Cyber Seminar Transcript  
Date: 11/01/2016  
Series: VIREC Databases & Methods  
Session: Requesting Approval for Access to VA Data  
Presenter: Linda Kok  
 *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*](http://www.hsrd.research.va.gov/cyberseminars/catalog-archive.cfm)

Cheryl: Welcome everyone to VIReC’s Database & Methods Series. Today’s session is requesting access to VA Data. Thank you to CIDER for providing technical and promotional support for this series. Today’s speaker is Linda Kok, MA. Linda is the Technical and Privacy Liaison at the VA Information Resource Center. She advises on policy for VIReC CMS project which provides VA researchers with data from Medicare, Medicaid and other CMS databases. Linda’s other focus is providing information about and improving access to data in the VA. Linda reviews start access requests that include real SSN identifiers on behalf of ORD. Linda is a member of the VHA Data Access Board and works with National Data Systems and other data stewards to operationalize policies that will streamline VHA data access processes.

Any questions you have for Linda will be monitored during the talk and I will present them to her at the end of the session. As a reminder, a brief evaluation questionnaire will pop up when we close the session. If possible, please stay until the very end and take a few minutes to complete it. I am pleased to welcome today’s speaker, Linda Kok.

Linda Kok: Thank you Cheryl. And thank you all for joining us by phone. I hope we will get through much of the data that I provided. I will have way too many slides. Some of which I’ve already assigned in the back. These seminars are given by VIReC staff and VA researchers in the field to provide information for you about data sources and systems used to access data in the VA. They also provide examples of how data can be applied to research and quality improvement questions. Other goals of the seminar are to address some of the limitations of secondary data use and to provide a glimpse into other resources to support data use.

This is our upcoming schedule for database and methods. You can register for any upcoming seminars through VIReC or CIDER. The seminars are given the first Monday of every month at 1 p.m. eastern except in the case of VA holidays. You can access previous seminars in the HSR&D Cyber Seminar archive.

Thank you for joining us as we talk about requesting access to data in the VA. While other sessions in the VIReC Database & Methods Cyber Seminar Series describe VA data sources, in this session we focus on clarifying procedures for requesting access approval for the most frequently used VA data sources. In this talk you will learn about the three categories of data access in the VA, general and specific requirements for requesting access to the data for each category and you’ll learn where to look for information about requesting access to the specific data sources that you need for your project.

We’ll pause for a moment to find out more about you and your roles in VA research and quality improvement activities. Heidi.

Heidi: Let me open that poll quick. And this is a please all that apply. Are you a new VA researcher with less than three years’ experience? Are you an experienced VA researcher? Quality improvement investigator? Project coordinator, data manager, data analyst? Or other? And with the other, please take a moment and let us know in that question box what your role is there. We’re always really interested to know who is here and what sort of information people are looking for so we would love to get that little bit of extra information. Responses are coming in nicely. I’m going to give everyone just a few moments. The numbers are still fluctuating a little bit here so I’ll wait for them to slow down before I close out and go through the responses. And I’m going to close that out. What we are seeing is 31% saying new VA researcher, 13% experienced VA researcher, 10% quality improvement investigator, 48% project coordinator, data manager, data analyst and 18% other. What we are receiving in the question box is up. \_\_\_\_\_ [00:05:18] concierge, research pharmacist, support clinical research, research assistant who has been working for the VA for less than two years, social worker currently pursuing a PhD program and providing program support to their service, human subject protocol manager, a role in VA research nonprofit foundation and association role, PCS intern informatics. Thank you everyone for participating.

Linda Kok: Thank you, Heidi. Well I think the breadth of the specialities of the audience is probably good for this presentation. We spend a lot of time talking about the various roles. If you are a new VA data user or your project will need data you haven’t used before, the material we’ll cover in this session should save you some time and headaches. If you’re a data steward, who are sometimes on the call, the session may help to clarify the distinctions between the different roles that individuals and researchers can fill.

My machine which has worked perfectly up until this moment.

Heidi: Nope, you just needed to click on your slides. So you should be good to go now.

Linda Kok: Okay. Thank you. Today we’ll start by looking at the general requirements for data access that apply to every data user. Next, we’ll identify the distinctions between the categories of VA data use. Then we’ll look at data access requirements specific to these categories. Then finally, we’ll identify the best available resources for locating the data request process and getting access to the data you need for your study. We’ve added bonus slides as I’ve said. They include a list of commonly used acronyms related to VA data sources and data access and a few slides to give you an idea of what the data access request tracker system or DART looks like.

We’ll begin by looking at the general requirements for data access that apply to every data user. Who can access VA data? Access to VA data containing protected health information or PHI is restricted to VA researchers employed either full-time or part-time by the VA and all other VA employees of course. This includes WOCs or without compensation appointees and VA contractors doing work on behalf of the VA.

Access requirements for VA data sources vary by the proposed use, the sensitivity of the data for example, does it include identifiable information, the physical location of the data and requirements specific to the data steward or access manager responsible for the data.

For those of you new to VA data, you may be unfamiliar with the terms data stewards or data access manager. Here are some of their responsibilities that are related to data access. They manage the data access request process. They may provision data or notify others to provision data to approved users. Examples of national VHA data stewards and access managers include National Data Systems or NDS, Patient Care Services or PCS, Pharmacy Benefits Management or PBM, the VA Information Resource Center where I work at VIReC and the Medicare Analysis Center, MAC. There are others of course. These are just examples. There are also local data access managers at your facility. These include the Facility Health Information Management Office also know as HIMS and the Customer Users Provisioning or CUPS Point of Contact for AITC mainframe access.

One of the first questions that you need to answer when you plan a new project is what kind of project is it. This may seem simple but the distinctions in the VA can sometimes seem complicated. Today we’ll look at the distinctions between the three categories of data use. Research projects, preparatory to research data activities and healthcare operations. It is important here that these distinctions are based on the use of the data, why the data are needed and what kind of project they will be used for. The distinctions are not based on who the user is. In the VA it is not uncommon for research investigators to be called upon to work on projects at their facilities or with program offices that will use VA data for non-research purposes. The reason we’re discussing this today is because data access procedures are different for each use category. If a project is misclassified, which happens occasionally, it can be time consuming to correct the classification later.

The VA data access or use category is research if the project has an IRB approved protocol for conducting research intended to contribute to generalizable knowledge. That is to add to a body of scientific knowledge. It is also research if the project is funded in the VA as research. VA research projects must have a research protocol approved by their institutional review board and/or their local research and development committee. The protocol should include a description of the data the study will use. Being specific and clear in the protocol about all the data sources that your study will use and why they will be used will make your data request approval process go much more quickly and smoothly. Remember the most important part of the definition of research is that it is to contribute to generalizable knowledge or it is funded as research.

So when is data activity, that is the use of data, classified as preparatory to research? When you are preparing a research proposal or protocol and you need to look at VA data to determine whether your idea for a study is feasible, for example, whether there are enough individuals that meet your inclusion criteria, you may need to access VA data for preparatory to research data activities. Prep to research data access does not require an IRB approval. It’s the only kind of research access that doesn’t. Nor does it need a HIPAA waiver or authorization in order to look at PHI. Prep to research is not the same as a pilot study which requires an IRB or R&D committee approval. The words to remember here are preparing a research protocol. If that is what you’re doing, then your data activities and data access are categorized as prep to research. If you’re preparing a proposal for a non-research study, then your data use will not be prep to research.

The healthcare operations data access category applies to analytic or data work that is not research. So what kinds of data activities are included in operations? There’s program administration, which is data analysis for administration of or planning for a VA facility or program office. Quality improvement projects are projects that seek to improve the quality of a process or activity for the VA. This could be a clinical care process or an administrative activity. Projects that evaluate VA programs that are conducted for the VA are also operations activities. A significant phrase for operations is for VA or VHA. If the analysis or other data activity is to administer, improve or evaluate a VHA process, activity or program solely for the VA, then your data access category is operations.

Remember, these categories are based on the data use not on the person’s usual role. For example, there are times when a program office needs to conduct an evaluation of its activities and a research investigator is asked to conduct or assist in the evaluation under the sponsorship of the program office. In those instances, even though the individual is usually a researcher, for this project data access would be for operations. As an example of how this can be complicated, there are recent efforts by research service and HSR&D to encourage proposals that partner VA researchers with VA program offices for evaluation projects. Sometimes these projects include both operations and research components. If those planning such projects are not clear about which components are research and which are operations or how these components are to be distinctly managed, it might be a good idea to seek official guidance.

Which brings me to my next slide. When you’re planning a new project and you’re not certain whether it is or is not research, consider checking with your IRB, your facility research & development committee, the program office supervising or sponsoring the project, the Office of Research & Development or ORD and the Office of Research Oversight or ORO. ORO has indicated in several presentations that they really appreciate it when complex projects check with them before they submit a protocol to the IRB so they can assist in clarifying how the research component should be managed to prevent compliance issues later. A link to their website is provided here.

As we said at the beginning, we will address your questions at the end of the presentation. But if you have any questions about the material so far, please feel free to enter them in the Q&A tab at any time.

I hope the last section has given you a clearer understanding of how to determine which data access category your project fits into. But what does that mean for the data request process? Next, we’ll look at data request requirements specific to each of the three categories. Here are the three data access category buckets. Notice that the one on the left, research, has a higher stack of papers in it. Requirements for IRB approved research include the IRB and R&D committee approvals, including IRB approved research projects and include several types of information and documentation. You’ll see later that some of these requirements are similar for all types of access. The requirements for research include a description of the data needed, the approval of your supervisor and local officials, access forms for specific data sources such as CAPRI or vital status file and for identifies such as real SSNs. For research requests an additional set of approval documents is usually required. Depending on the data source you will be asked to submit some or all of these: the protocol, IRB and R&D committee approval letters and the informed consent and HIPAA documents.

Next, we’ll look for the requirements for prep to research activities. Before we list the requirements, you should know that sometimes you may not need data access to find out whether the available data can support your planned research. For example, both VIReC and VINCI provide data counts. VIReC for medicare and medicaid data and VINCI for CDW data from the corporate data warehouse. If you do need access to data, you will be asked to describe the data you need. Your supervisor and other local officials, for example the facility ISO may be required to approve the data access request before you submit it. Other requirement details will vary depending on the data source. And there may be other specific data access forms required.

But regardless of the data source, investigators planning to access data to prepare a research protocol are required by HIPAA to complete a memo describing their prep to research data activities and how they will comply with HIPAA requirements. There may be different prep to research memo forms or templates that you’re required to fill out.

But in all cases the following assurances must be included. The investigator affirms that the data access is to prepare a research protocol or for a similar purpose. Access to the protected health information or PHI being requested is necessary to prepare the protocol. The PHI will not be removed from the VHA and will not be retained by the investigator or in project files once the protocol is approved. That is, you may not save the PHI such as contact information for potential research subjects for later use. Once you protocol has been approved by the IRB, access to PHI data that you used during prep to research end. However, aggregate data, of course, may be retained.

Operations data access requirements include a description of the data needed, justifications for its use, the signature of the supervisor and ISO and specific access forms that vary by data source. Examples include the VHA NDS healthcare operations ePAS form and the patient care services data transfer agreement or DTA forms for program evaluation and quality improvement studies. You may recall that for approved research projects, many of the forms are submitted once for the project. But for operations data access, most data request forms are specific to the individual seeking permission to access data. Many people think that if a project intends to publish its findings, in for example a peer review journal, that it is automatically categorized as research. This is not true. If you have an operations project such as but not limited to a quality improvement study, you may publish your findings in peer review journals or other publications. We’ve included some examples here of quality improvement journals on the side. If you intend to publish the results of your operations project, you should complete the forms shown on the left of your slide. This template or sample format provided by ORO for documenting that you are publishing non-research findings. You can download this sample format from the ORO website. The link is shown at the bottom of the slide.

As I mentioned earlier, if you have any questions about this section, you can enter them in the Q&A box at any time.

Finally, we’ll talk about how to locate information that will lead you to the correct request process for your study. We’ll pause a moment to ask you about your familiarity with the VHA data portal. Have you visited the VHA data portal for information about data access? The responses are yes, frequently, browsed a little, no, never, not sure or not applicable to your role or job. Heidi.

Heidi: And responses are coming in. We’ll give everyone just a few more moments to respond before we close it out and go through the responses here. And it looks like things are slowing down. What we are seeing is 16% saying yes, frequently, 39% browsed a little, 38% no, never and 7% not sure or not applicable. Thank you everyone.

Linda Kok: Thanks Heidi. Well I’m glad that we have that 16% that have been there already. I think we’ll let you know a little bit more about how it relates to data access. Thank you. The VHA data portal is the right place to find information and the right place to start to find information about VA data. The portal is an essential resource for all VHA data users. It will be your home base for finding out about many aspects of VA data but especially about how to access data. A link to the data portal home page is provided at the bottom of the slide. You may want to save it to your favorites or you can always find it on VIReC’s home page. This close-up view allows us to see a few of the features of the VHA data portal. If you haven’t already, you should take some time to click on the many links provided.

The data sources tab links us to information about the most frequently used data sources. The data access tab links us to information about the data access request information for those data sources. If we hoover on the data access, it opens up a selection list and you can link to information for operations, prep to research or research. The other two choices listed here are for data access for those providing medial advisory opinions and for veteran service offices that assist veterans who need access to their medical records. We’re focusing on the first three options, operations, prep to research and research data access.

We’ll begin with research. This slide shows most of the research access webpage. The format is the same on operations and prep to research pages. On the top section, shows the type of access. In this case research. In the middle, there are shortcuts to some of the request processes. At the bottom, there are lists of data sources and tools. We’ll take a closer look at this on the next slide.

The most frequently used data sources are listed on this page. Data tools are also listed. Tools include systems for accessing data such as CAPRI and vistAWeb and CPRS as well as data processing platforms such as VINCI, SAS/GRID and the AITC Mainframe. We’re going to use the CDW data as an example and we’ll click on the plus sign to open the CDW data request process table.

This slide shows the request process table for data stored on the CDW. There are tables like this of various sizes for each of the data sources listed in the data source request process list, this page. Each column in the table represents an element that will help you find the right process for the data you need. The first column lists the data sources available on CDW. The second column includes the level of access, national, VISN or local. The third column lists the identifiers available for each data source. Shifting over to the far right, the sixth or last column includes a list of data access request processes. For our CDW domain example all combinations of these elements, data source, level of access and identifier lead to DART. As I mentioned before, once you know the request process that you need, you can go directly to the data access page to find the shortcut. Look for the requesting access section and use these shortcut links to go directly to the instruction page for that request process. We’ll talk about the DART request process next.

The DART process page includes several sections. Instructions for getting started, including a list of basic research documents, the research request forms that are used, resources such as a link to the DART user guide, which is very helpful to use whenever you’re new to submitting requests in DART, a launch DART button and a link to VINCI for help with the DART application at [VINCI@va.gov](mailto:VINCI@va.gov). A link to the portal’s DART data access page is provided at the bottom of the slide. We’ll focus on the instructions and the data request form.

This slide shows the DART data request tracker request process page. The section called requesting access includes instructions and one of several links on the page to launch the DART application. Lower on the page is a list of all of the downloadable request forms, potentially required for DART requests. You can see them a little bit more clearly in this one. Your project will need some but not all of these request documents. You can find out which forms you need for your specific project before you download them by launching DART and initiating a DART request. It doesn’t matter how soon before you submit a request that you initiate it. You can initiate a request and then slowly include all the information and build your request until it’s complete. In DART if you complete the basic project information and the data source selection, DART will generate a list of just the documents and forms required for your request. That way you don’t waste time. Then you come back here and you can download any of these forms and complete them and they will be ready to submit with your DART request.

There are five types of DART pages. There is request, where you enter the study name and IRB dates. Participants where project team members’ names and research sites are entered if you have a multi site study. The data source page which has several sections. There is a section on whether you’ll use the data on VINCI or a local network server or both, what identifiers you’re going to need, data sources to be requested and whether or not your project requires informed consent and HIPAA authorization documents. On the documents page, all of the required documents and forms for the data source are listed. On this page, you can upload each required document from your network server files to the associated folder in DART using a browse function. Finally, you submit your request for review.

To give you a peek at what DART looks like, I’ve included a few of the pages on the bonus slides at the end of the talk. There is a VINCI cyber seminar on using DART. We have the link shown here. That’s really a good way to get a full introduction into the DART application.

This flowchart shows what happens to a DART request for NDS managed data. The process begins at the top with the requestor uploading the required documents and submitting a DART request. The request is sent to the NDS administrator who will review it for completeness. Once the documents are complete, the NDS reviewer sends the request to several offices for review. Shown here are privacy, data security and ORD. The ORD is for real SSN access. If NDS or any reviewer has questions about the request or needs more information, you’ll receive a change request notification through VA e-mail. When all reviewers have reviewed the request, NDS performs a last review and gives its final approval. At that point, both you, the data requestor, and the data provider receive a notice that your request has been approved.

On this slide, you see some of the data sources that Marie Soden talked about in last month’s Database & Methods Cyber Seminar, an overview of VA data informations system and national databases and research uses. As you can see, all of these sources are requested through DART. This slide shows other commonly used data sources including PBM data that is requested through PBM’s data access process, the VA’s CMS data for research use of CMS data, medicare and medicaid data for research use is requested from VIReC and patient care services data are requested through the PCS data access process, the DTA process. Remember links to these processes are available on the VHA data portal data access pages.

We’ve got a few seconds. I’m just going to give you a quick look at the process for requesting pharmacy data from PBM. This slide shows the instructions. In summary, you mail a copy of the IRB approval letter and a summary of the protocol to this PBM address and complete the form that appears when you click the button here. And we have a link to this form shown at the bottom on the screen.

For CMS data, medicare and medicaid data for research, a request process can be found by looking on the VIReC webpage. You begin by linking to VA/CMS home and then the standard data request process, shown in the circle. Links for both VIReC and Mac are shown at the bottom of the slide. Just as a reminder, VIReC provides research access to medicare data and MAC provides operations use to medicare data.

The patient care services request process has DTA forms for quality improvement and program evaluation and research. There are links to these forms available on the PCS webpage. It’s a sharepoint site. Even enlarging the image doesn’t make it easy to read but there’s a link here to the sharepoint site at the bottom so you can visit the page directly.

This slide shows the operations access page on the VHA data portal. It includes the same list of data sources under associated request processes that we saw previously for research access. The most frequently used processes NDS healthcare operations request process which we are pointing to here, the direct link to the NDS operations access process, you click this link and it will take you to its second page which says create document and then you click on create document to display the ePAS form. This is part of the ePAS forms. It’s like two sections that are important.

Depending on the data sources that you select down here, there may be different tabs that appear at the top of the page. So, if you click on corporate data warehouse, a corporate data warehouse tab will appear at the top of the form. You’re basically putting in your information. You put in the active directory link or e-mail for your supervisor and your ISO at your site and that automatically sends this form to them for their electronic help signature. And you also need to select one of the uses of the data. It could be in most cases for this audience it would be healthcare operations. So these are close-ups of those sections. These are by name and personal information. Then the supervisor, the ISO, the purpose of use and the data sources.

This slide concludes of overview of VA data access request processes. The takeaways are this: there are three categories of data access and use in the VA—research, preparatory to research and operations. Operations is any data use that isn’t research or prep to research. It includes quality improvement and program evaluation projects. The VHA data portal is your best source of information about data access. If the data source you want to request is not listed there and you can’t locate it anywhere else through the links we’ve provided, please contact the VIReC help desk and we will help you.

That will start the question section. I’ll bring up my contact information. Thank you all very much for joining us today. Cheryl.

Cheryl: Well the questions are a bit slow coming in. We don’t have any.

Heidi: Cheryl, I’m sorry. I just realized it didn’t give you permission. We actually have several pending questions.

Cheryl: Okay. Heidi, would you like to do the questions?

Heidi: I will start from the top and then turn it over to Cheryl once we’ve gotten through the first one here. Okay. Can you provide a very concrete definition of generalizable? Is it simply published?

Linda Kok: No. Generalizable has to do with a scientific body of knowledge that is capable of being useful to other researchers doing research or to healthcare providers across the world. It’s generalizable meaning that it’s for everybody but operations work is generally for VHA or VA use, the primary audience and the effect of the findings of the project affect VA processes or VA activities and they aren’t really necessarily generalizable to the outside world. They might have some value to the outside world but that wasn’t the intent of the project. The intent of the project was to improve quality in the VHA or to examine, to evaluate the quality of a program or evaluate its usefulness. Generalizable is worldwide and that’s research. Research findings should be useful to the real world and operations to the VHA and the VHA processes.

Cheryl: Okay, we have another question. What would be a good definition for pilot study?

Linda Kok: It’s very tricky. I think the best thing I can say is you should talk to your IRB about whether your study is a pilot study. You know prep to research activities is really just a fact finding, learning about the data, finding out whether you have the data in sufficient numbers to answer your research question. Pilot studies are generally trying out the research project on a smaller scale or with some limits to it, maybe testing some forms that are going to be used in a survey. Things like that. I would check with the IRB and if the IRB is not helpful enough for you, I would call ORO. The people at ORO are really quite helpful in helping you figure out what it is you’ve got, what your project is.

Cheryl: Okay. Regarding the differences between operations and research, if for a nationwide project, can we ask central IRB instead of the facility IRB for clarification whether something is research or operations?

Linda Kok: You can contact the central IRB but I think I would go to ORO instead. I think it would just be an easier thing to do. The central IRB deals mostly with multi-site studies. If you have a multi-site study, the certainly the CIRB would be the best source.

Cheryl: Do most facilities have a facility IRB and facility research group? In my VISN it is only the VISN and we have very little support at the local level for research.

Linda Kok: Not every facility has an IRB, in fact many, many facilities use the university affiliate IRB as their IRB. If you need additional support, you could contact the CIRB or you could I mean depending on what the question is, you could contact ORD, the office of research & development and ask for help.

Cheryl: How long does it take to get requested data from VINCI? I submitted about four months ago, inquired by e-mail but now realize no response yet.

Linda Kok: Okay. There are some data requests that go through within a few days and there are some requests that take a little longer. In some cases, they can be quite long. It depends on the data source. Some of the ones that are taking a little longer sometimes are the CAPRI and vistAWeb. There’s a VINCI check box for surgical data, for the surgical quality data, and both the CAPRI and VistAWeb types of data are taking a little bit longer. Keep sending e-mails every two weeks if you are finding you have a delay. If you have a change request that’s sitting in your e-mail box, you need to fix that before your request can be processed.

Cheryl: How can we find out further details on each database?

Linda Kok: I’m not sure what you mean by details. For details on the data access part, you go to the data access pages for research or operations or prep to research and at the bottom of every one of those pages, there is a long list of the data sources and data tools. If you open and close those boxes, you can see what data request process is specific to that data source.

Cheryl: Another question. I’m a new research fellow and I wanted to ask whether the researcher needs to have a specific budget allocation or funding source for the data request or is the data request fulfillment part of general research support that is not charged back to the researcher? Is there a limit to the complexity and size of the data request in DART?

Linda Kok: Part of that question is no, there is no charge back to the project. This is part of VA operations to provide data access to researchers in the VA working on VA-approved research projects. This does not apply to university research but just to VA-approved research projects. And what was the second part of that?

Cheryl: The second part, is there a limit to the complexity and size of the data request in DART?

Linda Kok: There are some extremely large and extremely large projects going on. So there is no real limit but if you’re new, all of the research experts that I talk to say that you should start very small and understand what the data are and learn about how the data are organized before you jump into extremely large or complex studies. But no, there are no real limit. You cannot just go ask for all of the CDW data to be downloaded to your local server, but you can use it on VINCI pretty well. There’s a difference between the kinds of access that researchers get for CDW data and for many data sources including medicare. Research requests get data that are pulled specifically for that project, for that cohort of people in the research project. Operations data access is usually to full data sets and is not limited to a set of individual cohort members.

Cheryl: We have some more questions. I have been advised to request a support letter from VIReC for grant proposals using VHA data. Who should I approach for this?

Linda Kok: VIReC supports all research projects. We really don’t provide support letters for individual research projects. We’ve gotten this request several times. But you can point to our website [virec@va.gov](mailto:virec@va.gov) and you can pick up, I mean our mission statement or any other about pages there and just use that. VIReC is well known by the reviewers and they know that we support all VA research.

Cheryl: When we request access to data, what format does it come in? For example, do we get access to the entire database and need to learn how to run a report or will they send us a report with just the specific information requested?

Linda Kok: For the most part you never receive a report. You’re going to receive data. Now if you’re on the ACTI mainframe, you will have access to entire data sets, years of data. But if you’re looking at CDW data or medicare and medicaid data, you’re going to have extracts of data. The extracts for medicare and medicaid data will be a set of SAS files for the years for the people in your cohort. On the CDS side, you will receive the extracts of the tables from the sequel tables of CDW that fit your specific project. Once you get approved for CDW data, that data provisioner contacts you via e-mail and then you set up a correspondent site so you can e-mail back and forth inside this correspondent site. You explain what you need and then you come to an agreement with them about what tables are needed and what rows or subjects you want data for. Then the data provisioner will pull those things and put it into a VINCI workspace for you if you’re working on VINCI or they will put the data, all the tables, into a transfer server and you can download them to your own local VA network server. It depends partially on what data you’re looking for. But on CDW data, you’re going to get extracts and in medicare data, you’re going to get extracts. On AITC mainframe, you’re going to get whole files.

Cheryl: For research, I’ve been told the only the VINCI data manager can extract a cohort from \_\_\_\_\_ [00:52:28] CDW with a given finder field or cohort criteria and the researcher can then use that. This seems difficult as a cohort can change and be updated with time. Is this true? Would you like me to repeat the beginning part?

Linda Kok: No, it’s okay. VINCI data managers are assigned to provision data for research requests for data that’s on CDW that have been approved in DART. As I mentioned this communication box, it is becoming clearer to VINCI data managers that developing a research cohort and building that information of who is in your project, who your study subjects are going to be, is an iterative process and you will need to be communicating back and forth with the VINCI data managers. It may take more than one or two passes through the data to get what you actually need. Once you’re approved for access to data on CDW, that’s when you begin that process. On medicare data served by VIReC you have a pre-request consultation where you talk to a VINCI analyst and they help you determine which data you need. Then there is a very specific data description form that you submit. It’s a little bit easier there. The VINCI data managers are getting better and better. It’s been a few years and they’re doing a pretty good just. It just takes time and a lot of … iterative process.

Cheryl: How do we learn the data fields in each data source? Is there data schema documentation?

Linda Kok: Oh yes. I haven’t mentioned it once during the entire presentation but the VIReC website has detailed information about the data sources. On the CDW data you can go to our website. You go to data sources CDW and then you can look at the documentation pages and there is an up-to-date list of all of the domains and all of the tables in them and all the data elements in them. There’s a lot of other information there. I’m not sure if I could go off and do a demo but go to the VIReC website and go to data sources CDW and documentation and you’ll find what you need.

Cheryl: We have two more questions right now. Does the size of the average national data set require high-powered computer resources or can most projects be handled on a local desktop or the sequel server webspace?

Linda Kok: This is a really good question. You’re never going to analyze data on your local desktop. It is not permitted. You’re going to analyze data either on CDW, in the VINCI workspaces, on the SASGRID at VINCI or you’re going to process your data on your local VA network server that is maintained and managed by OI&T, Office of information and technology. You never ever put any data on your desktop that has any PHI on it. You can also process data on the AITC mainframe at \_\_\_\_\_ [00:56:23] if you like to use job control language, IBM format.

Cheryl: We have a facility data support group to assist in most facility level data requests. However, we’ve recently been asked for further national level data not really in our group’s scope of work. Would these requests get from VINCI by req or other national groups? Most of these requests come from people who do not know how to extract data from CDW.

Linda Kok: Well I think most people that are working in the VHA don’t know how to extract data from CDW. It’s a very complex transactional system. You need to understand sequel. By the way, VIReC also has CDS cyber seminars, some of which are stored in our archive and on our website that will introduce you to the CDW and how to use it. They’re excellent cyber seminars and at the most basic, brand-new, beginner level so I would definitely recommend that all of you take a look at those cyber seminars. Without knowing more about what data you need, I can’t really tell you where you should go for it but if you go to the VHA data portal, data access pages, I think that would help you understand where data can be requested from.

Cheryl: The person who asked the question about using the local desktop actually amended the question to be does the size of the average national data set require high-powered resources or can they be handled on the local server?

Linda Kok: It depends on how big your server space is but there are lots of people doing national studies on their local facility server or in their regional server. A lot of facilities are losing their servers and t they’re going up to the OI&T regional level now. It’s changing across the country so there aren’t local facility servers everywhere anymore.

Cheryl: Well that was our last question.

Linda Kok: Well we timed that right.

Cheryl: We sure did. Well thank you Linda for taking time to present today’s session. To the audience, if your questions were not addressed during this presentation, you can contact Linda directly. You can also contact the VIReC help desk at [virec@va.gov](mailto:virec@va.gov). We have on your screen now a list of some additional VIReC resources to help you. Our next session for our Database & Methods Series is December 5, 2016 at 1 p.m. eastern. The session will cover healthcare utilization using VA data and will be presented by Peter Richardson of the Center for Innovations in Quality, Effectiveness and Safety, IQuest in Houston, Texas. We hope you can join us. Thank you once again for attending the session. Heidi will post the evaluation shortly. Please take a minute to answer those questions. Thank you.

Linda Kok: Thank you, Cheryl. Thank you, Heidi.

Heidi: Thank you everyone. For the audience, as Cheryl just said, when I close the meeting out you will be prompted with a feedback form. Please take a few moments to fill that out. Thank you everyone for joining us for today's HSR&D cyber seminar and we look forward to seeing you at a future session. Thank you.