with their data because they don't want to be complicit in certain sorts of research. The Havasupai did not want to be studied for schizophrenia. They felt that it was going to be a stigmatizing study no matter how it came out, and it wasn't an issue they wanted examined. Similarly, they did not give permission to and did not want to be part of a study of the history of their population because they believe they know where their population came from and had no interest in abetting other theories, including that sometimes referred to as the "BS hypothesis," the Bering Straits hypothesis about where their population came from.

Personally, I would be outraged if I discovered that material I had given for one research project was used, let's say, by English researchers to study the intelligence and genetics of the Irish, since I have a fairly good idea how that might come out in the hands of English researchers. But more generally, any sort of research into race and intelligence using material that I had given I would feel is a betrayal and had made me complicit in research that I did not want to take part in. Now, those are my sensitivities. Other people will have other sensitivities. They may not want their data or information about their family members to be used in research in mental illness, research in sexual orientation and genetics, research on alcoholism or addictive personalities, or a variety of other things.

But people will feel betrayed if their research is used for purposes that they think are bad purposes without their knowledge or consent.

Similarly, although I think this is a lesser issue but not a trivial one, people often, at least traditionally, take part in research in part because they trust the researchers who come to them. They trust that Dr. Collins is really going to look after their interests and be interested in cystic fibrosis. They've met him, they've shaken his hand, they've looked him in the eye, they trust him. I think very few subjects, very few research participants or research partners have any idea, regardless of what the informed consent form says, of how broadly their samples and data might get distributed by people whose eyes they haven't looked into and for whom they do not have that level of trust.

Now, in the context of large resources, the creation of libraries, resources like this would be, what I once tried unsuccessfully to get termed genotype/phenotype resources, this produces a real dilemma because you can't successfully ask people about each and every research project that happens throughout the history of the databank or throughout the history of the resource and give them full informed consent, get their full informed consent about each and every use. It seems highly impractical because there will be hundreds of uses, if not thousands, spread over time. And even if you were able to have the budget to go back and individually reconsent people on each one of those, I think you'd quickly find that people were sick and tired of seeing you and didn't want to be reconsented after a while on each additional molecule involved in pancreatic cancer or involved in asthma.

On the other hand, I do think that there's something phony about the idea of informed consent for these kinds of resources. The idea of informed consent, both in general and as laid out in the common rule, is consent in which the research participants -- I'm trying to avoid the word "subjects," which I agree is a bad word -- the research participants are informed about the specific risks and benefits of the particular research that's going to be done with them or with their data. How can you do that with a resource like this? No one has any idea what specific research will be done, what particular diseases or genes or environmental effects will be examined. The whole idea of the resource is to make it available for people to do everything that seems important and useful over time.

So even calling it informed consent I think is a misnomer. I'm not going to take the position that as a result none of this should be allowed by IRBs under the common rule, but it is a real problem with the issues of consent around participation in resources like this. The consent is not truly informed, cannot be truly informed, and yet on the other hand, for practical reasons, when we know enough about the specific projects, it really doesn't make sense for us to be able to go back and reconsent everybody on every specific detail. It's a dilemma.

Possible solutions? Well, the first thing to say about any possible solutions is they're certainly not perfect. This is a real dilemma, a real problem, and there are no perfect solutions to it. But there are ways, I think, where some of these can be mitigated, involving two steps. First, at the beginning of the process, try to find out if there are specific issues that the research participant does not want his material or her data used for. You could do that in an open-ended affirmative way: "Is there any research you don't want done with your material?" I think that's unlikely to be very successful or very realistic.

One might also imagine an opt-in, give somebody a 12-page list of different research topics and ask them to check all the ones they're interested in. That doesn't seem very meaningful to me either.

A shorter, more targeted opt-out might actually be meaningful, listing things that you have some reason to believe might be sensitive, might be issues that some of your research participants might not want their materials used for, and ask them to check yes or no in advance.

Even that I think is only a moderate step in the direction of protecting these interests of people who may not even know what interests -- know that they've got interests in these issues until something comes up.

Another alternative and one that a Veterans Department project that I've been involved in has endorsed is to have continual monitoring of the subjects of the research topics that are involved either, or I think better, both by an IRB and by a group drawn from the research participants themselves, and have them discuss the new protocols that are proposed and see if they think there's anything here that's particularly sensitive. If they think there is something that a significant number -- weasel word; what percentage is a significant number; how do they know if it's a significant number. But if there's something they think a significant number of the research participants might object to, then they'd have the power to require individual reconsent.

Now that I strongly suspect would happen in a very, very small number of these projects. I suspect that no one in the country is going to be particularly personally involved or emotionally attached to issues of pancreatitis, or issues of asthma, the pharmacogenomics of different asthma drugs. But when we get into behavioral genetics issues, I think then the likelihood is much greater, and the alternative to doing something like this is to have a situation where you've got a research participant who, years after signing up for this good, noble thing, discovers that his DNA or his family history or his health records were used for something that he finds abhorrent, in which case I put it to you that he feels cheated, betrayed, unhappy, and he has some grounds to do that, some appropriate grounds for that.

Now, if you take my position that some sort of control mechanism is appropriate, it does rule out one alternative. It rules out the alternative Dr. Collins mentioned a moment ago. I'm not sure this was his full plan, to make the material open to anybody who has IRB approval, or putting it even more broadly on the Web. You would need some sort of check, at least a listing to make sure that data and DNA, data and materials from people who had said they didn't want to be involved in this particular research topic wasn't involved in that topic, and I think you should also, for sensitive issues, put it before a participants board, as well as an IRB. So it rules out one alternative.

If that alternative is really important, if that's what you need to do to make this successful or to make it as successful as you hope that it will be, so be it. But make sure that the informed consent for it warns people up front that you have no control over what your data and your

materials are used for. They may be used for things that you disagree with, and if that happens and you find out about it, don't complain to us. Put that in English, not in informed consent-ese, and don't hide it at the back of a 20-page consent form.

Second issue, confidentiality. This is another issue where there is an enormous problem. Americans are enamored of privacy, enamored especially of health privacy, and it's confronting an issue of which this is just one small part, an economic and technical reality that is, in the words of one Silicon Valley mogul, "Privacy is dead. Get over it." The push which I think is inevitable for more computerization of data, inevitable I think because of all the advantages that come from that computerization and networking and access of data, invariably undercuts the possibility of promising people complete or even very full confidentiality.

Now, in terms of confidentiality, most research goes forward in sort of a key system where the specific researchers may not know the identity, somebody somewhere knows the identity but it's hidden behind a code. There is a key holder someplace that has, of course, possibilities of abuse if the key holder somehow cheats and decides to use this information for bad purposes. Personally, I think the odds of that are extraordinarily low and can be made lower with appropriate sanctions, but they cannot be taken to zero. People who take part in this research cannot be promised confidentiality. They can be promised the best confidentiality we can offer them.

Despite what I've heard occasionally from the computer folks, it doesn't look like there's a technical fix for this. I've heard a lot about one-way encryption or hashing encryption followed by a lot of movement of hands as it comes to be explained, and as far as I can tell from cross-examining computer scientists in some of my classes, it's not going to be a useful technique particularly for a project like this where additional data will be added longitudinally. One-way encryption works all right if you're only putting data in once and there's no way to ever decrypt who that is. But if you've got data from me and you've put it into the database, and later you want to add more data from me, there's got to be a key somewhere. Somebody, somehow, has to know that file 17648G is Hank Greely. Once you've got that, this one-way encryption idea no longer will provide the technical fix that people hope for.

But there's a more fundamental confidentiality problem. Useful data sets are rich data sets. Rich data sets are identifying data sets. Professor Latanya Sweeney has published some nice work on this. I was born on June 25, 1952 in Columbus, Ohio. I actually should go back to Franklin County and look sometime, but my guess is, given the demographics of the era, there were probably seven kids born that day, of whom I'm guessing there were four males, three white, one black, three females. If you know that information about me, that I was born then and there, and you know my sex and my race, you're down to three people in the world. If you're really interested, you can find out which one of them is me. I happen to know that one of them is my cousin Mike. We were born on the same day, the first grandchildren of our grandparents. He will never let me forget that I'm 20 hours older. He's 5'6", 130 pounds. It wouldn't take much more data to distinguish which one of us was me and which one of us was him.

If you're born in a small town, the identifiability becomes even easier. If you're a famous person, your identifiability becomes even easier. If you live at the zip code for the White House, it's not going to be all that hard to identify you with just a little bit of data.

Now, as the world becomes more and more wired, this becomes a bigger problem, because more of this data is put online. Interestingly, I think the thing you've got to worry about for a lot of this place and birth date data is the genealogists, who are busy as beavers online, putting all sorts of

databases online. Genealogy is apparently second only to sex or pornography in terms of its interest level on the Web, and genealogists are constantly putting new data sets online. So right now you'd have to go to the Franklin County records to look up my birthday, but that's probably not going to be true for very much longer, and once that becomes possible, the ability to identify people with deidentified data sets becomes much stronger.

Now, you can fuzz the data sets. You can say not born on June 25, 1952 but born in 1952. You can say not born in Columbus, Ohio but born in the Midwest. Every time you do that, you lose something of potential scientific value. The real harm there, the real problem is you don't know how much value you're necessarily losing. There are seasonal variations in disease incidence based on what season somebody is born in. There are some things like schizophrenia which have a higher or lower rate depending on what season you're born in. There are regional variations and issues. I went to my ophthalmologist, who looked at my retinas, a normal exam, and said were you born in the Ohio River Valley? I said, well, close. He said, well, you've got histoplasmosis scarring on your retina, which is very common in people from the Ohio River Valley.

Now, you can try fuzzing the data. As you fuzz the data, you lose some medical and scientific value, and you don't know how much you lose.

So the dilemma here is, the more useful you make the data, either in terms of the completeness of the data set or the wide breadth of people who are able to get it, the less you can promise people confidentiality in it, and even anonymity. Even if you try to make it completely anonymous, there is no key anywhere. You can't successfully do it.

Is there a solution? Not much of one. The only solution that I can recommend is complete and total honesty, but that will be expensive. People are leery enough about their privacy that if you tell them we can't promise you confidentiality, and even if it's anonymous, somebody who cared enough might be able to look at all this data and figure out who you are; we don't think anybody is likely to do that, we think the odds are low, but in good conscience we have to mention to you that that's a possibility, you will lose some research subjects I predict, and probably not a trivial number of them.

The alternative, though, is to not tell them that, let them sign up based on their understanding that there's broad confidentiality protections, and then feel betrayed when they discover that their identity has somehow been blown and that their confidentiality is not there. As I say, I think this is a much bigger problem than just a problem for large population research resources. It's a problem all of American health care has to deal with that stems from a mismatch between our public expectations of privacy and the realities of the society we live in with respect to privacy.

Well, there are a number of other important issues, but I suspect I've already gone over my time. Let me just close, though, by saying I think this is really important. I think it's really important to line up the ethics of projects like this so that people do not feel betrayed, do not feel that they've been lied to or mistreated, and I think it's important for two reasons. One is because it's the right thing to do. If you've got subjects who feel that you have mistreated them, lied to them, deceived them, betrayed them, then at the very least you've probably done something wrong. You may not have been evil, but at the very least you didn't communicate as well as you could have, and that's an unethical result.

Secondly and more pragmatically, it's bad for science. Any research subject who feels betrayed and mistreated is not a research subject who is likely to sign up for more research. The Havasupai aren't likely to do a lot more research anytime soon. They're also not research subjects

who are likely in their role as citizens to lobby for, vote for or support biomedical research. Treating research subjects well is ethically important for science and for scientists. Treating research subjects well is politically and pragmatically important for science and scientists as well.

Thank you.

DR. WILLARD: Thank you very much, Mr. Greely. I appreciate that, and we'll hear from you again once the full panel has spoken.