Cyber Seminar Transcript  
Date: 09/15/2015

Series: VIReC Good Data Practices  
Session: Planning for Data: Early, Often, and Ongoing

Presenter: Sarah Krein  
 *This is an unedited transcript of this session. As such, it may contain omissions or errors due to sound quality or misinterpretation. For clarification or verification of any points in the transcript, please refer to the audio version posted at www.hsrd.research.va.gov/cyberseminars/catalog-archive.cfm.*

Linda: Hello and welcome to VIReC Cyber Seminar Series, Good Data Practices. The purpose of this series is to examine data issues that arise during research and to see how experienced researchers have addressed the issues. Before we begin, I would like to take a moment to acknowledge and thank some of those who have contributed to the development of the series. This series has had active and valuable advisory group that guides our choice of topics and the concepts to be covered in each session and these include Matt Jajusky (PH), Peter Grunaveld (PH), Jennifer Garvin, and Jim Burgess. We really appreciate all of the wisdom that they bring to us. For the Good Data Practices 2015, we are going to start today with Sarah Krein discussing planning for data, early, often, and ongoing during the research study. On Thursday, September 17th, we have a session called Mind the Gap using administrative and claims data to answer your research question and that will be presented by Todd Wagner. On next Tuesday, we will have a presentation by Denise Hines called Decisions, Decisions, Decisions, Selecting Methods and Tools for Data Analysis and we will finish out this four session series on a topic that we have been getting a lot of questions about on the VIReC Help Desk. On September 24th, a week from Thursday, Steven Deppin (PH) will discuss Using Red Cap, Data Management and Studies Linking Primary and Secondary Data. So, with that I will turn the microphone back over to Hera. Thank you very much!

Hera: Thanks Linda. Today’s speaker is Dr. Sarah Krein is a Research Health Science Specialist at the VA Ann Arbor Center for Clinical Management Research and she is a research professor of Internal Medicine at the University of Michigan. Her research interests include management of complex chronic conditions such as chronic pain and organizational behavior and implementation research with the particular focus on enhancing patient safety and preventing healthcare-associated complications. Questions will be monitored during the session and I will present them to Dr. Krein at the end. I am pleased to welcome today’s speaker, Dr. Sarah Krein.

Sarah: Thanks Hera it is really a pleasure to have the opportunity today to talk about planning for data and this go along with early, often, and ongoing since throughout a research this is something that we have to deal with. Just to begin, I have no disclosures and we are going to start off with a poll question for all of you so we know who is on the line. So, the question that we are going to ask about you is what is your type of research and level of experience? So, this is broken up into two parts actually, so, what is your role in research? Are you are research investigator, a data manager analyst, project coordinator, or other and if it is other, you can let us know more specificity in the questions session of the chat.

Molly: Thank you Dr. Krein. We do have a nice responsive audience. About two-thirds of our audience has replied so far and the answers are still streaming in so we will give people a little more time. One answer that has come in through the question section is a Postdoctorate Fellow. So, that is from somebody remarking that they fell under the other category. And it looks like we have capped off at about 78%, so at this time I am going to close the poll and share those results. So, it looks like 29% of our audience replied as research investigator, 40% of our respondents are data managers or analysts, 18% project coordinators, and 14% other. So, thank you to those respondents and we will jump right into the second portion of that question which is how many years of experience do you have with VA Data? Answer option number one, less than one year, option number two, one to three years, option three, three to six years, or seven plus years. And again, we have had about two-thirds of our audience vote and those are still streaming in so we will give people a little more time. Some other answers have come in regarding the last question about what is your role. One person says that are a research informatics manager who oversees data managers, research informatics software development and operations. Another person writes in that they are an interviewer. So, thank you to those respondents. Regarding this poll question, it looks like we have capped on 80% response rate. So, I am going to go ahead and close the poll and share those results. It looks like just over a third of our audience respondents have less than one year of experience with VA Data, 28% have one to three years experience, 18% three to six years, and 17% seven plus years. So, a nice spread across the board. Dr. Krein, I am going to turn it back over to you now so you will see that pop up again. There we go.

Sarah: Thanks a lot Molly and thanks to the audience. It is very helpful to have a sense of who folks are and what they do in VA and VA research. So, I just want to mention that this session by design is intended to be pretty introductory. So very basic but for those of you who have a lot of experience, hopefully, you can also contribute in the question and answer portion of this session or let me know if there are things that I have missed along the way that you think might be important for our more junior folks to know about as they go forward with their research in VA. So, the course of the next 40 to 45 minutes is going to include talking about factors that influence data needs. We are then going to discuss some data related questions that need to be answered early in protocol development. Part of what I am going to do during that section of the talk is introduce an example and kind of walk through some issues using an actual research example which hopefully will be useful again for those of you who are new to VA and VA research in particular. And then we will talk some special issues related to data privacy and security and data for research versus operations or quality improvement since many of us may not only be doing research in VA but also working with operational partners or have funding that is operational in nature and more improvement type of work.

So, to begin I just want to address the question of why is advanced planning related to data important? And I have outlined a couple of things that I think are important but again there could be some other things that I do not have on this list that others can contribute. But, for me, thinking about data in advance is important to determine whether or not my project is even feasible and whether or not I am going to be able to do what I am intending to do and answer the questions that I like to answer in my study. It is also important when we think about developing our study protocol. Often we will start out very broad with our research but then we have to get very specific about what kind of data are we using even down to the variable level so we are going to need have that in depth knowledge to develop our study protocol. Thinking about data can be useful to identify potential issues before you start so making sure again that what you are planning to do can be done before you are half way through on your project or have written up most of your project and find out that is really not going to work the way you intended. Now, that is not to say that does not happen, but it is something if you are thinking about some of these issues early may help you make things smoother on the backside. Advanced planning as it relates to data is also important for study team composition so I find this really useful because often I would like to bring members of my team together early on. They can help me to develop my protocol so I need to know about what kind of data am I going to be using and what kind of study am I going to be conducting. Is it going to be qualitative? Is it going to be quantitative and who do I need on my study team? Budgeting, of course we have financial constraints, so we are going to have to think about what our budget can hold with respect to the data that we might need to collect and then our study timeline. Often data issues can drive what our timeline is going to look like for our project. But, perhaps more broadly, I think one of my heroes has described really the reason why advanced planning is important and that is because if you do not know where you are going, you might not get there so data really comes into play if we think about what our end game is and how we are going to get there. All right, let us talk about some factors that influence data needs. So, there are three things that I think are critical with respect to thinking about the data that we are going to use and the data that is going to be necessary for our project. First and foremost of course is the goals of the study. What are our research questions? What are our study aims? A related issue of course is the study design so is this going to be a prospective study, a retrospective study, interventional study, cross sectional study, longitudinal study, quantitative study, or qualitative study. All of those things have implications as they relate to data and then resources. Again, we talked about budgeting but even broadly what kind of personnel, what kind of finances do I need in order to do my study. But, the other thing to keep in mind is that while I am thinking about my data needs and identifying potential data sources. That is also going to help to reshape and shape my goals and designs so there is really an iterative process that comes into play. I may start out with certain study goals, a certain study design, look at the data that might be available, go back to the drawing board, adapt, modify my design, maybe adapt modify my goals, look back at the data, and it can go on for some time but it is really an important part of the process when it comes to doing research.

So, now let us talk about some data related questions that need to be answered early in protocol development and at the end of the session we have these all listed in kind of a checklist format. Again, just things to keep in mind as you begin your work. So, the first question that I generally ask after I have outlined what my research goals might be or my research questions, is can I use existing data sources and getting more specific, is there VA data sources or maybe even non-VA data sources that might lend themselves to answering my research question and I need to be thinking about what is the scope or the scale of the project that I want to do and is there data at that level so is it going to be a national project where I want data across VA. Is it going to be a regional or a \_\_\_\_\_ [00:11:20] level project or will local data be sufficient? I need to be thinking about what is my level or unit of analysis. So am I looking at a patient specific analysis or an encounter-specific analysis or maybe I am thinking about something at the provider level or at the facility level. What is the timeframe? Is there going to be sufficient data out there to answer the questions that I want to answer so do I have enough data points and if I am looking at perhaps doing something that includes VA and non-VA data, do I have concurrent data in order to answer my research questions? A related question of course is do I need to collect data directly? Maybe I have decided there is no existing data or maybe I have a study that needs to have different types of data collected in order to answer my research questions. So, if I needed to collect data directly or primary data collection, now again there is certain specific questions that I am going to have to delve into and this includes how am I going to collect my data. Is it going to be a survey or will this be a qualitative project? Will I be doing interviews, focus groups, or observations? Is there the need for direct measurement? So, an example in this domain might be we have done a lot of research around diabetes so maybe I need to do point of care hemoglobin A1c testing and get glycemic measurements on patients or functional measures. We often see that in work as well, where you have patients come in for some type of functional or physiologic measurement and then how frequently do I need to collect those data. So, is it going to be just a baseline measurement? Do I have to do a follow up measurement? Do I have interim time points? All of these things have implications for my research and for the data that I am going to be utilizing. And getting even more pragmatic now, the question of how will I obtain my data? So, maybe I have settled on there are some existing data that I can utilize. There is a secondary data source but there is even more questions that we now need to address. So, are those data already extracting and do they reside in an existing data repository. So, many of us may be familiar with the Corporate Data Warehouse or the SAS Medical Data Sets which are existing data resources that pull data out of the VA Electronic Medical Records System. So, perhaps there is data already residing there in an extractable format that I can utilize to answer my research questions. But, sometimes that is not the case and this often comes up when I am working with clinical colleagues who may be familiar with patient care and we start to talk about a particular research topic and they say of course those data are there. I can see it when I go into CPRS. Well simply because you can see it in CPRS does not mean that it already resides in an extractable format so in some cases you may think the data are there but would require manual extraction of some sort in order to use those data for your research study. Obviously, the implications are very different when you think about resources, personnel and study timeline depending on how you are planning to obtain those data. So, if it is a manual chart extraction process, it is probably going to take a lot more time and may actually change the scope of the project that you are planning to do. Then, non-VA data sets. Are there existing data that the VA already has that I can utilize? So, an example would be the VA CMS data or perhaps the National Death Index Data that the VA has already been collecting that I might be able to get access to or do I need to think about purchasing data outside the VA for example the American Hospital Association Survey Data. We have used that in several of our studies but it is something that I would need to perhaps get on my own and purchase. With primary data collection, again we have a lot of logistical issues that we need to address as we think about what kind of data we might have and this includes things like how am I going to identify the sample for whatever primary data collection I intend to do. What kind of methods am I going to be using? Does that require certain software of personnel? Do I need specific equipment or tools? So, again if I am doing something with a direct measurement, maybe I need to have a point of care measurement system, maybe I need to have some type of technology to help me do physiologic measures whether it is a blood pressure or other things and those all have implications for my data collection.

So, next we are going to focus again on a specific research topic and just look at this in kind of a step-by-step progression just to give you some sense of how this would look when you get to your research project. This is a very specific example but I think it will show you some of the steps that you might take.

So, the topic that I wanted to focus on is something that my colleagues and I have actually been working on recently which is to look at a particular device called the peripherally inserted central catheter or PIC line. The reason why we are interested in this device is we know that there is increasing use and this is based primarily on data from outside of the VA. We know that the use of this device is highly variable and perhaps not always appropriate and also that the device itself is associated with complications so there is patient safety implications. I can also tell you that little is known about PIC use practices or appropriateness in VA so it seems like something that would be good for us to begin to research. With that in mind, we came up with a number of research questions. Again, these are just some examples that things hopefully going forward will actually have an opportunity to study further but one of the questions that we had which is very basic is what is the volume of PIC use at VA Medical Center? A second question, what factors influence PIC use care and management in VA Medical Centers? So, if we are thinking more about something interventional down the road. We would want to know more about that topic. A third question, what are staff perceptions in practices related to PIC use in VA and we think that is helpful to know since physicians may order the device but there is other staff involved in insertion and maintenance of the device. Number four, what is the Veteran experience related to PIC use in VA? This is important since many patients go home with these devices so they are also involved in care and maintenance. As you can see, each of these questions is going to have different data implications. So, next what I would like to do is just look at research question number one which is what is the volume of PIC use at VA Medical Centers and actually open up another poll and ask you what would be your choice of data to answer that research question. Again, the question is what is the volume of PIC use at VA Medical Centers and would you use option one existing VA data. Option two, would you conduct a survey of VA Medical Centers. Option three, would you create a new data repository? Option four is other. Again, if you have any suggestions perhaps type that in.

All right, looks like we have some results. So, it looks like about 85% said they would you existing VA data, 9% survey of VA Medical Centers, 2% create a new data repository, and 3% have other suggestions. So, I am looking forward to seeing what those are as we go forward with the presentation. Well, as the majority of you suggested, our answer was to use existing VA data sources and the reason we thought this would be a good strategy is one, we know that the insertion of this device is a procedure and that there is data on procedure codes so it seemed pretty likely that we might be able to find some information. We also know that the device itself is a product and that the VA has data on products so that might also help us and we also know that the device is associated with infectious complications and that in VA there is data being collected and reported outside the VA as well on certain infectious complications. So, it seemed to us as well that existing data sources might help us in addressing that question.

But, now that we know there might be some existing data sources, we actually have to identify what those data sources are to answer our question. So, some ways that we might go about that would be to look in the literature and find out has anybody else already looked at this topic either in VA or outside of VA and what data sources were they using and what specific data sources were they using? I am usually inclined to also talk with my colleagues so this would be both research and clinical colleagues to find out if they have suggestions. There may be other researchers that you know who worked with different data sources and could point you in a certain direction but also you may want to talk with your clinical colleagues as they may be able to tell you about how they document about a particular issue and in this case, for me, it was about PIC insertion. A third strategy that you might want to consider for those of you again who have not been in the VA very long, is posting an inquiry on the HSRData Listserv and this is really a nice resource because it allows you to sort of pick the brains if you will of researchers across VA and often can get some really good feedback that will help lead you in the right directions. Even more specifically now, I need to know where can I find information about those VA data sources because I am going to have to do some due diligence and really find out what is in these data sources and which data sources are going to best be utilized for my question. So, some of the sources that I am most inclined to use is what I have listed on this particular slide and that includes the VA information resource center or VIReC website, the Health Economics Resource Center or HERC website the VHA Data Portal, the VA Support Services Center or VSSC website, and also sometimes VHA Program Offices. We will look at each of these in a little bit more detail and also there are links at the end of the presentation for those of you have downloaded it to each of these websites.

So, the first website that I often will look at when I am thinking about existing VA data is the VIReC website and as you can see just from this screenshot you can find out a lot about working with VA data from VIReC. You can also get to research user guides that have very detailed information so I can start to think are there certain variables and in which data sources are those variables that might be useful for answering my research question and their summary information as well so I can figure out how many years of data might be available and a number of different aspects about working with these different VA data sources. Another website again is the Health Economics Resource Center and I have also found this to be quite useful in this case because I am working with something that is a product and for which there is cost implications. I know that HERC has done a lot of work in documenting certain data sources so they may also be able to help me figure out which of those specific data sources would be useful to answer my questions. And then the VHA Data Portal and this is described on the slide as becoming the one stop shop for data users needs. So, this is also an extremely helpful resource with respect to talking about different data sources and it will link you into even more information about these different data sources but also for data access because you are going to need to know once you have identified those sources, how do I actually access them. So, the VHA Data Portal is going to help you in that regard.

Another resource is the VHA Support Service Center, VSSC, and I found this to be useful also in that over time, VA has been collecting a lot of data for different operational purposes and some of those data reside in data cubes and you may find out more about those through the VSSC website. Then last but not least, I have also found some of the VHA program offices to have information about different data sources and so for my purposes I happen to know and after talking with some of my colleagues, that analytics in business intelligence which houses the inpatient evaluation center might be a possibility for the work that I am proposing to do since that is the entity that has been collecting data and reporting data about infectious complications. So, sometimes you have to be a little creative and do a little bit of poking around to find some of these different resources but again when talking with colleagues you may find things that you did not know existed and get to different potential data sources.

So, in terms of answering my research question of what is the volume of PIC use at VA Medical Centers, I have now narrowed it down to existing national VHA data. I have decided I want to do this across VA. Looking at the VRIC website and the VHA Data Portal, I have settled on data from the Corporate Data Warehouse and from SAS medical data sets. In looking at the program office information and the inpatient evaluation center data, I have now learned I can pretty much take that one off the list. It is not going to be sufficient for answering my question even though they have information about central line associated bloodstream infections and they also have information about number of device days, it is not specific to the particular device that I am interested in. It is actually for all central venous catheters. So, it is not going to be useful to answer my question.

So, my next step of course, is I now have some data sources and I could certainly haphazardly go about writing up my protocol and saying I am going to use the CDW and the SAS medical data sets and even name some of the variables that I think are going to let me answer my research question. But, I am often a little bit more cautious than that and so now, I am beginning to think about data access. So, I would go back to the VHA Data Portal to learn about how I am going to be able to access these data sources. Now, there are two different types of access. So, if I was going to go forward with a pilot study for example and wanted to actually do research with the data that I am going to access, I would be inclined to write a protocol, have an IRB approval. At that point, I could go to the data access request tracker request process, the DART process as you may hear about it. But, because I have decided that I still do not know enough about the data that I am going to be working with, I am going to use a preparatory to research data access request instead. I think this is really a useful strategy. Again, if you think the data might work but you are really not sure to get a sense of how useful will it be and can I really answer the questions that I have in mind. So, using a preparatory to research process which is a provision of the HIPPA privacy rule in VA handbook 1200.05, the VA can provide personal health information to develop a research protocol and that is without patient authorization or without an IRB approval but there are some limitations. Remember this is only for preparatory work. I cannot publish on this. I cannot use this for my actual research. So, back to VHA Data Portal, we would go where it tells us about this preparatory to research process and the request process and I have to say the tools have gotten a lot better over time that there actually is a process for this. It use to be a little bit more difficult to figure out how you would access data for these purposes but they are really doing a nice job now with the VHA Data Portal to help lead you through the different steps that you are going to need to take. So, you can see on this slide the data sources I have identified, the CDW and the Medical SAS data sets. The scope of those data, the types of identifiers and then the process itself which is simply in this case a memo where I have to specify the type of data I need, where the data are going to be stored, a little bit about what data I am proposing to access. I then need to get a few signoffs and in this case, the Information Security Officer is one of those so depending on what facility you are at, you may need to have a little time in order to get all of those signoffs but it is really a fairly expedited process in order to get access to the data for preparatory purposes. So, now having done due diligence with my data and still looking at my initial research question, what is the volume of PIC use at VA Medical Centers, I can tell you that the data are not going to help me answer that question or not sufficient to help me answer that question. I have gone through and looked at data from the SAS medical data sources. We pulled out the CPT codes. These are procedural codes that are PIC specific but they generally reside in the outpatient data and do not necessarily capture all of the insertions on the inpatient side so I am getting an undercount. We also looked at the SAS medical data sets on the inpatient side and looked at procedure codes once again. In this case, ICD-9 codes although I guess now they will be ICD-10 going forward. Those are not specific enough. They give me all central venous catheters. So, now I have an over count. Then, looking at the CDW data, we were able to find some event capture data, which I had identified after talking with our PIC team, and about how they actually document their insertion and we thought, oh, we found the perfect data source. Unfortunately, not all facilities use that particular package and so we were again not getting full capture across the VA system. So, having done all of this work, sometimes what happens is you end up having to go to plan B and figure out another strategy to answer your question. But, this is again highlighting why it is so important to be thinking about data early, often, and ongoing.

So, we just want to go through one other quick example again with a different type of data. So, in this case, what I am going to do is ask you about research question number three, which is what, are staff perceptions and practices related to PIC use in VHA and again I would like to know from you how would you obtain or collect data to answer this research question? So, oops, it looks like our quick poll went a little bit too quick. Well the choices are to use existing VA data to conduct a survey to use qualitative methods which could be interviews or focus groups or again other if people have other suggestions for how they would collect data to answer that research question. There we go. So, I will give people just a few seconds to answer the poll. Great so, it looks like our results are in. So, about 9% of you suggest using existing VA data, 47% would conduct a survey, 41% would use qualitative methods, and 2% have an other response. Well I think most of you are on the same wavelength, as we were in terms of thinking about this question. So, our answer to how we might address the question of what are staff perceptions and practices related to PIC use in VA were to conduct a survey and use qualitative methods. So, I think most folks were heading down that same road. But in thinking about a survey, I just want to talk a little bit more about some of the considerations now that we have settled on that as a potential strategy that you are going to need to take into account. So of course, in order to do a survey, I am going to have to figure out who am I going to survey and how am I going to identify those people. Is there a feasible way for me to get to the most likely respondents for my survey? In this case, the survey might be directed to those who insert the device but again I am going to have to figure out how can I identify them. The next consideration of course is how am I going to conduct my survey so even if I can identify these individuals can I conduct my survey electronically. Would I need to do a paper base survey? Would I need to do a perhaps a telephone-assisted survey of some sort and what kind of survey instrument might I use? I am indicating red cap here just because I know that there is an upcoming seminar on red cap use and that is one of the strategies in technology that is currently being utilized to conduct surveys in VA. Then another consideration as I am thinking about my survey is potential regulatory consideration. So again for those of you who have perhaps not been in VA very long when we are talking about surveying VA staff especially at a national level, there are certain regulations that come into play so I am going to need to build all of these things into my protocol and make sure that I have sufficient time to get these approvals. This includes the organizational assessment subcommittee review and national union notification. So, all things that have implications for collecting those types of data. All right, so we are going to move on to just a few more questions that we are going to have to address as we are still preparing our research protocol. Once I have settled on the type of data that I am going to collect and how I might obtain some of those data, the next question is how will I store and manage my data? Now for many of us, the traditional mechanism has been to use our local server. I think this is still a possibility in certain sites and still something that is being done but another consideration might be to use the VINCI workspace and I know there is some emphasis now on using VINCI and a lot of potential benefits to using the VINCI workspace as well. So, I certainly encourage you to look at that as an option. We also need to think about what kind of security we are going to need for our data whether it is stored electronically but also data that might be stored in paper form. Is there enough physical storage for me to conduct my project? Do I need to have more filing cabinets or do I have enough storage if it is on the server as well and then what accessible to collaborators in \_\_\_\_\_ [00:34:41] might I need? This is something again for many of us who are collaborating with folks across the VA or even outside the VA, if they are going to need to have access to data, this is going to have to be something that is written into my protocol. I have to make sure that I have a mechanism for them to have access to the right level of data. Another consideration of course is how will I analyze my data and the two questions that I often think about in this domain is what kind of software do I need and what kind of personnel do I need. The reason that I wanted to just mention this is often in the work that I have reviewed when there is quantitative work you see software mentioned whether it is SAS or Stata or ARB. You see personnel mentioned. There is often the file statistician. There is a data analyst and the data manager but sometimes when I review qualitative work people forget that they may also want to use software whether it is In-Vivo or ATLAS.ti or something else and also that you need to have a qualitative methodologist, maybe you need to have qualitative data collectors. You may need to have qualitative analysts so these are important considerations again in who I am going to have on my team and what kind of budget I am going to need for my project.

All right, we are just going to end by talking about two smaller issues but not less important issues. The first is privacy and security consideration. So even in the development stage, this is something that comes into play and for those of you have written grant applications before you know that if it is a human studies application that you have to have a human studies section as part of that application. So, thinking about privacy and security consideration is very important and also can help you as you even further define your protocol. So one question that I have listed here is who are the study subjects and that may sound like kind of strange question but again just based on experience and reviewing a lot of applications in studies where there is the involvement of patients as well as staff. Often what I will find in the human studies section is a very nice write up about patient confidentiality and protection but people sometimes forget that if it is a research project and they are actually doing something with staff as well, there may have to be a section about staff as research subjects.

There are a number of steps that can be taken to ensure protection with respect to privacy and confidentiality. I am not going to go into any detail but I did want to give you a few resources again to help as you think about what it is going to take in the projects that you are proposing and that includes looking at VHA handbook 1200.05 the requirements for the protection of human subjects in research. The program for research integrity, development and education are the pride program. There are linkages at the end of the presentation to these resources. Then also, I would highly encourage you to consider talking with your local IRB or Human Studies Coordinator, or the Chair of the Committee or Committee Member if you have any questions as they may be able to help you also understand some of the considerations that you need to take and some of the strategies that you can use to ensure privacy, security, and confidentiality. A somewhat related issue, which I have to admit I do not know a lot about, but I know that there is a lot of focus and increasing focus on is data sharing and so this is just an example. This is an editorial from the BMJ but I also know that in NIH applications of a certain size, you have to have a data-sharing plan. So, again something that may be a consideration for your project or maybe not but something that you may need to think about again early rather than later in protocol development. Then a final consideration is whether the data are for research or for operational or quality improvement purposes. This is important again early on because especially as you get to the data request process and if you go to the VHA Data Portal, you are going to see that there are actually differences in the data request process. Now I do not know that there is necessarily differences in data management and storage whether the data is for research or operations but I can tell you that at least in our center we have made the decision to keep those two data sources separated. Part of that is because there is different approval processes for those data and at some point, some of the data that we have collected for operational or quality improvement projects could be utilized for research and so we want to make sure that we are distinguishing those two types of data. Then I would suggest that there is probably no difference when it comes to data security that whether the data are collected for research purposes or for quality improvement purposes I would say that you still need to ensure security, privacy and confidentiality.

You know as I mentioned earlier in the presentation, we put together just a bit of a checklist for you that we hope will be useful in thinking about what these advanced planning questions would look like as you begin develop your research plan, your research grant application, your research protocol. So, the questions include what are my study goals and proposed design and at that point you may want to think about is it research or operations based on those goals and potential design. Next kind of question is about using existing data sources and if so what are potential data sources. Is it VA data that I am going to be using or non-VA data and are the data feasible for addressing my study aims? So at this point, you might want to consider that preparatory to research data access. A third question, do I need to collect data directly and if so, what type of data am I going to be collecting and from whom am I going to collect those data. How will I obtain the data so whether it is secondary or primary I need to get very specific about kind of the nuts and bolts about is this data that I can extract? Do I need to do a chart review? Do I need to have special tools and technology to collect those data? Another question, how will store and manage my data? Again as this point, you may want to look at the VINCI workspace as a potential option. How will I analyze my data? Again, this is going to help me think about software and the next question, who do I need on my study team? So, do I need to have qualitative and quantitative folks and getting them involved early in the process can be really critical because then maybe it will help lead you into some direction with respect to your design based on the data that are available or that you need to collect. Then of course, when the rubber hits the road it is what is required study budget and timeline. Is it affordable? They may have the best plan ever but if it is not going to sit within the funding limits, I may have a bit of a problem. The next two slides are just to again remind folks that you need to be fairly specific in your research protocol and in some of the subsequent documents that support that protocol. This includes when it is being submitted initially for review by a review committee but also subsequently when you go through the steps of IRB approval and then if you are requesting data access there is some additional approvals that come along and people may be looking for some very specific pieces of information. So, you have to be specific about the primary data collection elements, about secondary data sources that might be needed about whether you are using electronic health record data, about identifiers needed and how they will be used. This is one that I think sometimes I do not do enough due diligence on in thinking that we need to have truly identifiable data when maybe we really do not. Is there other protected health information that you are going to be utilizing in your project? What level of data will you be using? Is it national, \_\_\_\_\_ [00:42:48] level or local? Are there secondary research sites that are going to need access to the data and if so which sites are those. Again, this is something that comes into play as you get into that access process. Some of the things that you are going to need if you are going forward with the data access request include if you have a HIPPA waiver which again has to outline some of the issues related to protected health information and be very specific about data sources and in scope. Be as specific as possible if you are asking for national electronic health record data access so for those of you who are familiar with CAPRI and VistA web these are mechanisms to be able to access electronic medical record data across the VA but with the approval processes you have to be very specific about what kind of data you are going to be requesting. So, my final question which I often ask myself but also would ask to the audience today as have I covered all of my bases? And with that, thank you very much. My contact information if anybody wants to contact me after the talk and I would love to hear any questions that have come up and any suggestions about data sources during the course of this presentation. So, thank you very much!

Hera: Thank you Sarah for your presentation. As of right now, we do not have any questions from the audience but if you do have questions later you can of course email Sarah at her email address or you can contact the VIReC Help Desk at virec@va.gov. Thanks again Sarah for putting together this presentation. Our next session in the series will be presented by Dr. Todd Wagner from HERC, the Health Economics Resource Center. His session is entitled Mind the Gap. Using administrative and claims data to answer your research questions. It is scheduled for Thursday, September 17th at 2 p.m. Eastern. We hope to see you there. Molly will be posting the session evaluation shortly. Please take a few moments to answer those questions. We really do go through all of them and take your feedback into consideration in planning these sessions. Molly can I turn things over to you? Molly? It looks like we just got a question in. Sarah, how do you know which data sources to start with in the VA?

Sarah: That is good question, so again my suggestion is generally to go and look at some of the websites that I had mentioned and actually I am now sharing with you some of the data information resources and the website linkages. So, I often will start with the VIReC website, the HERC website, and the VHA Data Portal and start to look at some of the different data sources that are available and they are really a wealth of information in telling you what kinds of information you might be able to get out of those data sources. My general strategy, I do not know if others on the call have other strategies, again and I also talk with other people if I have colleagues either research or clinical to get their sense of potential data sources.

Molly: This is Molly. I am sorry about that here. I did not mean to go radio silence on you. My mike button got stuck. Anyways, yes I want to thank Dr. Krein for presenting for us and for Linda and Hera in your help getting this series together. As Hera mentioned, I am going to close out the session in just a second and put up the feedback form and so please do take just a moment to look closely at those questions and answers and we will take them into consideration when planning future sessions. Hera do you want to take the last question that just came in or should we wrap up and have them directed.

Hera: Yeah we take it, let us go over it.

Molly: Okay.

Hera: So, Sarah, one question for you. You did not cover what happens at the end of the project. What happens to the data or programs, keep, destroy, \_\_\_\_\_ [00:47:16]? How long do you keep data programs after the end of the project?

Sarah: Wow! Yeah, that is a great question. I mean so I was only thinking about the preparatory stage. I guess I was not thinking about the end stage. Maybe we need to have a session on that. Part of it was traditionally driven by the record control system and what we could or could not destroy which was pretty much could not destroy anything. So, I know there have been some changes and probably need to think about that a little bit more. I mean we are inclined to again most of our work at least the electronic work we have been archiving away even when the project is closed but not destroying the data since we were not allowed to and from many of the other things you know having a system for just maintaining those pieces of information but I would be interested if others also have some suggestions. I know I have not thought about this probably enough to give you a really good answer at this point but it is something again. You are right, we probably should be thinking about that more up front as much up front as we are on the backside since we have to have a process in place for maintaining any information that is critical to the project and also by regulation that needs to be maintained.

Hera: It sounds like a new RCS 10-1 requires retention of PI records and data for six years.

Sarah: Great, thanks. I knew there was a new one out but I could not have recited that information so thanks.

Hera: All right thank you.

Molly: Excellent, great. Well, thank you for the great questions and a lot of people are writing in and asking about the slides. You actually already have access to them. Just refer back to the reminder email you got this morning from HSRD Cyber Seminar and you will find the hyperlink leading to the slides right there and the questions keep coming in here. I will leave it up to you if you want to take it or not.

Hera: Enough time, so let us go ahead. You can access data via VINCI website as well. Is that correct?

Sarah: You can access data through the VINCI workspace correct but there is a request process as well so if you go to the VHA data portal again there is information there about VINCI and how you can work within VINCI and I think VIReC, is that correct Hera, you guys also have some information about VINCI and the VINCI workspace I believe on the VIReC website.

Hera: Yes we do.

Sarah: There is also some VINCI Cyber Seminars. I know that there has been several talking about VINCI and working in the VINCI web space. So, watch for those. I think they are really insightful.

Molly: Indeed they are. Thank you we have two years worth of VINCI Cyber Seminars going over data access and those can be found on our online archive catalog. You can just go to the right hand side of the page and use the drop down menu to filter by series and select VINCI and we do hold our VINCI Cyber Seminars the first Thursday of each month at 3 p.m. Eastern and we encourage you to go to the registration catalog and sign up for those. So, thank you for plugging that Sarah. Great, well thanks again everyone. I am going to go ahead and close out the session now. So, good luck with gathering your data. Have a good one everybody!