Siamack Ayandeh: Okay great. Thanks Ryan and Brandon for hosting this seminar. My name is Siamack Ayandeh, I am the Chief Architect for Big Data Science. What we are going to talk about today is how cloud technologies are helping and impacting research, in particular, the prospect of VA Enterprise Cloud Enclave built for research. Also, how Prospect compliments other research analytic enclave in VA, how can you get access to VAEC Prospect, and the summary of Prospect value add. Let’s start by looking at VA research priorities and how they are evolving with time.  
  
With research today, data is fueling the research, a lot of data is fueling the research. We see two phenomena. One is that the rates of data, the volumes of data are increasing with an accelerated pace. And the second phenomenon is that we see there are different types of data that are being produced and there are new sources of data. Also, analytics, both for research and operations has historically centered around structure of data. As we have these new types of data, whether it’s general mixed imaging or patient generated, we need new ways of storing and handling this data.   
  
Research also by nature is a collaborative effort. We are partnering more and more and somehow we have to pool this partner data and store it within our research environment. Using the storage capabilities of the cloud is one way to do this.   
  
Also, data sharing has been and is becoming more and more indispensable in conducting research. This can take different forms, whether it is using learning from an observational study to derive the main study question or whether it is linked simulations as a first step towards experimental, let’s say, drug discovery in a lab. Or whether it’s in the form of sharing data to do derived studies.   
  
Also, a phenomenon which is not specific to research but is more generic is that the cloud is becoming the platform for consumer software and applications. A lot of these applications need specialized computer environments. VAEC Prospect, which is an Enclave in VA Enterprise Cloud is built to support all these requirements, offers a place where research data can be stored irrespective of the type or volume of the data. We scale up the volumes from gigabytes, which is traditionally what has been available, to terabytes and petabytes. We offer a specialized compute, graphical processor units, or high number of cores and memory intensive applications, and also new services which are offered as part of the cloud offering for AI machine learning, natural language process, text extraction, and the like. In summary, cloud technologies are enriching existing and enabling new models of conducting research, whether it’s observational or experimental or simulation or derived, the cloud has capabilities to support that.   
  
VA, historically, has had many good data sources. But not until now adequate options to store the data and to perform computations and analyses. Existing enclaves are often attached to more or less single data sources. And we need to move beyond this. Here is an overview of Prospect and a research enclave, which is built, owned, and operated by research, is home to several existing research projects. It’s an analytic environment where you can take your data and do analysis, and as I said, can store different types of data, large volumes of data, and perform the specialized computations. It is paid for by research dollars using the scientific computing provision. It can house KII, THI, and has a moderate ATO, is connected to the VA network and can pull from \_\_\_\_\_ [0:06:39] data sources and can store any type of data.   
  
Also, we can get results which are from vendors and partners. Some of them also operate in the cloud. There are large amounts of both fast local storage as well as what I call our limited bulk storage. And of course, unlimited is a relative term here, relative to our requirements. There are several tiers of storage and this allows for us to balance workload requirements versus cost versus time to results. And of course as I said before, we have specialized compute and access to the latest generation of technology. On premise, usually hardware lags in terms of setup times and by the time we had access to the hardware, it can be several generations behind.   
  
Services is also a big part of the picture. Cloud service providers are beginning to offer services for, let’s say, natural language processing. In the case of Prospect since it is building WAS, we have comprehensive medical, it’s an AI environment which allows you to perform NLP on medical text and notes. It indexes the notes and allows us to process large amounts of notes and index those notes. The notes form a big part of electronic health records, which are currently not fully utilized.  
  
We can also build high availability solutions using native cloud technologies and projects, which often have difficulty for one reason or another, print can find a new life in the cloud as we will see in the next slide. In this slide, what I’m showing is how Prospect compliments some of the existing enclaves within the VA. Many research assistants starting at the top left corner with on prep servers or medical devices, many research systems across VA involve some form of equipment, which we broadly refer to as medical devices. These devices generate large amounts of data and this data needs to be stored and processed, something which is often missing at the location of the device. Prospect can become an extension of these systems where the data is transformed for purposes of further analysis.  
  
Another example is leveraging study data which is generated in a Vinci workspace. Vinci is the place to go for accessing the centra data warehouse where we have all our electronic health records. Once you have your study database, you can export from this database and push it to the prospect study mark for further analysis. This allows you to get access to more storage, specialize computations, et cetera. Or any other limitations that you may be facing in this environment including software applications that, let’s say, eyes on frameworks that a lot of research seems to need nowadays.  
  
We also can pull from third parties and also other data sources within the VA. Some of the VA data sources have plans and some of the applications have plans to move to the VA Enterprise Cloud. Once this happens, they would become accessible locally, meaning within the cloud, which would make like a lot easier. But in due course, this data movement is over the VA network. So it is very secure and the data is simply moving within the VA itself for the most part.  
  
Prospect currently holds terabytes of study data with PHI and PII and the funding and operational model of prospect just like any other enclave in the cloud is pay for use. We can add resources as needed. The infrastructure is so called elastic and caters to burst research workloads. Research by nature tends to burst in intervals. There’s intervals of high activity and then there’s periods of downtime. This is difficult to design on premise because you would have to design to maximum capacity, which would end up being very expensive. And it’s very difficult to predict when these bursts are going to occur.   
  
The advantage of pay for use is also that you're only paying for what you're using for as long as you're using it. For computations, for instance, once you stop your computations, you don’t pay for that machine anymore. If you're on premise, it’s a completely different way of funding and operating these environments. Let’s say if you need the storage, you have to get funding to buy petabytes of storage, incur large capital expense, has daily run rate and operational expense and would take years because you're looking ahead, say, five years would take you five years to fully utilize what you have.  
  
How can you get access to VAEC Prospect? There is a link, I think this link is already provided in the chat, where you can apply for access to Prospect. Basically the application gathers information on the regulatory framework of the project, also what data sets are going to be used, and what is the application stack that you need. If there is need for custom network connectivity, we would have to understand that and set that up, what the storage requirements are, explain how access management process, and the steps needed to gain access are described, what would be the approximate cost and if you're getting paid from, or the leadership supported by that funding, then you need to get the appropriate approval for that.  
  
Provisioning and onboarding will follow. The period is often measured in days and weeks. The study teams may need approval to transfer their data depending on what data they have, to Prospect. If they are coming from a Vinci study space, the transfer from Vinci is often considered a service provided by Vinci staff. This is a good time to thank the Vinci team for supporting the work that we do here in Prospect. If there’s a third party transfer, it is based on an approved cradle or DUA or OMU. It’s up to the research teams and the relationship with their partner to figure out what regulatory requirements they have.   
  
Our final slide, what is VAEC Prospect’s value add, just a summary of what we talked about. It’s built, owned, and operated by research. And it’s paid for by research dollars. It has fast onboarding, offloads to YNT and Enterprise Cloud Solution Office and reduces their workloads. It is worthwhile to point out that when you come to Prospect, you immediately begin to benefit from five years of development work, which has gone into building Prospect and getting it to where it is today. This is work that research teams don’t have to repeat or duplicate.  
  
Currently, we have eight active studies, also five studies are completed or are dormant currently. And we have seven studies which are in various stages of onboarding in our pipeline. Since the middle of 2022, these studies have generated 12 paper admissions, graph applications, so on and so forth. These numbers of course change and change with time. The last snapshot I had a while ago.   
  
In Prospect, within reason, we try to fit infrastructure to the needs of the research methodology and algorithms and support researchers bursting in operation. This leads to faster times to results, as showcased by several studies. We’ve had studies producing results within a few months.   
  
We can also extend, on premises, storage, whether it’s an output from a medical device or backups or whatever it may be. Many projects run short of storage and depending on their applications, Prospect can extend their storage capabilities. Prospect is also open to any VA approved software as the cloud has become the new venue for consuming software via pay for use model. If the software is available as a service, then you can pay for it with cloud credits. If it requires a license, then the study teams have to have access to commercial license for the software that they need, or get that from some books in VAIT.   
  
In terms of network connectivity, we have connectivity to On Prem and can set up connectivity, or I should say transfers to third parties. The nice thing about Prospect is that it is infrastructure as a service and that allows us to fine tune and optimize the infrastructure and therefore the cost. We have only so far used cloud credits to pay for infrastructure and spent zero dollars on provisional services over capital expense.  
  
That’s about it. I just want to, at this point, say a special thanks to our ORD leadership, the Enterprise Cloud Solution Office, and the Vinci Services team for their continuous support of Prospect. And now I hand it over to Brandon for questions. I guess you want me to read the questions, right Brandon?

Brandon Alexander: Yes.

Siamack Ayandeh: So the question is can Prospect be used for archiving study data for closed research studies paid by ORD research dollars or local station research dollars. I’m not sure what is meant by a closed research study. We can archive study data, that’s not a problem. And because it is archiving, it uses storage, the cost is very low. Prospect is currently paid for by ORD research dollars so we are not accepting local station research dollars right now. Next question please.  
  
What is the records management policy set for closed studies and notification of the status of set closed studies. Again I’m not sure, if someone can put in the chat what they mean by closed studies, I think all research studies are regulated. I’m not following the terminology here. The record management policy, the house studies which are governed by some form of protocol or document, and a research body which oversees that research, like IRB as an example. All these record management policies, et cetera, are set or are specific to the protocol that is coming to Prospect and we will follow those policies. Next question please.  
  
Does it, or how does it interact with CDW consumer experience data warehouse, and with Genesis. CDW I think is Central Data Warehouse. And as I mentioned in my presentation and one of my slides, the place to get access to Central Data Warehouse is of course the Vinci Study Space. Once you have created your study database in Vinci, we can take exports from that database and bring it to Prospect for further analysis. Genesis is the environment that supports the Million Veteran Program. It is primarily for conducting general mix analysis, secondary analysis, having genomic pipelines which are already set up. Very power environment, Genesis and MVPR are probably the crown jewel of VA research. At this point, we do not have any connectivity or relationship with Genesis. But to have such relationship is up to the MVP leadership team. Next question.  
  
Is it technically possible that we can run models on Prospect while keeping some communication channels to external cloud environments, parameter updating. The answer is no, we don’t allow external communications to other locations. The connectivity to any location, any computing environment in VA is controlled through what is called a Tick. And there are groups who control what goes in and comes out of a Tick. Any kind of connection that you want to have outside of the VA has to follow a certain process and protocol to be able to establish those connections. This has to be done on a very specific case by case basis but right now we do not support that. Next question please.  
  
This is two questions or a two part question. Closed studies are those which are no longer open with the committees and data cannot be touched any longer. Okay yes. Well that’s something which technically we can store and archive the data but we have to get clarification from our ORD leadership on that.

Brandon Alexander: Right now that’s all the questions in the queue. We’re going to wait for a few more.

Siamack Ayandeh: Okay. This question is the research project has completed its work and decides not to create or deposit the study data, a research data repository for reuse, sorry, and decides not to create or deposit the study data. The research data repository for reuse, the data on the project is \_\_\_\_\_ [0:25:32] securely obtained for the scheduled time period of six years. Is this a question or a statement? I’m not sure if this is a statement or question.

Ryan Sohizad: They said that it is a statement.

Siamack Ayandeh: Okay. Just for me to comment on this statement, yes this is all covered in the protocol, which is governed in the particular research study and protocols have to follow VA policy of course, which in turn follows federal policy. I think we are running out of questions Brandon. One more. Where do these archived files go? We can archive them in bulk storage for as long as necessary. So we have that capability. They remain within Prospect? Yes, it’s still in Prospect. For archived data, it is the Prospect team which would dispose of the records and also the specific protocol or algorithms that you would have to use for disposing of the data records would follow what is specified in the protocol.

Ryan Sohizad: No more questions as of yet.

Siamack Ayandeh: The question is what is the process for archiving these closed studies. Right now, we haven’t had any projects that have reached this stage so we don’t necessarily have particular processing clicks. We work with study teams, we have copies of their protocol, and we would comply by whatever requirements they have in their protocol.  
  
Where should I start when I want to apply for use of Prospect. We put a link in the chat, which is where you can fill in a form for onboarding. And if you need help with that form or before you fill in the form, you can send an email to VAECProspect@va.gov with any questions that you have. I’ll be happy to meet with you before you fill in your form to help you better understand how you can take advantage of Prospect and what we have here. At this point we do not have metadata management software.  
  
Question is can we scan paper research documents for storage. I’m not sure I understand the question. To scan, you need a device which is outside of the Enterprise cloud. You can scan any documents and provided we’ve accepted that you can store them in Prospect, we can transfer them to Prospect if that’s what you're asking. We have a few projects which are doing the natural language processing on various types of documents. I’m not sure what the context of the question is or what exactly you have in mind.  
  
There is an SOP, anybody who goes through. The question is the process for archiving the closed study records must have some sort of processing place as their SOP. Anybody who goes through an ATO like we had to go and has to spend several years getting it would go through questions and right now we have dozens of SOPs. Off the top of my head, I’m not going to say yes or no if we have an SOP. If we don’t have an SOP we are probably in the process of developing one. Prospect has been reviewed by the authorizing official four times so far over the last five years and we have an ATU until March of next year and we are hoping to extend that at that time.   
  
Question is can we customize document image or customize the environments in Prospect. We don’t support containers at this point. The reasons are to do with VA policy more than with regard to bringing software from outside because the image would fall into that category and VA has certain policies and ways and processes for digesting software into VA. That hasn’t been sorted out yet. In terms of customizing environments, you can create your own environment in Canda. We have limited support for that.  
  
Next question is, is it imagined that this would be used to support active enrolling in clinical trials or primarily record research and large computer work. No, we are not supporting enrollment of clinical trials at this point. We don’t have such a project. There are several other systems for enrollment of clinical trials within the VA, which are specific, some of them to specific cohorts which are targeted. We don’t have any of that in Prospect. Right now, Prospect is being used as an analytic environment, meaning we bring in and store research data, we get this data, we don’t interact with primary sources. We interact with Central Data Warehouse, which is the secondary source. And then similar for imaging and for general mix. We bring in this data, different types of data, which is challenging to do elsewhere in VA. We are agnostic to data types and that’s the capability of the storage services that are available in the cloud. And then enable analysis to do computational work. Next question.  
  
Can the study teams use this to store all the study records and set up a local share drive? I’m not sure what you mean by all the study records. Again, any data that you want to bring into Prospect, the data has some kind of data steward somewhere within the VA. You would need the approval of your data steward. For instance, for CDW records and the data stewards just to give you an example, if you are locally generating your own data, then someone at your local facility has to act as that data steward. It all depends on what you mean by studying all the study records. We are focusing on solving problems which are hard to solve on premise.  
  
In other words, fill in the gaps which exist right now in conducting research in VA. Recordkeeping, I think we are good at that as it is. That’s not that urgent or pressing problem that I have seen. What I have seen is study teams often run out of storage for their study data and that is where we come into the picture. We can offer shared drives that you can mount in your environment and simply copy and paste data that would be stored in Prospect. We have different types of storage that we can offer to control the cost of the storage.   
  
That’s a long question. Okay let’s go through this question. It’s a fairly long question so bear with me. If it connects to CDW, CXW is different than CDW. The data link would also contain military personnel service data, veteran contact, health eligibility. Okay, no we don’t connect with CXW.   
  
Question is what is the policy on research compliance. It’s the same policy that exists. This has nothing to do with Prospect. Your research compliance has to do with your IRB and your protocol. Once that audit takes place, if they want to see what you have stored in Prospect, how you isolated that data, how we limit access to that data to study personnel who are listed on the protocol et cetera, we will become part of that audit. We just follow the requirements of VA research and their policies. We don’t set any specific policy of our own.

Brandon Alexander: Still waiting on some at the low end.

Siamack Ayandeh: Well I think at some point we need to wrap it up. If there are more questions, people are welcome to send them to me in an email. I’ll be happy to respond.

Ryan Sohizad: Since we have no other questions, thank you everybody for joining. Please reach out to Siamack with any questions. Slides will be posted on the webinar archives sometime this week. Thank you so much.

Siamack Ayandeh: Thank you.