Cyber Seminar Transcript  
Date: 9/24/2015

Series: VIReC Good Data Practices  
Session: Using RedCAP: Data management in studies linking primary and secondary data

Presenter: Stephen Deppen  
 *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)*.*

Moderator: Hello and welcome to VIReC cyber seminar series; good day to practice this. The purpose of this series is to present issues related to good data practices and to discuss examples of how these practices are applied to VA research. Before we begin I want to take a moment to acknowledge and thank some of those who have contributed to the series. We have an active informal advisory group that helps guide the choice of topics and the concepts covered in each session based on suggestions from participants like you in previous years. This year’s advisory group included Matt Masacheski (ph), Peter Groveville, Jennifer Garvin and Jim [Inaudible 00:00:43]. If you missed any of the good data practices cyber seminar series as Heidi was just discussing you can access them through the VIReC cyber seminar archive page at the URL shown on this side. You can paste this URL into your browser and then filter for good data practices. It HSR & D archives exist.

This year’s good data practices series as Heidi said has four sessions presented on Tuesday’s and Thursday’s last week and this. Last Tuesday Sara Krine (ph) in her session planning for data early and often discussed the benefits of making decisions on data early. Last Thursday Todd Wagner discussed approaches he used to handle data limitations and data linkages in his session entitled *Mind the Gap* using administrative and claims data to answer your research question. On Tuesday [Inaudible 00:01:54] session entitled *Decisions, Decisions, Decisions* and she talks about how to select methods and analytic tools for research, health services research.

Today is the final session for this year’s good data practices [Inaudible 00:02:13] Dr. Stephen Deppen will tell us about REDCap data collections and management applications. And it’s used for longitudinal studies and studies that combine primary and secondary data in his session entitled *Using REDCap Data Management and Studies linking primary and secondary data.* Dr. Deppen is a clinical epidemiologist with experience in the evaluation diagnostic test for cancer and health services research and quality improvement. His research focuses on improving the diagnosis, treatment and outcomes of cancer and evaluating the efficacy of cancer treatments with specific interests in lung cancer. He is an assistant professor in the thoracic surgery department at Vanderbilt University Medical Center and is investigating how benign fungal diseases affect imaging for lung cancer screening and diagnosis. He is the lead author on a number of influenza manuscripts including the accuracy of FDG TET to diagnose lung cancer in areas with infectious disease lung disease, a meta-analysis published in Jama.

As a reminder any questions you submit will be monitored during the talk and I will present them to Stephen at the end of the session. Also at the end of the session a brief evaluation will pop up when Heidi closes out the session. Please, please stay until the very end and take a few minutes to complete it. We really value your input on topics for next year.

And now I am pleased to welcome Dr. Stephen Deppen.

Dr. Stephen Deppen: Thank you Linda and thanks everyone for showing up. We’ll go ahead and get started.

So as was mentioned I’m an epidemiologist and I was one of the original folks to be approached to how REDCap should be set up. And so I wanted to first [Inaudible 00:04:25] as well as the [Inaudible 00:04:26] and the Translation Science Award as well as my collaborators.

So there’s a poll going out what is your role in research and if you could designate which role you choose. I’ll give you just a minute.

Unidentified female: Yep responses are coming in and we’ll give you all just a few more seconds before I close it out and go through the results here. Okay I’m going to shut things down and what we are seeing is 28% research investigator, 39% data manager or analyst, 25% project coordinator, 8% other and there we are seeing research informatics director, IRB administrator, statistician and laboratory. Thank you everyone for participating.

Dr. Stephen Deppen: Great. And a second question is how many years of experience you’ve had working with VA data.

Unidentified female: And our options here are less than one, one to two, three to six or more than seven. Give everyone again, just a few more moments here. Okay it looks like we’ve slowed down. So what we are seeing is 44% saying less than one year, 9% saying one to two years, 28% three to six years and 19% seven or more years. Thank you everyone for participating.

Dr. Stephen Deppen: By the end of this webinar you should be familiarized with some of the basic capabilities of REDCap. Now this is not intended to train you, to make you a user. I just want to kind of wet your whistle and get you interested or thinking about REDCap as your choice when you know that you’re going to be collecting data. I’m also going to be presenting some of the methods of combining primary and secondary data for the capture for research as well as just give you some resources and examples when you start using REDCap.

So REDCap was developed by the REDCap consortium, developed at Vanderbilt to CTSA and the driving purpose behind REDCap was to get essentially IT out of the database creation business and my previous job before being – getting a PhD was actually being one of those IT guys. And the basis of REDCap was to create a method to generate databases and surveys which are technically easy, secure, IRB compatible and compliant and are responsive so you can do a rapid cycle change and not have to, I know personally you get an idea, you have a set of questions, you type them up, you present them to someone who creates a database. A week later you find out there’s issues, you go back to them, it’s another week and that airative process draws out your research at nauseum. But with REDCap no programming experience is specifically needed.

So three of the sort of basic things that REDCap can do which are interest is it handles a wide variety of data. And also because it’s an open source it’s a very broad application is that it can take those data and different combination permutations for a wide range of applications; so it has actually some more data capture capabilities than you would think of say than if you’ve done survey monkey or the like.

The other aspect, and the back end aspect which is very important is it does an excellent job of securing access controlling groups, being able to define groups and individual roles which then defines their access to the data as well as activity tracking. This makes it extremely useful both from a single institution standpoint as well as from a multi-institutional setting. Because many, over 3,000 other institutions actually have REDCap locally so you can either create a situation where a copier database is run locally at the other institution which is part of the IRB for your multi-institutional trial or they submit their data in a comprehensive coated way so it immediately goes into your existing database which is centralized – if you’re say a central data repository.

And finally as broad base or reporting capabilities and for me as someone who does analysis I really enjoy the fact that it exports data with the ability to fully de-identifying that export and to be directly pulled into the most common statistical packages or as a simple flat file for archival purposes.

So here we see just kind of the home page after you decide to set up a project and the other nice thing about REDCap is it really walks you through the process of creating your own database. It gives you – it asks what is the purpose, do you need some specialized aspects like is this longitudinal, is this the survey and then are there specific nodules that you need? Do you need auto numbering randomization and if you are using longitudinal data or if you wish to create say email generated activity. So if one of your participants requires to have a separate survey 30 days into or 30 days after their consent, it can actually generate an email that says these people need to do form one to your study coordinator.

Second and third aspect is to test your project thoroughly and once you’ve done that you move to production and finally once you have your data production you like the way it looks, you then assign rights and permissions. And we’ll go through each of these steps briefly but the focus is going to be more on – and there will be a demonstration towards the end where you can ask some of your questions.

So an easy modification, and again we’ll do this a little bit more in the demonstration you create forms and each form can have different sections. And as you see here there are two forms. It says Z4031 and this is actually the randomization criteria to see whether or not the patient was eligible for this study. And then you see the nodule study which is a separate form which once they’ve been randomized they then fill out the secondary aspect. And these might be decided upon by an outside entity, in this case ASCO or it could be this you internally generate and define.

The other strong aspect of this is what you see here it’s in an Excel spreadsheet is the actual code that occurs behind the scenes. And the expectation is not viewed to be able to do this or generate this code; it does that automatically with an online web designer. But because this code is there you can then take this, email it to anyone else in the world who has REDCap up and running, they can pull that data in, create a project and their project – their database will look identical to yours. And I’ll go through some of the ramifications of that, especially for multi-center trials in the future towards the end of the study – towards the end of the talk.

So the second significant aspect of this is the security aspect and the first step of security is access. So once you have individuals who are defined at your institution when this case met the VA, once you get access to REDCap, to request access they’ll give you a user ID and so forth. And then you’ll be in the system. I create a database, I find you in the system and then I grant you access at the level that I choose and an email goes to you that says “You are now a member of this”. You – it tells you where to go, what that is and when you pull up REDCap all databases that you have access to show up. So the second aspect of this is you grant access and then you define the roles, you define the security. So for example if we have a multi-center trial and you have data coordinators at each center the data coordinators might be the individuals who are able to see all data while the PI’s at each center will not be able to see all data. So if you define individual roles and also can define groups so Hopkins is one group with four participants of which one is the data coordinator.

The other aspect of it is that it features a detailed audit trail and here you see the date that the database was manipulated, who did it, that’s me and what was – and you also see exports and what data was exported and every time a field is changed it will give you this level of information. You can see a record number of 630A for example, the ID was updated.

So here you see this is what the user rights and access have looks like. It gives you just a brief idea of some of the aspects that you can control and manipulate so you can define, as I mentioned earlier you can define the group and assign roles. You determine expiration dates so if you have a temp that’s doing some data entry you can – and they leave Friday the 16th you can’t cut it off that date. They don’t have access after that date, you can also see the different aspects of the database and things are green at the top, one of the data users and then Melissa Protter is one our data coordinators and you see what level of access as well as exporting and importing [Inaudible 00:14:58]

So the other aspect of this that I like is the report creation. And so the features of report [Inaudible 00:15:09] are general so once you have created your database it will create a pdf of each of the questions, all the values and all of the – what the tool looks like in a pdf form, which can then be added to your IRB submission and IRB will be able to see both what the questions are, what the inclusion/exclusion criteria are and it’s also useful if you have someone who is requiring to use paper to do your data entry. And then entering into the database later, this is also a way to collect that data and purely paper form. There’s also established canned reporting which is primarily focused on internal quality assurance as well as some basic external review, values and summation of values and some basic graphing. With respect to REDCap at the VA, it’s behind Davinci Firewall. As I mentioned before you have these various access controls and it’s easy to create deidentified data as well as archiving that data.

So having done a kind of brief introduction I want to talk now about some specific examples of how REDCap has been used and how I’ve used REDCap in capturing data. So the first example is going to be our treat lung cancer model, which is a predictive model. It was initially a single institution, retrospective chart review. After publication we went to a multi-center data collection which had three sites in other states, one in VA and one international site. The populations between these five sites was very different. [Inaudible 00:17:03] population one, pulmonary clinic in one and surgical patients in the other and meta-data if you remember back to that Excel spreadsheet earlier that had the actual code. The meta-data is what that sheet is and that’s how the impressions are defined, the variables are defined and relationships as well as any sort of guidance as to how to answer a specific question. So that meta-data was used very differently for each one of these sites.

So for example here we have what I call the top part of the form which is some of the basic demographics and we have an institutional ID, they’re missing HMP but basic coded data that you would expect say from your enterprise data warehouse, from the CEW or from some coded environment where you have that. So age, gender, race and in this case height and weight and in red you see the height and weight are used for a calculated field. Some institutions don’t have height and weight but they have BMI and that’s available on a separate line. And again each of these is list as a common to limited plat file and they were then imported into this database. We have the other aspects of the same abstraction tool and this case this is the data abstract or the clinician will take that information from the top, go into the electronic medical record, go into the say imaging module [Inaudible 00:18:41] and abstract the rest of the data collection tool. You see these are examples of some results, FEV-1 as well as results of looking specifically at some CT scans as well as a PET scan.

So our second example is a meta-analysis and the reason why I wanted to choose this is very similar to the first but just to give you an idea of how broad your data import can be. So here the data that we imported was from an internal data source and that data source had or has a very common data model, something we use every day pub med (ph) and there’s slight differences between some of these data sets. Pub med has one that – and some of the others, especially the web of science and gray’s literature has different underlying data structure. We also have data entry by two different abstractors and the database is set up to support systematic collection as well as expert analysis and as you may be able to guess meta-analysis. So the importance of reproducible research in any systematic review is paramount that the data needs to be able to – the methods and the data need to be able to hand out to anyone for them to replicate our results.

So here we have an example of the pub med import and if you look the study ID is actually the PMIV, so a clinician can be cover with that, cut and paste it and put that directly into pub med and bring in the pub med reference. So if they want it for the article they can do that directly. You see the full end note citation, they actually took this from pub med, use the XML design from pub med, put it into end note in this case which is what we’re using and then from end note we actually exported it into a flat file directly into this database. You see the [Inaudible 00:20:37] that abstract – full abstract of that [Inaudible 00:20:43] some other coded here and then finally there’s an example of a part of the review form and these are – if you’re familiar with structured systematic review which is the quality and outcomes of the – of an individual study to measure the quality of the study that was performed is our – is standardized questions. We adapted for our purposes and if you look underneath each one of the – in the blue underneath each one of the response you see some guidance as to how we are defining that as specific to this study.

And finally you get an idea of the export of the data in this case just a flat file XL spreadsheet that pulls out all the data that we’re interested in and you can see across the top we have that coded data through the first author and then a decision whether or not to include or exclude agreement – what level of agreement was there between two reviewers showing that there was disagreement and then what exclusion of criteria were used.

And finally I want to talk about a longitudinal study. Now longitudinal studies are different from say a single source or single side or one off sort of database because in these instances our participant can either have multiple encounters so you can think of someone who goes into the drug trial and they get randomized to one arm and they might, if it’s say chemotherapy drug, they might have induction over the course of weeks or months and they might – we want one to be measuring some physiologic as well as quality of life measurements at each time. So because of this we can have one participant that can be encounter multiple forms across time and furthermore we can use multiple forms for one encounter; so we might – if on visit three we might include their vital statistics, the drug that was included reactions and then [Inaudible 00:22:46] as well as the patient reported quality of life.

All those four forms are collected into one encounter and we would track that patient across encounters and those forms might vary based upon different encounters. So it’s highly – it’s a very flexible in how you want to collect data from that aspect. And so some examples and you can have the class [Inaudible 00:23:10] where you do a baseline huge questionnaire of different diseases and so on and then you do – periodically you will go back to the individual participant, say they were a diabetic and someone is doing diabetes research and you’ll go back and you may do a diabetic specific questionnaire or set of measurements, getting the last four HB1C values or something similar.

And so it can also be used on the QA side, the National Neurosurgical Outcomes Database has over 200 individual surgeons who are recording their results into a common database and I – there’s 70 plus sites I believe. At any rate that gives you an example of just some of the breadth of that it can be used.

So here’s an example of the front page for the longitudinal study and just like with single side study or one off study REDCap walks you through each step of the way of creating and designing your data capture tool. So in this case you would choose a longitudinal collection tool and then with these collection tools you gather different forms into events. So this is an example of a form, this case it’s an example and enrollment is an event. Enrollment as an event might have consent, SF36 with randomization result [Inaudible 00:24:46] randomization and initial results. So those three or four forms might be [Inaudible 00:24:53] with the enrollment of that – other different events which mean coded and defined. And this is an example of an individual participant view so now we’re looking at patient number one and you see that they were enrolled and which forms are filled out, so green means it’s complete, with yellow being unverified. You also see that visit one for some reason we’re missing our initial lab data. It’s a great way of figuring out what is missing for that individual participant.

And then here we have an example which lists all records which is – it keeps going on [Inaudible 00:25:32] and you get a quick idea as to relative compliance [Inaudible 00:25:37] purports off this as well. So if you go and look at [Inaudible 00:25:43] data for a specific person again does a good job of orienting you as to what you’re doing with and whom you are dealing with. In this case we have a participant and we’re looking at event one and event one if you look over on the left you see event that has four different forms associated with it and the form that we’re looking at right now, which is highlighted is the patient morale questionnaire. And you can see the form status, which is actually – that should be a little bit lower and that should be the complete – looks like and then where you are with respect to patient and over longitudinal make up.

The longitudinal study, like other studies have a set of essentially canned reports and so you can see here the classic missing values reports and missing values of any values or required fields, validation, outliers is it’s someone entering data beyond the data that was each – you can limit the data ranges but as it being a warning and alert that this is out of range or you can force an individual who is doing data entry that has to lie within a range. This just gives you some example of all again, that are available for you day one.

So now I want to spend some time looking at reporting and research aspect. And to do that you pretty much have to do a demonstration; so first I want to talk and look at some of the basic exporting. I want to get to that eventually. So people might not be as familiar with this but it’s one of the more powerful aspects of REDCap. So just give me a minute.

And so here we have the initial tab is the list of your databases that you are an active member or participant in. Here is the status – the status is red, that’s inactive. Being active and in production and here it’s in development; so we’re just going to look at one that’s in development. So again here we have that initial screen, we’re going to go in and add new records so you can see and here it gives you an example of just what these screens look like and each of these are defined and there’s branching logic available. You can see that if the diagnosis is benign for example, we want the benign results. If malignant we’re talking about it’s in the staging and then finally if we’re finished with our data entry, complete and continue. And if there are any required fields that will bring that up. So again that makes very easy for having good, clean data. Okay.

So the other aspect I wanted to show you was some report writing. First – go back to that meta-analysis. We have fortunately no PHI of the meta-analysis. You can see – here’s a report that was created looking at each one of the articles that requires further review. This was defined as a filter field. You can do a fairly sophisticated filtering system as well as saying what it is you wish to report out. I want to export this to hand off to the statistician, I want everything. And here you see the variety of the packages you can export to; we’ll do R. And on the right here there’s some really important points I wanted to make.

So when we talk about the security aspect of it as you are making your database you can define different fields as being PHI and therefore you want to restrict those from being viewed or from being exported. And here these are some examples of what REDCap can do in order to deal with the deidentification question. So it can remove all identifier fields based upon how you define them. It can hash things like record field, let’s say that’s a Social Security Number, you can hash that. It can move – we don’t have any dates here but we can also – we can remove all dates or we can shift the dates as specific amount so let’s say you’re dong time to event data that amount will be shifted consistently for each record; so your time to event analysis will be correct just the dates will not be right. The specific dates will not be right.

And this is a fair amount of data so it might take a minute. So export to R and one of the important things is we have all that coded data or we have all that code from the meta-data behind it and what REDCap does is it takes all that meta-data and it uses it to generate all of this R code and it does the same thing for each one of the statistical packages; so that now each one of those variables is defined and you have the full value; so prospective cohort, retrospective cohort and so you can see that and to say the least it would take me a significant amount of time to code all that up to hand off to my [Inaudible 00:32:42] so really the time saver once you’ve done your work. And it also makes life easier for the statistician who is trying to work with data that they may not be familiar with. And then we also have [Inaudible 00:32:51] data, which I can pull up with just a flat file.

So it just gives you an idea of some of the reporting. And some of the other capabilities that I really haven’t talked about or spent a lot of time with, there’s a randomization module, there’s multi-side studies you can do. We’ll give some examples of that. So going back to how we did our multi-site with that meta-data the VA it was actually done before REDCap was available at the VA. So the meta-data was used to create an XL spreadsheet with verification tabs so that the data say a male is one and a female is two, so that data is coded within a VA, is coded in a way that it will be able to be immediately pulled into REDCap database. Second the meta-data was sent off to Australia. They have REDCap at local universities and a graduate student just basically pulled it up, day one we talked about it. Talked through what their data looks like because it was on a separate flat file, it had a map but their existing data was to the new data scheme. And then here we had it in two different places here at Vanderbilt from the same meta-data.

There are a number of really, really good resources. If you’ve not used REDCap before there are a number of videos which are available under REDCap website, I strongly encourage you to use them. I actually use them frequently especially when I go back and brush up on my chops, something I’ve not done very often. Like randomization [Inaudible 00:34:42]. And going back if we go into project set up you can see here I mention that meta-data but you can also have an online designer. So as I mentioned you do not have to be a programmer in order to use REDCap; so here we have an existing data base and I can easily change the different values and what that data and that question looks like because well as adding annotation or help notes etc. And then after I do the changes here it will go in, be reviewed and show up magically in that meta-data which then can be used in spread. REDCap also has a large library – let me find it real quick. But anyway there’s a large library of other studies and other databases that have been used and those can be cut and paste, you can see here. So if for example I wanted to put an apache score at the bottom or in the middle of my data collection tool there it is, I can view it. And that’s what it would look like. And then I can pull that directly in. I can do that right now because it’s in production. But it gives you an idea what you can do and see the BRFS and a number of these tools are also available in Spanish.

So that’s the end of the presentation; I will be happy to take any questions.

Unidentified female: Okay I have several questions. Can you hear me okay?

Dr. Stephen Deppen: Yes I can.

Unidentified female: Great. The first one is REDCap open source and I think I know the answer to that but can you also explain how it exists in the VA?

Dr. Stephen Deppen: So [Inaudible 00:37:37] be able to perhaps better answer some of those questions with respect to VA; so REDCap is housed [Inaudible 00:37:45] and it is administered by ViREC and it is available behind the VA firewall; so it looks for the VA demand and you can access REDCap from any web browser anywhere in the VA. So that’s the open aspect of that. It’s open source in the other aspect outside the VA is that you have what’s called the REDCap consortium and the consortium is basically for a minimal fee, a few thousand dollars you become a member of the consortium and you run a local instance of REDCap and it is open source in that you have thousands of these local instances occurring with individual researchers delivering data to databases and able to share those across the board.

Unidentified female: Thank you. The next question is kind of the same question two different ways. One of them is are there any limits on who can be surveyed using REDCap? Can veterans be surveyed, can employees be surveyed?

Dr. Stephen Deppen: Good question; so that is going to be more of an IRB question. So REDCap is used in two different ways. One is classic research, that’s what you saw here. Some of the things you’ve not seen because it had PHI is particularly for a QA, so an activity associated with clinicians, associated with patients and that instance because it’s a QA, we’re not doing directly per say surveying patients. But on the research side if the – your IRB would really have to answer those sorts of questions. And let me just say from previous experience is that once your local IRB gets used to working with and seeing REDCap the expectation will be [Inaudible 00:40:00] that REDCap is going to be your method of acquiring data because of all the security aspects behind it. I’ve been on the negative end of that where I’ve had to essentially shut down a study because of security breach and it literally took me seconds to shut it down and essentially three or four hours to figure out what the issue was and that it wasn’t related to REDCap. It was actually related to someone printing it out in Excel spreadsheet and handing it out. So what I found is that IRB’s once they get used to seeing REDCap being in use per research you have to jump through more hoops when you decide not to use REDCap or when you have a relationship with an outside entity that is not using REDCap.

Unidentified female: Thank you Dr. Deppen. The next is similarly is it possible for the participants in this study to take the survey outside the VA firewall? For example VA research projects send a survey link out to a veterans email for him to take it at home?

Dt. Stephen Deppen: As far as I know the survey module only works behind the VA firewall.

Unidentified female: Thank you. The next question what back end database is used by REDCap?

Dr. Stephen Deppen: I believe it’s my sequel.

Unidentified female: Okay these are going pretty quick. This one might take a little longer. How do you import CDW data into REDCap?

Dr. Stephen Deppen: I knew I was going to get that one. Give me a minute. So there is a – the short answer is there is an import tool so import tool. And you decide how your data is resulting. Are they in rows or in columns? And there it is. So you do is let’s say I’ve got – I have – if you look for the first four or five columns here I will take my patient data with the zip code, the living status, [Inaudible 00:42:27] and so this would be coded in that I would get from the CDW and what I would want to do is make sure either the database that I set up because I know I was going to use this data; however that data shows up in the CDW would make it appear the same. So let’s say gender is a 01 with 0 being male. I would make sure that gender is defined in my database the same way. Let’s say I’m using the US census definition of race and ethnicity, 0, 1, 2, 3, 4, 5 I would make sure it matches whatever that data structure is. So as you make your database if you go through those steps that end piece becomes really easy.

Unidentified female: So column names would have to match what the imported data base [Inaudible 00:43:16]

Dr. Stephen Deppen: They don’t; so what I will do is – I will either import the CDW data and then I will rename the column names into a kind of staging area to what REDCap requires. So the MRN for example, that’s something I define but let’s say last name is something that’s specific to REDCap. I would in that staging area I would change the CDW data to last name for that appropriate column.

Unidentified female: So are you mapping each of your columns to a column in the CDW data?

Dr. Stephen Deppen: Correct.

Unidentified female: All right and then matching all of the coding structure?

Dr. Stephen Deppen: Right, right. And the other thing which I’ve not shown here is actually there’s actually an importing tool and when things show up that don’t match it gives you a detailed record by record sell by sell report of that; so you can go back and invariably especially when you have a large database, invariably you make mistakes on one side or the other and you can then go in and fix all that.

Unidentified female: Thank you. Okay I think you talked about this a little bit with the library that maybe you could explain how people do it a little bit more. Are there projects already populated in REDCap or do you have to start from the ground up with each project?

Dr. Stephen Deppen: So the – by populated do you – if you mean that there’s live data I mean there are existing data sets out there that are live. But if you’re talking about data structure the library gives you those data structures so for example you saw the template – the template that was in the presentation was this one. If I go to share data library and I’m just going to go ahead and pull in the bartel index so import this. I’m just going to add it and now I go back and look at this instrument – here’s that data. See it here on the other side and there it is. So –

Unidentified female: The structure and everything else can be copied over.

Dr. Stephen Deppen: Correct, correct so if you’re trying to copy a database as well as a data set then you would import your data set.

Unidentified female: So the data comes into import, the library is one way that you can import the structure?

Dr. Stephen Deppen: Correct, correct. And once you import it you can then easily modify it, okay?

Unidentified female: Great, thanks. If you create your own survey or data collection tool and you don’t want anybody else to change it unless you share it with them particularly can you limit modification of your tools?

Dr. Stephen Deppen: Yes. Excellent question and that’s under user rights. So here you add user with custom rights and here you see the list of rights that are allowed. So there’s specific data, I’m not going to – only if this is checked do they – are they able to do any sort of designer set up; so I’m not going to give them access, I’m not going to give them user – they can’t change users. They can only export deidentified data or I can say they can’t do any exports, all right?

Unidentified female: You could change the settings if you wanted to send it to somebody privately.

Dr. Stephen Deppen: That’s correct and I can even change it based on just overall, I can even limit each one of the forms that they see.

Unidentified female: All right, great. I have another – I’ve got several more. Please explain what hashing a field means.

Dr. Stephen Deppen: What hashing a field means. I’m not entirely sure what you mean by that.

Unidentified female: They want to know that it’s crypted –

Dr. Stephen Deppen: Oh hashing field, right, right, right. Okay so let’s see if I can think of a data base that don’t have PHI that – so hashing a field would be so for example have a social security number, okay? So hashing a field would be you can have it completely hashed in that it no longer has any relationship to the previous values; so it is completely a random number is added to randomly to each social security number; so that’s one way of doing it. So basically it’s useless.

The other one is and similar with the date is every social security number could have a specific number added to it; so for example if you wanted – and in some cases the researcher would know what that number – 3,000 point or 62 is added to each or script is added to each and here’s the script so you can then deidentify – you can reidentify it if necessary. Generally speaking especially social security numbers you do a strong encryption; you do the full randomization. So it’s – dates it’s usually – not usually, you either hash it in which case each date value, each individual so there’s eight values, month, day and year – four digit year. So each one of those numbers becomes a random number so that’s what’s called fully hashed. Or there’s some other aspects of that as well. Basically you’re making it a useless number or you can move each number a specific value of 0 to 36, 365 excuse me.

Unidentified female: Sort of data shifting.

Dr. Stephen Deppen: That’s correct, data shifting. So those are each ways of dealing with deidentifying assuming you still want to report it.

Unidentified female: Great, okay.

Dr. Stephen Deppen: There are others as well that you can choose. But that’s just – those are some examples.

Unidentified female: All right, this question says I saw sequel queries being used in the calculations field in the data dictionary, how is that done? Can you refer me to any documentation on how to do it?

Dr. Stephen Deppen: I – that’s the advantage of living at Vanderbilt I actually wanted them to calculate something – doing the calculation is easy it’s just a calculation between two things and the SQL was the equation and that is again automatically done as part of the design view. So I will type in the equation for BMI and said you’re playing from – and then I define each of the variables that went into that equation and so that is part of – you do that as part of any time you define a function within it. There’s also – they separate SQL function which is if you type in the first letter of – or the first number of say an ICD-10 code it will do that. And for the ICD-10 code piece that’s actually a – I don’t know if that is a current feature or a future feature for REDCap but that is something that is available in the version of REDCap that I’m using that is at Vanderbilt. So I’m not sure if that’s been pushed off to the VA yet.

Unidentified female: Okay this one is a little different; do you need permission from the developers of tools such as the BRSSS in the library to use them?

Dr. Stephen Deppen: Some of them – if you look at that. I just went through a quick because Apache and the reason why I chose Apache is that it’s public. For some of the others it does require that if you do any reporting out and there’s also a quick agreement piece as part of that. Most of those that are – there are a couple that I no longer see like the beck, which [Inaudible 00:52:32] not there anymore.

Unidentified female: Okay so if it’s a permission thing it will show up on the things that you have to agree to it?

Dr. Stephen Deppen: That’s correct; I agree not to by like the copyright, etc.

Unidentified female: Okay. So REDCap or can REDCap handle genetic data?

Dr. Stephen Deppen: Okay – kind of. So another study that I’ve done and actually one of these studies so we have patient demographics, we had – we imported ICD-9 codes, we also imported [Inaudible 00:53:15] codes and CMS data as well to determine the quality of care. And then separate from that each one of these patients also had an exon shifts (ph) done; so the exon shifts in combination were over a terabyte of data. So no, importing that level of data into REDCap, would that work? But separately we used a set of genetic programs and a set of our scripts to essentially scroll down our interesting results in the genetic side. So you know this person have these snips, was the add mixture of the individual across the chip or so on and essentially half a dozen fields of coded data from the genetic data was then pulled in. So directly I don’t think so. Ours wasn’t just because it was more efficient, it was not the right tool to do it in. The long/short answer.

Another aspect of it is however for anyone of those questions when I talk about – when I mention you know we can give – you can give guidance to the bottom, you can also link in a pdf or a hot link so you can actually say here the guidance as to the definition of this thing. Or you can bring in a pdf which gives you the description if it is something that’s – or say the Apache score documentation can be brought in.

Unidentified female: Okay. Great. And a follow up to that he says so four levels of data that you might have, a person, the host. The host genotype the visits and the virus genotype.

Dr. Stephen Deppen: Right, right. So those might be – those might be categorical variables or in combination or you might even be listening specific proteins or a combination score based upon some other things. So those – you can think of it as I don’t want to combine that data and this is a good repository for certain things but I think especially when you look at base genetic data it probably is not the right tool for that.

Unidentified female: Okay good, the limitation about veterans taking the survey they have to be behind the VA firewall, it is this a rule at the national level or is it a local level? It’s national –

Dr. Stephen Deppen: I think it’s national and I’m not the one – I don’t know the answer to that.

Unidentified female: Yeah it’s in Davinci so therefore it stays inside the VA firewall. Okay it’s a national level restriction. Could you say more about using REDCap for systematic review?

Dr. Stephen Deppen: So if you actually if you go to the supplemental junk of that jama paper we kind of go at nauseum into that, that was such – is it because you were going, we were trying to change the guidelines. So when you think about systematic review. You go through your picos so decide what your population impact and so on. And then you work with a reference librarian and let me just say with respect to the meta-analysis I did not pull that data. I didn’t pull the [Inaudible 00:57:01] and I didn’t get the coding. She helped me greatly to get it into a structured coded fashion and once I saw that then I knew it was going to be the easiest thing to do to deal with that data because you were having to go through over 1,800 abstracts. [Inaudible 00:57:23] because in REDCap actually had dual – so two investigators were looking at it so we’re now at 3,600 individual records. And because it was coded – whenever you had coded data if the underlying data is coded there is going to be a good way to map it to a database. And in this instance it’s relatively easy because it’s something that we’re used to. And then it’s just a matter of doing what you do in any other systematic review and define the population, define the question, we have a data tool – if I can pull up that data tool. And you can see I’m actually done – there’s a couple other meta-analysis I’ve now done because of that.

All right and so we then have – this is what took a chunk of time and sitting down with the researchers and sitting down in this case with radiologists and other experts to decide what the – trivial and it’s just a line of code but that’s where the real science comes in and for me as a researcher having this in front of researcher forces us to be very clear as to what we’re asking and because it makes – forces us to be clear concise and consistent the data that we have is going to reflect the question better or answer the question that we think we’re trying to answer and the only person you can blame is yourself if you didn’t do definitions right or didn’t take in other aspects. It’s easier to do all these manipulations and tests; you can do it in production. You just to be more careful as you may well guess because it doesn’t like it if you drop data after you change something in production. But there have been instances; there have been a number of instances where it’s like “Oh I need to include this other category because we have some portion of the population which has the category which we never even thought of”. So –

Unidentified female: Okay – yeah I think that – and they can refer to your paper as you said. One last question before we tie this up is there a cost associated for REDCap? No there is not. And I think we have the VIReC address on your final slide and people can contact VIReC at that email address and apply to get started using REDCap –

Dr. Stephen Deppen: And I strongly suggest using those videos are great and help desk for further questions. The help desk and national consortium of help desk for incredible sticky questions.

Unidentified female: And I think that it’s that time I can – I think we’re pretty much out of time. Thank you very much Dr. Deppen for taking the time to present today’s session. Please contact the VIReC help desk at [VIReC@VA.gov](mailto:VIReC@VA.gov) if you have any additional questions. Heidi will be posting these evaluations shortly, please take a minute to answer those questions. We take your feedback into consideration when we’re planning new sessions so if you like what you hear, if you want to hear more about a topic please let us know. And thank you all very much for attending. Heidi?

Heidi: Thank you everyone for joining us. When I close the meeting out here you will be prompted for a feedback form as Linda said we really do read all your feedback and appreciate everything that we get in. I want to thank Stephen for presenting today and I want to thank our audience for joining us for today’s HSR & D cyber seminar and we look forward to seeing you at a future session. Thank you.