Feinstein Kean Healthcare
caBIG® Bio IT World Presentations
Lecture by Ken Buetow and Stefan Baumann
Thurs 11:30 a.m. Beacon Hill Track 9

[Applause]

Male: It's my pleasure next to introduce a colleague, Stefan Baumann who is the head of imaging infrastructure at the Novartis Institutes for Biomedical Research and is presenting today on a collaborative effort between the NCI and Novartis to develop a network to support image data exchange in pharma driven clinical research.

Stefan Baumann: Thanks for the introduction. So my name is Stefan Baumann. I'm working in the clinical imaging group in Novartis and I have the very big pleasure to do a co-presentation with Ken on a collaboration that we have started with the NCI. So I'm not going to go into too much detail on what is quoted here but it will suffice to say we are in the imaging domain and imaging is very important for advancing drug development. We have put a lot of effort and resources around systems and processes that support our clinical imaging management in clinical imaging trials. So what are the pharma image exchange challenges if we are doing image trials. Let me give you a little bit of background. So we're talking about a scenario where one sponsor is working with multiple partners that on our behalf collect data as they are acquired at hospitals and they make sure they are of similar quality. They do all the operational work of communicating with the sites and with the hospitals in order to get those images timely, and then they do the analysis, most of the time on behalf of us, and they send to us the results. In recent years Novartis and other pharma companies were interested to not only get those results of the images but actually get a hand of those images because there are a lot of very nice information hidden in those images which we would like to take advantage of. In terms of these needs and processes -- a quick list of what we would like to do with those images: on one hand we would like to do an image quality review to understand that the statistical power of our imaging trials from the quality of the images, we would like to archive those images for various purposes. Obviously we would also like to do method development which means going back to an image that has been analyzed using certain ways and find other ways with which those images could be analyzed with much better signal or finding other things in those images that are important. In terms of the actual analysis, also other pharma companies and Novartis are in some cases interested to do internal in house analysis, especially in exploratory trials that are looking at features and images that are not going to be submitted to the health authorities as part of a trial submission but potentially for internal decision making.

The big problem is the data access, as we noted before. So if you can see that, that is a stack of CDs. That is the traditional way of transporting images from hospitals to CROs; from CROs to sponsors. That's now being changed but a couple of years ago this was the gold standard transfer. We have a couple of issues with that. We have a lot of data. We have no standard organizing principles and it's very hard to find the needle in the haystack here. We have no possibility to encourage quality standards to be maintained at the hospital acquisition site because it's very hard to actually check something in this stack against the quality standard. We, across a couple of pharma companies are looking at this together, decided that this is a pretty competitive field because no one is going to have any advantage of looking through that

haystack and replace it with a better methodology unless everyone is actually doing that and we bring the whole technology back on a better level. So the Novartis approach on that, historically, is we first started on our internal imaging hub. That was a nice way for us to exchange and receive the data through the internet. It did have two challenges, if you want. The one challenge is, since we're talking gigabytes to let's say on the order of terabytes – small, half a terabyte of size for a typical imaging study, there are special interfaces to get access to these transfer methods to get data into the imaging hub. That means it's relatively expensive for us to ramp up every partner one after the other to our specific methods of how we transfer the images even though we think it's nice. The other disadvantage of that approach is that the innovation time to do new stuff is pretty long. We have to ramp up our partners with our specific requirements if we want to do something more interesting like not only transferring all the images from a trial, but doing things like keeping the trial image data at a partner but on demand access of the image for reviewing an exceptional case or something like that. Therefore we've decided to go with a more service oriented architecture type of approach and we have done a proof of concept which is a software called ImageEDC and that software was based on a product that the NCI has leveraged. It was based on the caGrid framework, and this image that you see is a proof of concept application to show that it is actually possible using frameworks out there to do service oriented architectural web based web service based transfers of images from A to B and more. We have put out this code at imageedcgooglecode.com and put a non-bsd style license on top of that. So with this success in mind, we are now moving to a new phase of this where we want to sit together with our pharma peers. We're sitting together with them every now and then and we agree on an isolated process that everyone thinks has a priority in implementing in the imaging trial field and we want to move that forward to a really productive end to end inclusive GSP regulatory aspect, etc. etc.

So the outlook here is how to make this productive and how to scale it up to a lot of trials, for example, and across a lot of partners. This is a very quick view. I won't go too much into detail about this prototype. Just pointing out that the red piece, which is the middleware -- the transport middleware security, is handled by the NCI caGrid. That comes out of caBIG® -- identity provision, authorization transport, service creation. A lot of things that are handled in that layer and we don't need to recreate. On top of this we have, looking at our business processes that we could come up with in the next two years, say identify a few objects and a few, let's say, primitive transactions that could be orchestrated together to actually build the whole space of processes that we want to implement, and then on top of this there is the example ImageEDC application with is an end user application. That has been presented by my colleague Josh Snyder on Wednesday, so if you'd like more information, there are certainly sites available on the internet.

Bottom line, we have recognized that we at Novartis would like to deliver reference applications and processes instead of staying on just the data standardization commodity level so the hope is that we can deliver reference applications and open source as an enabling principle to speed up the innovation process together with our peers. And we hope, using that approach, we can also leverage quick adoption across our pharma peers. And there's a big advantage for us coming out of this is if we talk with one voice to our partners which are mostly commercial research organizations we can also be academic colleagues or other partners -- for them it's much—they're much higher incentivized to actually build against these specifications. They're non-trivial specifications and we wouldn't be able to incentivize them if that was a purely Novartis project.

A few things that remain to do and we're now coming into this piece where we recognize that it would be very nice to do a collaboration with the NCI and so the NCI has, as you may hear later on in Ken's piece of this presentation, already established this paradigm of actually delivering applications instead of just standardization. With that in mind we are now coming closer together. We are helping the NCI with our view on how to regulate the software codes, how to regulate services hosting in a GXP world. That is going on. The outlook here is that we would like to get to a stage where we have a kind of an agile innovation lifestyle where we can really build up a product, roll it out, test it, and then also feedback to the NCI and their vendors our requirements for future iterations of this software so it's not a static thing but it's really a live innovation process.

We want to put a great deal of attention into adding our pharma peers into this so we really have a broad agreement as to what it is that we want in terms of processes to solve our clinical imaging trial challenges. This, at the risk of cluttering the presentation, is a very simple single slide to say that we also have other projects outside of the world of clinical imaging where we want to go with the same approach of delivering a reference open source application. Here we are looking at a application that is programmed by Ed, a colleague of mine who is also sitting in the room, so if you'd like more information on that -- it's a biological integration system that can pair imaging targets with biological pathways and gene information in order to make that a searchable hub and we are in the process of open sourcing this application and here that is just to say that we want to bring this paradigm into a broader level. With this I want to thank you and hand it over to Ken.

Ken Buetow: I'm just going to add a couple of quick comments and actually context setting around this and then we'll open it up for questions overall for the broader group. This represents from our perspective at the NCI a card carrying example of the ecosystem class infrastructure that we think is going to be critical to establishing this new generation paradigm. Both the fact that we, as government, have the opportunity to work and partner with groups like the pharmaceutical industry, but also the fact that a leadership role and other pharmaceutical partners such as Novartis are working together to bring to other members of the industry together so that we can actually all be working together around common sets of standards; common sets of infrastructure. Shown on this slide is just the core community that we're hoping to be able to start to bring together. It's both the industry partners, such as Novartis, but also bringing together other members of—other government sponsors and other members of both the IT infrastructure as well as the healthcare delivery infrastructure. As I mentioned in the earlier talk, we're still struggling with, as a broader community, getting to the point where we have all of the information in translatable form. I think this type of partnership is, in a certain sense, at the center of what makes the internet a successful endeavor in the sense that it's not just about people sitting around tables and, again -- we're big believers in standards; we're big believers in specification, but not just stopping at the level of saying "geez, wouldn't it be cool if we had a standard that did this," and actually taking and, relatively rapidly converting those standards and/or specifications into reference implementation and into code. So that we can actually do this agile cycles of developing and producing things that we find, hopefully, before the standards have hundreds of millions of dollars invested in them in order to be able to find out where they don't work or where they work well. We can actually catch, in early implementation, some of those individual components. Our strategy at the NCI continues to be, as we've done through the caBIG® program to work aggressively with the standards community. In imaging, we continue

to work with RSNA and other groups that are establishing the imaging standards. We work very closely with CDISC to develop to have implementable standards associated with the regulatory submissions. We work very closely to bridge, HL7 with CDISC so that we can leverage the information in clinical encounter in order to be able to capture information in a routine and recyclable way. But again, having just the standards, we think, is an important prerequisite but certainly not the goal line. We take this to a next level of creating service specification so -- how does one implement at the standards in a detailed enough fashion that these service specifications can then actually describe the interfaces between applications? So we're not trying to dictate how people do internal representations or what technology platform one uses in order to instantiate something and again, I think our Novartis collaboration is an example of a place where essentially Novartis can go out and develop their own particular applications and/or frameworks but then just agrees at the interface how to interconnect and how to translate between their pieces and other pieces. Lastly, again, like the internet model, we believe that instantiation in reference implementations is critical so the caGrid framework that Stefan mentioned is our first iteration of a national to international framework. We actually now have folks in nineteen different countries who are using the caGrid framework as a way to interconnect data and actually it's, again, platform neutral in its underlying web services framework and then can be extended to include richer semantics and other components. So my slide piece of this closes with just saying that I think we need to be creating this broader framework and we're very excited about the partnership that we have with Novartis but actually look forward to this being the first step in a much broader collaboration with other members of industry both in the pharmaceutical, biotechnology, and other space. So with that—I'll stop and I'll open the floor to questions. Questions for Stefan or on the project? Thank you.

[Applause]

Male: So you mentioned with other organizations like Novartis -- How about other institutes, research institutes? Not for profits?

KB: One of the big places that we've been working aggressively is with, for instance, NCI designated cancer centers and with other groups who are all working in partnership through our community to try to connect these. So for instance I mentioned this morning one of our activities around leveraging a national biospecimen network. That has more than forty five institutions, the vast majority of them being academic and not for profit institutions. We have groups of—I mentioned the I-SPY trial that's being run a multi-site trial being run at twenty different locations so that's—it almost goes what has been not traditionally recognized I guess that we were excited about this is that this is not just a sort of a research academic toy but in fact has the robustness that would support industry grade activities over and above just what academics may be interested in in experimenting with.

Male: [Inaudible] I have a question for you, Ken. So in the spirit of what you said about large pharma joining in this effort, has there been a concerted effort to identify areas of the research that can be put before competition realm to encourage big pharma's to join such efforts because I imagine that it would be difficult for individual pharma's to decide on IP issue what they're going to lose and what they're going to win. Is there a unified view of that?

KB: I can give the short answer that this is just a new area for us, that we're beginning to explore. I think we have a lot of interest in listening to the industrial partners as to what are the areas that they think are trivially precompetitive. We might speculate that some of those resources might be around the emerging large scale infrastructures necessary to support genetics, genomics and other things that the NCI and NIH are investing tons of money in to build both information resources as well as infrastructure to support so some of that we think could be straightforwardly leveraged, but I think that what we've also found in interesting niche in maybe in the tools that would facilitate doing large scale clinical research the patient recruitment, patient pop—creating a new model of CRO that facilitates bringing patients into the research endeavor in a manner that no single pharma can easily leverage partially because of all the complexity of how one interacts with patients and how one would and concerns, regulatory concerns as to how recruitment is performed. So I think that may be another precompetitive place that one could explore that.

Male: How do you see this specifications or protocols being articulated or talking to [inaudible] constellations of sparkle end points?

KB: So help me understand a little bit of what you're—so our stuff is all actually represented. Our specifications are all well represented in standard semantics on ontologies common data elements that could easily—in our humble opinion could be easily translated into these other larger scale sparkle class types of information representations. Our representations are already codified and broken down into elemental levels either at an ontology level and/or data elements that actually sometimes rise above the specific ontologic representations because we can add additional characteristics to them. But it's already highly structured that actually facilitate the translation into that space.

Male: And we'll be making that bridge between the existing well annotated [inaudible] elements and the [inaudible]?

KB: We work very closely with our federal partners in terms of trying to share our all of the infrastructure that's accessible so the short answer is yes. The slightly longer answer is it's always a time dependant variable by which one goes through and starts to submit out these different pieces, but we have worked very closely with the other federal efforts in data.gov and other efforts.

Male: Thank you.

KB: Please.

Male: I think one of the biggest issues for most of the institutes is how are they going to share their data and what's the reward model for sharing the data? In this echo environment? What is your view on that? I can't imagine what we are looking into say actually if we would work properly we should as academic medical center become publishers of our own data, but there is no model to support that basically although there is huge potential of course to do something like that.

KB: I actually have two answers to that question but I'd also be interested in hearing my industrial partner's concept of that because I think that part of the challenge has to do with the intellectual property associated with the data because as it gets translated into information there are different concerns and different communities. One of the things that Stefan mentioned is that when we talk about the grid and this interconnectivity framework it actually has an elaborate and sophisticated set of access controls so that one of the things that sometimes gets miscommunicated when we talk about sharing the ability to do this is that it sounds like we're immediately posting everything up on data.gov free access for anybody, anytime, anyplace. Anyplace that we can do that we do do that, but an awful lot of the data that we spin have concerns of human subjects protections or has intellectual property concerns. So at the CAGrid infrastructure we have entity level attribute based security models where one can define role based access as defined by the creation of virtual communities, so it can—the access control for any particular element can be ratcheted all the way down to only the person who submitted that data element has access to it, ratcheted to there is a virtual community of people who would maybe be your collaborators or the folks performing a multi-institutional trial or all the way up to I would share it with this particular class of organizations but not with other groups of organizations. So in a certain sense, you could create the access controls necessary and in theory because we indexed who accesses what data, one could create almost virtual publications by actually checking how often datasets are used and who's using them. I think that's an exciting model that's been sort of slow to be adopted so far.

SB: It's interesting that you mentioned this and I completely agree with Ken's answer and can maybe add anecdotal point which is we have the same question, we have among all the images that are in our database a received inquiries from academic partners to say under which circumstances are you able to share that data with us to do research and the question to us is well under what technological and social constraints can we share data and we're currently in the process of actually looking into this question, for myself it would be very highly interesting to actually point out data somewhere on the internet and say whoever has a nice method to do something with that data should try it against this data and then we can start calibrations, but since we're dealing in the clinical space we have to very well align that with questions of who owns the image and it's use and especially privacy. There are a couple of constraints there.

KB: Any other questions? Okay. Thank you.

[Applause]

[End of Recording]