>> So, the next segment is talking about the cost report download. So now we're getting into the actual data. So the objectives for this segment will be to review where to find the cost report data files themselves, understand what's contained within each download, understand the definitions used to create each download, identify the resource documents that are necessary to download and set up the data files -- so we're not going through all the resource documents; just the ones to get us started with the download and setting up your data file -- and then understand the relationship between the data files. Okay?

So where do you find the data? They can be found on the cost report website. So if you are on the CMS website currently, you could go back to the Cost Reports section, and go to the Cost Report by Fiscal Year. For certain facility types -- home health agencies, hospitals, and skilled nursing facilities specifically -- they are so small that they provide the download to the cost reports in one file. And it's found within their cost report page. Or -- sorry. I'm sorry, I'm telling you the wrong -- I can't read the slides. For renal facilities, hospices, and rural health clinics; those are the facilities that have such small downloads, that they are found within their Cost Report Page themselves. Otherwise, home health, hospitals, and SNFs are broken out by fiscal year, and they're found within the Cost Report Download by Fiscal Year. There's a specific website for the downloads. We've already looked at the resource documents, but they are found on their respective Cost Report pages.

So let's take a look at the Cost Reports by Fiscal Year, and that area of the CMS website. And again this is for home health agencies, hospitals and skilled nursing facilities. This is just to show you where you would find the downloads for the other facility types that have smaller files. It's within the download section. Again, we did talk about what "HCRIS" means, but just to say it again. It's within their Healthcare Cost Reporting Information Systems database, otherwise known as HCRIS. The data are updated on a quarterly basis. What that means is that CMS overwrites the database each quarter with the latest information. So, even if there might be changes happening, they might submit their first one with -- but between that first one and the next quarter, they've submitted a settlement; they have a settlement -- you're not going to see two in the database. It's going to be completely overwritten with whatever is the most current. So you don't have to work with understanding, which one should I be looking at for a provider? Every quarter it's refreshed with what is the most current data available. The number of years available will vary depending on the facility that we're looking at. So some of them go back to '94. Others go back to '96. So it all depends on the facility type.

And then we have talked about this before, but there is a lag as far as how complete a file is, and generally it takes 12 to 18 months to have a complete fiscal year. So, fiscal year 2012 data, it's not complete right now, but perhaps within the next six months it might be.

You can probably guess this by looking at some of the resource documents, but the data are in relational database format. So it's not a nice, easy,

flat file where you've got the column headings telling you what each variable is. You do have to link files together to get out any information. The three files that you saw on some of the resource documents include the RPT file; and that contains information about the cost report that was submitted. There is another file called ALPHA, and that contains all of the character variables. And the numeric file contains all of the number values, or variables. So, you have to know where the variable resides within the forms, but you also need to know "is this variable a character or alpha variable, or is it a numeric variable?" to know how to get it out.

So again we're going to go back and talk a little bit about, what does this mean when we're looking at the cost reports by fiscal year? It includes cost reports for most of the facilities. We'd already talked about that. The other facilities, hospice, renal, and health clinics have it on their page. And again, it's refreshed quarterly.

Okay, what does fiscal year mean? The fiscal year definition is, the CMS has a federal fiscal year. So each -- when it says fiscal year, that is referring to the federal fiscal year - October first beginning, and the next year ending on September 30th. Each provider -- hospital or any facility -- can pick their own fiscal year. They don't have to have a fiscal year that matches the federal fiscal year. So then that creates a challenge of, where do you put those cost reports then? And so it is based on -- it can be based either on the begin or end date. For the facility-types -- hospitals, SNF, home health, hospice, and home health clinics -- the definition is federal fiscal year, so CMS uses that October 1st through September 30th of the next year for defining the fiscal year. The CMS will look at the provider's fiscal year begin date for those facility types that I've listed to determine which year it's going to fall into. So this is the important piece that I had. I'll take the fall for that. I had given you the incorrect information on how to even deem where it falls. So the example is, the provider fiscal year begin date is 7/1 of 2011. That would be found within the federal fiscal year of 2011, because it falls in between 10/1 of 2010 and 9/30 of 2011, which is fiscal year 2011. So that's -- when you're looking at the cost reports, and you see the fiscal year, that's how these things are divvied up. That's how it get's determined where it goes. Renal dialysis is different. So if you're working with this group, it's different. They look at CMS, renal department defines their fiscal year on the calendar year, so January through December. And further, not only do they define their fiscal year differently; they also chose to select the end dates for the provider cost reports as the determining factor as to what year it belongs to. So renal is completely different and uses the calendar year and then bases it off of the provider's end date to determine where you'd find the cost report.

Okay, now facility type. This one, you'd probably look at it and go, "This should be easy, right? I get what a hospital is; I understand what a SNF is; and I understand home health. So what is there to know about facility type that's so important?" So the downloads that you find here are for free-standing facilities submitting cost reports. And that is the key piece of information. Free-standing facilities will submit cost reports for all of their components.

So, let's talk about a hospital. Let's say a hospital owns a skilled nursing facility, or even better, they own home health agencies -- an agency or agencies -- so, well then, wouldn't the skilled nursing facility cost report just fall within the skilled nursing facility download? No. It's only if it's a free-standing skilled nursing facility will it be in the Skilled Nursing Facility download. If a hospital owns it and manages it, it's found within the Hospital cost reports.

So let's talk about -- is anybody studying hospice care? Or interested in hospice care? Oh, that's good. Or rural health clinics? Because those can be found in about four different places. So the -- hospice is one that you can find it in hospitals, as part of a hospital; you can find it as part of a SNF, home health agency, or a free-standing. So there's four. Another thing, when you're looking at the download -- I think we've covered this, but just so that everybody knows what's going on -- you will see a facility type of hospital in Hospital-2010, and you can also see that for SNF. But anytime you see the facility itself, by itself, that's using the old form, any old form that was used previously. Anytime you'd see the facility name and "dash" and then the year, that's indicating that these data are being submitted under the new form. Okay, so now, let's hope nobody's studying hospice or rural health clinics, but then also, what about if you have a changeover with the form? So, during the time periods when you have a changeover, you could have -- you could be working with multiple downloads for a given year to get the complete picture. Or -- yeah. So it really gets crazy with the amount of data files that you might be needing to look at to get the full picture. So, hospitals you'd be having -- 2010 was the year -- so you'll be working with two downloads, and it seems like you should expect at least two years where you're going to have to work with two downloads for that, to cover the changeover.

I've listed out the form numbers again. So now, as we're getting ready to download the data files and start working with the data, here's what we're going to do. We're going to take a look and get ready; we're going to look at the record counts. We're going to look at -- I say the data model -- but we're going to look at either the corrected data model or the SQL document that gives you the variable positions. Identify the space requirements that are needed for the download. The hospital cost report data files can be quite large. And then identify the software that you would use to manipulate the data. So there are -- you can use several different types of software. For the exercises today we're going to use MS Access, and you don't need to know MS Access to do the exercises; we'll walk you through it. But that is just the easiest one for today. But I have other users that are using SAS or SPSS to manipulate the data.

So again, here's where we go to get our resource documents. We're pulling out the record counts. Here it is again. We're going to pull up the data model or, better yet, the SQL, the HCRIS table description is an SQL document. There it is again, the incorrect version. And the main thing that we will -- I wanted to point out on this slide is that the linking key across all of the files that we'll be using is called the report record number. So that's a pretty important field.

Alright. So let's take a look at what we're going to do with downloading
and linking the data. We're going to the cost report by fiscal year page.
We're going to look at the download for fiscal year 2012. When it opens
up, this is what you would see. You'd see these three files. We've already
talked about this, but just let's go back -- I'll go back one. You'll always
see in the download, it'll be the facility type. If it's the form for the
new one, it will be listed there, underscore the fiscal year that you're
looking at, and then the name of the file, the alphanumeric or report file.
For the 96 forms for hospitals, they'll have a file that's called a roll-up
file. That contains summed or total values. CMS, in the past, didn't
supply every single variable that was available in the forms. Some of the
forms include a total line. And CMS thought that just took up too much
space, and people could sum these values on their own. But then they found
that there were people that really valued this file, because it was a
resource that provided all the totals. I generally, I have not used the
roll-up file very much, but it is available for the older data sets. And
CMS is no longer producing a roll-up file for the newer data files. All
right, so let's take a look at what does one look like? What does a report
file look like?

So if you were to just open it up in Notepad, this is what you see. It's
a CSV file. The first column is your report file, and then -- it doesn't
have very much across. And so this report file, it contains information
about the cost report that was submitted. So the record, the unit of
analysis here is the cost report. And some of the elements include the
provider number, the fiscal year begin and end dates, the status of the
cost report, and the ownership status to name a few.

Let's take a look at the ALPHA file; what does that look like? Not very
much in here as far as columns go. Report record number. Then we have this
thing that looks -- that's the worksheet codes. We'll talk more about what
comes next, the line and column numbers, and then it looks like a bunch
of text. So there's one, two, three, four, five columns in the ALPHA file.
And each record in the ALPHA file is a cost report variable. So it's not
a cost report; it's a cost report variable. Some of the data elements in
the ALPHA file would include the hospital name. Yes, the provider number
is an ALPHA variable, because it has a leading zero. And yes, no variables
are also examples of what would be found in the alphanumeric file.

The numeric file looks almost identical to the ALPHA file. One, two, three,
four, five columns, except we've got all numbers on the end. This file,
again, is -- only contains the numeric variables and some of the data
elements that you might see in this would be beds, discharges, net income.
But from looking at this, could you tell me... which one is net income?
So it's not evident by just looking at this to pull out something like beds.
For this, what we're going to do is just talk about what resources we need
to load up the data to get started. So in this example I had opened up
the RPT file, and I opened it in Excel. And the last record was 846, and
I opened up the record counts, and I looked over here and, yep, it matches.
I have 846 records, and that's what CMS says I'm supposed to have for this
download. So that's a good sign. Here's the continued error on the

alphanumeric table. Again, we would use this if it was correct to give us the location and the variable types. So now, in this example here, I'm in Access. I'm loading up my data for the report file. I'm listing out all the variables, listing out the data type. And here, I wanted to link to the ALPHA, and the linking key is the report record number. So just remember that.