Feinstein Kean Healthcare
caBIG® Bio IT World Presentations
Lecture by Philip Payne
Wed 11:30 am.
Waterfront 2 Track 2

Philip Payne: All right. Can everyone hear me okay? All right. So thank you for that introduction. I will admit this is in fact my first Bio IT World that I've ever attended so always a pleasure to meet a new community of people innovating in the health and bioinformatics spaces. What I'm going to attempt to cover in the thirty minutes or hopefully less given our current scheduling is a very broad topic of: how do we go about increasing our ability to conduct impactful translational science in order to advice biology and healthcare, and what is the role of biomedical informatics, in particular, integrative platforms in achieving that sort of aim?

I always like to give people a roadmap of what I'm going to be talking about so broadly I'm going to talk about three really closely interwoven and motivating factors that serve to provide context for this talk. First I'll give you just a very brief, and what will sound like marketing speak but is just to give you an idea of where I'm coming from, description of the living laboratory that we have at the Ohio State University Medical Center. Then I'll focus more on this concept of P4, or personalized medicine, in the role of evidence generation that brings together biology and healthcare delivery, and the role of translational science in bringing those two disciplines together to affect positive change in healthcare delivery and outcomes. Then I'll talk more about our efforts to create what we call a learning healthcare system, a popular term right now both conceptually, and then how in particular we've been taking a foundational stack of technologies from the caBIG® program to move forward and realize this aim. Then finally I'll round out with a few lessons learned and next steps in terms of where we're going.

I always like to give people a sense of how I came to stand up here behind a lectern and talk about these sorts of topics, and I would argue that how I've arrived here is not a lot different than a lot of other people that work, in particular, in the academic side of the biomedical informatics community in that a lot of us sort of, for lack of a better term, stumble into it by accident. So in my case, I began doing very prototypical software development work in a biomedical setting at UC San Diego working in ophthalmology research and, based upon that experience, moved on and pursued a masters in informatics at Columbia really, again, by accident. I really didn't know what the term was until the last minute; I applied to the program late, showed up and thought -- this could be cool, and ended up finding a new career path. When I completed that masters level training I went back to UC San Diego, worked in the cancer program there and then got talked into returning back to New York in pursuing a Ph.D., the alphabet soup argument, and then move to Columbus, Ohio where I've been at since 2006 where I began as an Assistant Professor and now I chair the Department of Biomedical Informatics. My particular interest areas really lie in these intersecting domains of knowledge discovery, multisite electronic data interchange platforms, workflow monitoring and human factors, which I think a lot of us realize are very tightly related to one another when we're trying to affect these sorts of large scale systems level efforts to deliver personalized healthcare and perform translational research, etc.

With that you know my viewpoint of the world; let's talk about this living laboratory concept. So for those of you that are not familiar, other than perhaps the obnoxious use of the word 'The' in front of Ohio State whenever we say the name of our institution; if you're

interested after this session, there's a long history as to why that's done. But nonetheless, it's a relatively large academic health center comprised of multiple constituent hospitals including a very large cancer and heart hospital. As you can see from the pertinent statistics, we have a very high number of patient encounters and we've regularly been ranked in all of these sort of prototypical, sort of, points of pride venues, including U.S. News and World Reports, the most non-scientific ranking system in the world that we all live and die by in the healthcare industry. We are currently in the midst of a billion dollar expansion program, including the development of this finely illustrated building here which is a million square foot facility which brings together patient care, education and research in every single floor. What that, in fact, means is that in this new hospital building, I think it's particularly interesting to this crowd that, on every floor where there are patient care rooms, there are also wet and dry lab space for research where we can put things like next generation sequencing devices or at that point by the time it's finished in 2015, next, next, next generation sequencing devices. The idea is really to break down those artificial barriers between what we do in the research lab and what we do in terms of patient care. I think it's a very exciting time for us and of course we're the home of both a NCI-Designated Comprehensive Cancer Center, the only such program right now in the state of Ohio, as well as a CTSA funded program that seeks to establish a professional home for clinical and translational science. Enough marketing speak.

One of the things that we've done at Ohio State in order to think about how we bring together research, education and clinical care is really provide a common theme by which we can link these areas together, and that is delivering personalized medicine. And we think about this as much more than just an advertising tagline. We think about: how are we going to transform the way we interact with our patients, accelerate the process by which discoveries in our basic science laboratories move into clinical research and new evidence based practice guidelines, and how do we, hopefully, although it's not necessarily in our financial best interest, keep the patients out of our facilities. How do we actually promote wellness as opposed to acute care episodes? And we do so through this concept which you may be familiar with of focusing on bringing together research, education and care around this P4 medicine paradigm in which medicine is predictive; for example, where we can use biomarker technologies to predict the risk of disease. It's preventative in that we can use these risk profiles to actually plan preventative care. It's participatory in that we use tools like personalized health records. We are also looking at the use of consumer genomics technologies in order to involve our patients directly in healthcare and, finally, it's personalized in that we actually design and deliver adaptive therapies so we don't just give you the same thing that the last dozen patients that came through the door had in terms of treatment for the disease. We give you the treatment that we think is optimally designed to give you the best possible outcome. And I will argue throughout this talk that informatics and this concept of P4 medicine are very, very complementary, and absolutely necessary to coexist, and that in order to deliver personalized healthcare, we actually have to have the evidence that defines how we deliver those four p's. So the idea here is that research, as well as the delivery and observation of clinical care, and then reasoning over the findings that we find in both of those places in the integrated matter is essential to developing that sort of body of well-validated knowledge and evidence to drive P4 medicine.

But there are a number of challenges that exist in order to realize that sort of goal which include really at a high level: how do we capture, represent and manage all this high throughput multidimensional data, and I will say, and I believe in this very strongly; it's not just about bimolecular data. We are now facing the same sort of challenges when we talk about clinical

phenotype data as a result of the explosion in the use of electronic health record and clinical trial management platforms at almost every single care setting. So now we have sort of a confluence of those factors and even if we can capture and manage that data -- how do we even reason upon it? How do we know that we're asking and answering the best questions possible to generate evidence, and how do we then take that evidence and deliver it in terms of decision support at the point of care that's useful? I would point out, for example, right now we have a pilot project where we're looking at using a number of direct to consumer sequencing technologies to look at the genomic profile of our patients, and then take that data and integrate it with our electronic health record so we can deliver genomic decision support at the point of care. For example, Warfarin, a common anticoagulant therapy -- we actually know that there are a number of key markers that we can use at a bimolecular level to predict how patients will metabolize the drug, and if we look at those markers in advance, we can plan their treatment with Warfarin and prevent a number of very undesirable adverse outcomes with the disease. However, being able to sequence the necessary markers, deliver that in the electronic health record and then deliver that alert at the point of care so that we are actually making that decision, is incredibly difficult. Like many institutions, we have one of the most common electronic health record platforms in the country right now, Epic, and there really is no capacity to take that sort of genomic data and imbed it as a type of data point that we actually include with a patient encounter or a patient's longitudinal health record. So we have lot of work to do to even be able to connect the dots between them. Then finally, in order to keep on finding these sorts of evidence based practices that deliver or support personalized healthcare -- how do we rapidly execute research?

I don't think anyone here in the room hasn't seen, much like the prior speaker mentioned in the joint session, all the scary graphs that talk about how long it takes us to get something from idea to clinical study to delivery at the point of care. If you don't know all the scary numbers it's around seventeen or eighteen years, which I don't think anyone would be particularly happy about when you think about that. So given this concept that we have this living laboratory at Ohio State and that we're very focused on personalized or P4 medicine and delivering the evidence necessary to make that possible. I've talked to you about this concept of translational science in terms of connecting those dots -- how do we actually do that? What's the sort of technical infrastructure, the workflows, the processes that are necessary to realize that sort of approach to basic science research, clinical research, and clinical care?

I'm going to talk about what we're doing leveraging our living laboratory in exactly this capacity to create a learning healthcare system. So the foundational premise for this is that we want to learn from every single patient encounter so we can improve care. So can you imagine if we begin with the patient encounter at the top of this cycle, we want to instrument that encounter so that we can get data from it that's useful for us for these sorts of purposes, as well as tissue, right? Because we're not going to have all the data we need up front just based upon what we do in the clinic. And then, using that data in potentially test or assays we can apply that tissue, we want to generate hypotheses, we want to verify and validate those hypotheses, formalize evidence based upon the research that we're conducting there, apply that evidence within the clinical environment, improve patient care, and that's both quality and outcomes and that definitely includes costs as much as it does the clinical outcome because these are closely related to one another, and then deliver that all back to additional patient encounters that we instrument to understand the impact of these interventions.

For those of you that don't work in the biomedical informatics community as we sort of think about it in the academic side of the house, the right hand side of this model we would refer

to as the conduct of translational bioinformatics and clinical research informatics. That's where informatics comes to pass in this cycle and, similarly, on the other side of the cycle, that's really where clinical informatics and public health informatics come into play, and there's a very robust body of well-validated theoretical and methodologic frameworks that give us the tools that we need to realize if we can combine them in the right way.

At the end of the day, and to recapitulate, what we're trying to do here is learn from every single patient encounter so we can improve their care, their family's care and the care of their community. This is not about making our patients guinea pigs. This is one of the big pushbacks about making research an integral part of the clinical care process. There's a reason why we call it practicing medicine and not performing medicine. And the reality is it is a constant process of improvement and the patients need to be part of that process of improvement because it has direct impact, not only on them, but their families and their communities. This is a very much a sort of mutually beneficial cycle. So how do we do this? There are a lot of potential data and knowledge resources that we can leverage from the clinical enterprise, like our electronic health records; the research enterprise, such as biospecimen management systems; any number of databases that we develop for various research programs; clinical trials management systems; the educational enterprise often has a lot to deliver here, including: synthesized data sets that we use for teaching purposes, as well as any number of knowledge bases, and then there are a tremendous number of external resources including: public databases, literature repositories that we can mine using text mining and natural language processing tools, and any number of other knowledge collection, such as ontologies and terminologies that contain a high amount of conceptual knowledge.

However, in order to realize that learning healthcare model and to leverage all those data and knowledge resources, we need to overcome, as I had indicated earlier, a number of key informatics challenges, including our ability to integrate all these resources, to reason upon them and then apply them at the point of care or in the community at large. I would argue that all three of these categories: integration, reasoning and application, are open areas of research, right? We do not have, sort of, cookbook solutions for how to do this particularly well. We have limited examples from pilots at any number of high profile institutions that show that it is feasible, oftentimes. We have not shown how we scale this at a high level if we want to do this, at say, the community, regional or national level and I think that's both a tremendous challenge and a very exciting opportunity for us, especially when we get beyond hypothetical or theoretical research and get more into the applied science end of the spectrum.

I would argue, at the end of the day, that what we're trying to do with this model and, in terms of creating the learning healthcare system, is to develop these bodies of translational biomedical knowledge that bring together patient population phenotypes, domain knowledge, any number of bimolecular markers, and then make sense of them using biological models, technologies and algorithms. At the core of this is the use of informatics as sort of this integration engine. Right? So that's where the basic and applied informatics science comes to bear.

Recently we published upon, and I think this is broadly generalizable, a design pattern that can be used for the development of computational systems intending to address this concept of developing translational biomedical knowledge in support of personalized healthcare in a learning healthcare system. It seems very simple, but you would be amazed if you look at, sort of, the published literature, both in the applied and basic sciences how often we sort of miss the mark in terms of bringing together all the known best practices from the computational and

information sciences, the business domain, the social sciences, and informatics in order to realize optimal outcomes with these informatics interventions. At the end of the day this is about first and foremost beginning with stakeholder engagement and knowledge acquisition, modeling our available data and knowledge resources, applying any number of algorithms and agents in order to be able to thus reason across those integrated collections, and then analyzing and disseminating the output of those processes. Right? And at each one of these steps there are people, there are workflow and there are technical activities that need to come together in order to realize those aims. I'm going to give you specific example from our efforts to create this learning healthcare system that's based upon that exact design paradigm, which we call the "TRIAD Project" at Ohio State. This is focusing on data federation, shared semantics and supporting the integration of clinical phenotype and biospecimen management data across the entire enterprise in traditional organizational boundaries, and providing knowledge discovery tools so people can ask and answer important questions about those. TRIAD stands for the Translational Research Informatics and Data Management Grid because, if you have ever heard me say this before, no informatics project starts without a good acronym. Everything else falls in line from the acronym. I knew I could get you guys to laugh eventually. So the triad approach at its core is taking a set of established, standards based, open source technologies developed by the caBIG® program and extending and enhancing and adapting them to meet the specific needs of our organization to achieve this goal of creating a learning healthcare system. And for those of you who are not familiar, which would be surprising given this community the idea here is that caGrid really, at it's core, provides us a set of grid based informatics platforms that had been shown in a lot of, I think, compelling pilot projects to have some capabilities that are very analogous to what I just talked about in that ability to link together disparate data and knowledge resources. It's not to say that there's not additional work to be done and you'll hear me say throughout this talk that the idea here is that we're using caBIG® technology as a foundation. We're not just taking a turnkey solution and installing it in our environment, and what we've done, through our CTSA funded Center for Clinical and Translational Science, is enhance and extend caGrid in particular to support this, sort of, domain agnostic data federation of shared semantics across the enterprise. And I sort of affectionately refer to the TRIAD project as caGrid cured of cancer. Because it no longer is a cancer specific platform; it's a platform that is designed to support sort of this more readily adoptable working interoperability where we are able to target specific application domains rapidly.

For those that aren't familiar -- again, the idea here is that we're using this core set of service oriented architectures to affect secure data transfer and integration across a lot of different targeted data resources, and we're doing that based fundamentally on this set of shared data models and dictionaries and their mappings among the various systems so that we can provide these real time query and data integration tools. And the idea here is that we're trying to decrease data redundancy. So data warehouses are great, right? Everybody says, "let's just build a data warehouse and put all this data in there," and there are a lot of places where that is the right solution. But there are an equal number of places where either sociotechnical or policy barriers, or the size or complexity of the data, or the computational cost of doing so, are actually quite a significant barrier to creating these warehouses. And so the reality is we have to deal with a heterogeneous environment where we have a collection of warehouses, project specific databases, tissue management systems, electronic health records, but our end users don't care about that. They just want to ask and answer questions that span all of these resources and I think going back to the introductory remarks, a lot of times we use very low tech solutions even

now that's sort of the equivalent of the sneaker network where people go and query all the source systems and then they mash them together in excel on their local machine or something like that.

There is a definite cost in terms of the replication of data when we start putting it into these new integrated repositories, and so if we can avoid that that would be very positive. There's also a factor here of overcoming those sociotechnical barriers that we can increase local control and accountability, right? So we're able to give people the ability to say that, "I have a local sort of control over data and I can say who or when I'm going to share that data," and that is a big incentive to get people to actually engage in these sorts of activities. And they're a low cost for evolutionary change. If one of these data resources falls off the map or we add a new one, I don't really have to do anything to the rest of my system, right? It is very much an organic system. I would tell people broadly, as an anecdotal comment, at Ohio State, we're a big organization. We have about 16,000 faculty and staff just in our hospital and our medical center, academically. Getting all of those people connected together is very hard. So getting them to agree on things across all those various groups, I think, is a microcosm of what we encounter, at sort of the regional or the national level, to achieve similar kinds of data integration.

So the TRIAD architecture at Ohio State is really quite simple. We have a number of shared tools, including: analytical services that are really these biostatistical and data mining tools, knowledge management platforms, common electronic data capture tools, we are primarily leveraging for clinical phenotype data at a very comprehensive clinical warehouse that large components of have been grid enabled, we, again, provide interfaces to our biostaticstics team, and we have a set of both API's and web portals that allow individual researchers or research groups or large groups or centers to interact with this data in varying levels of complexity. An example case is that we've been linking biospecimens and clinical phenotype data so that we can actually use this residual tissue and characterize it using that patient's longitudinal clinical phenotype to ask and answer important questions. So this schematic basically shows the flow chart, but the idea here is we're consenting patients at the point of care, they're getting normal clinical care, we're capturing tissue from them, we're also getting all of their clinical phenotype data through the electronic health record system, we use a robust system of de-identification services, we manage a residual biospecimens behind the scenes using caTissue, we have a broad variety of data research repositories, and then we use this TRIAD infrastructure to actually query across all those resources in real time. So what does that look like? From a functional standpoint, it gives us the ability to say things like, "Rind me a cohort of patients with phenotype X. Let's say they've had a prior heart attack and they're on an anticoagulant therapy, and tell me, do we have any blood or other tissue available for them? Where is that stored? Who has that tissue? How can I request it?" That's the sort of model. Under the hood we're talking to lots of different tissue repositories, lots of different sources of data, but our end users don't know that and that's the whole point of this federated approach. We're breaking down these traditional organizational boundaries between the clinical enterprise where the electronic health record lives, our data warehouse which lives in sort of that quasi gray area between that and research, biospecimens repositories that are managed by either the clinical enterprise or research groups, and then research databases.

Again, to our end users, they don't know that because we're acutely using this robust integrative informatics platform. Here's just a screen shot of a prototypical interface where you can see we've just use off the shelf web portal technologies and now people can start looking at these resources with these de-identified cohorts of samples and patients and pull back this phenotype data. Is there tissue available? Is there not? These are designed to be tools that don't

require an informatician or a data analyst. They're tools that are designed for end users. Basic scientists, clinical researchers, clinicians, to actually be able to interact with and understand the data. Lowering the barriers between the end users and the data that they're trying to access.

Let's in the last couple of minutes come around to a few discussion points. So I always like to circle back around because lots of people ask me, you know, since I'm an academic and I work in biomedical informatics -- what's the difference between that and IT computational sciences, etc? The reality is that they're all very closely related to one another and very complimentary. But, at the end of the day, as a biomedical informatician, you know, we're really focused on: how do we deliver timely and contextually appropriate data information knowledge to support basic science, clinical and translational research, clinical care and public health, right? It's about going from data to knowledge. And we use tools and methods that we share with our colleagues in the computational information sciences, in the social sciences, and in the business domains, for example, in order to achieve that. I would argue that the example that I just gave you is a prototypical case where we are, in fact, trying to do exactly that -- break down organizational boundaries, use advanced integrative informatics platforms and link together these data information and knowledge resources so that we can ask and answer important questions that are pertinent to accelerating evidence generation in support of personalized healthcare. And, as I said earlier, one of the key things that we did and why are we able to actually achieve this, because lots of people talk about this, but we actually have end users that are integrating and using these sorts of data and information knowledge resources. It is it because we began with caBIG® technology as a foundation. I think there's a lot of misunderstanding in the community about the caBIG® program and how its technologies are useful in practical use cases. You know, they are not turnkey solutions, right? It is not something that you simply, sort of stick, in your USB key or download from the internet and start using, but I would argue that they greatly accelerate and catalyze by giving us data structures, algorithms, methods, interfaces, and other software components that we can actually then leverage to build these sort of integrative solutions. At the end of the day, they give us tremendous capabilities in terms of our ability to improve clinical research capacity, support these team science paradigms and in particular accelerate T1 translation which is a very problematic area relative to that sort of seventeen to eighteen year gap from initial idea to widespread clinical practice.

I'll also take a minute to just say the primary challenges to doing what I described to you here are not really technical. They are sociotechnical or cultural challenges and they come down to understanding your organizational needs, being able to communicate to your end users so they understand what you're trying to do, being able to train those individuals, and actually bringing those understanding of requirement, workflows and attitudes back to the design process. And these are the fuzzy social science or software engineering or cognitive science activities that are a lot less tidy than what we do when we're talking about just building software. This comes down to things like engaging our leaders, having strategic plans, doing workflow analysis and robust requirements analysis, and perhaps most importantly, having those champions that can serve as the sort of people level intermediaries between you and the community you were trying to affect with this intervention. I'd also say that it's incredibly important in these sorts of projects that we find ways to combine these, sort of, service of delivering these sorts of platforms with science, so at the end of the day we want to bring together innovative platform development and these sort of service related activities in evaluation. And I think there's a lot of opportunities I have enumerated a couple on these slides, I won't belabor them, at these intersection points and

it's something that's the responsibility of everyone involved in these processes to do, because much like we want to create a learning healthcare system, there's an equal opportunity to create learning informatics or IT systems in which we are constantly learning from their implementation and improving them so that we can affect our downstream clients or targeted workflows and processes. And evaluation is key, I always remind people. It's not sufficient just to say, "I built a thing." That's not really sufficient to be able to ask and answer questions about the impact of an informatics intervention. We need to robustly and systematically evaluate these technologies. So with that, and if people don't know, it's a state law in Ohio that every slide deck ends with a football reference.

[Laughter]

PP: So, actually some of my colleagues in the room would probably agree with me, I will say just briefly that's my e-mail address and my web page if people want additional details. It sounds to me like you want to do questions and answers at the end of the session, so unless you want me to take a few now, it's up to you as the chair.

Male: You can take a few now.

PP: Sure.

[Applause]

PP: So any questions? None whatsoever. I have stunned you into silence. Yes.

Female: I'm curious if you're doing this, creating this platform that sounds like it's available for widespread use, what kind of permissions do you have to get from the patients to be able to utilize this information?

PP: So for example the case that I gave you when I talked about the prospective process of getting tissue and data every one of those patients signs an informed consent at the point of care that indicates their understanding that they are using residual tissue in data for research purposes. Now that data is de-identified and there is no way to go back to the patient if we find something clinically relevant to them. It's sort of a secure one way hashing process, but it's a prototypical IRB approved process. It's very costly. I'm not going to lie to you. And there's a tremendous debate. I didn't even get into this today between opt in versus opt out for that. So if you were going to talk to my colleagues at Vanderbilt University for example, they use an opt-out process where it is assumed that every patient will participate in this unless they explicitly say they do not want to. They get information advances that says it's taking place and it is an option for them. But the assumption is they are participants. Other organizations go with opt-in where you assume everybody is not a participant until they expressly say that they will. I think this is an area of great debate that ultimately will have to be satisfied at the federal level in order for us to say which is the preferred approach. The problem with opt-in with what we use is the cost of doing so. So we add an additional step. We go if you've been to the doctor and I'm sure everyone in the room has been to the doctor, hopefully and when you go and register in one of our clinics where we're doing this, right? You register for your encounter and before you see

your physician, you go to a second person and this second person then says, did you get our flyer. There's a little brochure and there's a webpage and a phone number and are you willing to do so. And I will say that of the patients that we approach in this pilot right now, over sort of 85% of those patients say yes. It's not really hard once you explain to patients in an appropriate way what you want to do but I think—

Female: It's for broad use.

PP: Yes.

Female: You're not asking for a specific.

PP: Right. It is a broad umbrella agreement.

Female: Great. Thanks.

[End of Recording]