5. Managing Users and Permissions - Transcript

Introduction [00:00]

Welcome to the fifth video about the RECAP Preterm data platform. In this video, I will be demonstrating how to create user accounts on a node and also how to grant access to data on a node. As usual, I advise you to watch our previous videos first, before this one, if you haven't yet. And again, I will be following our wiki, right here. So you can find a link to it in the description of the video.

Creating Users and Groups [00:26]

So in this video, we're going to cover this part right here, Part D: “Managing Users and Permissions”. And remember, just in summary, we started by, on our second video, we started by importing data into our node. And then we went up one level to the Study Manager just to describe our study or the cohort that we were using, which was EPICE-PT. So it's just to give a little bit of context to the data that we put on the repository. So just to contextualise... in what context was that data collected. And then on the third video, we harmonised the data. And now presumably after you have harmonised data, there was a purpose to you doing that. And typically that would be for someone then to pick up that harmonised data and do some analysis with it. And that will be our last video about how to perform analysis on the platform. But before we do that, we have to go over this first because whoever is going to do the analysis, say you harmonise the data like we did on the previous video. And now someone is going to take that data and do some analysis. That person is going to need an account on the node because... how did we put data in the repository and then harmonise the data? We went into the Data Repository we logged in, right? In order to be able to do all of that. So the person who's going to do the analysis is also going to need an account with which to log in to the Data Repository. So that's why we need to go over this first, how to create users, we're going to see how to create a user on a node. And then how to grant that user access to a particular table on our Data Repository. So let's do that.

So the way you create users is... let me just go back here to the homepage of our testing node. And the only application that we haven't talked about... so we've talked about the Data Repository, where the data is, talked about the Study Manager, where we describe the cohort and every relevant information about the study to which that data relates. And then we haven't talked about this one yet. So this is your Authentication Server. This is

where you create users that will be able to access the other two applications. So let's go there. Let's go to our Authentication Server. This is it. So the first thing to do is login. So we can click here on administration. I'm just going to log in, as usual, I'm using the administrator account. Okay, so “Welcome to the Authentication Server”. So how do we create a user? It's fairly simple, we just go to users, here. And there are no users yet. Well, there is one user, which is the account that I'm using. But that's the administration account, that sort of a special account that doesn't appear here, that's always the default account that always exists on a node. But now let's create another one, let's create an account for the hypothetical person that's going to do the analysis and that's going to need access to our data. Let's create an account for that person.

“Add User”. Now, not all of this is required. But, so we have to specify a username, I'll just use my own name, “goncalo”. So right there, and then an email, I'll just use this one. Which by the way, is the email you can use if you have any questions about anything on these videos. And then a password and confirmation of the password, then we could write the first and last names, but it’s not required, so I'm just going to leave that empty. And then the other thing, the most important thing here are these two fields right here. So this is what's going to define which parts of our node this user will have access to. So let's, for now let's leave those empty. And I'm just going to click Save. Okay, and that's it. So the account is created, it exists here.

So I'm going to try to use this new account. So for that, I'm going to open a private window here, just to put them side by side. So I'm going to access the node. So here's the node that we're using. So now what I'm going to do is try to use that new account to log into the Data Repository of this node. So I'm going to go to “Applications” and “Data Repository”. Okay, here I am. So now I'm going to try to login using that account. But remember, this is... on the left side, we're logged in as the administrator, where we created this account. And then on the right side... this is why I use the private window. This is as if this is a different session. It's as if I was on a different computer, right and if I was, you know, this person now trying to log in. And I tried to log in and the authentication failed.

Granting Access to Data in the Data Repository [07:20]

And that's because the account is there, it exists, but we haven't specified what parts of our node we want this user to be able to access. And the way we do that, let's click on Edit here. and scroll down here. So these are the two important fields that I mentioned before. So we need to specify here, the application. So we want this user to be able to go to the Data

Repository. And the Data Repository... that's what we call it the Data Repository, because it's more of a descriptive name, but the actual name of the application is Opal, we can see right there, the application is called Opal. So the Data Repository is called Opal, so if we want this user to be able to access the Data Repository, we have to go down here to “Applications” and select Opal. Okay, and now if I click save... so the account is still there, but now it has this extra permission here. So I'm granting this account permission to log into Opal, which is the Data Repository.

So if I come back here to the right side, and try to log in again, I can log in. So because of this, right? So this user is now able to log into our Data Repository. But now if this user goes to projects, to look at the data, it can't see any data. And that's because the only thing this does is allow this user to log in to the Data Repository, but nothing else. Okay, so we know that there's data here, we've put data there, it's just this user cannot see it, because we haven't explicitly said that this user must be able to see that data. So this does not allow the user to access any data, it just allows the user to log in to the Data Repository. So the next step would be... so this is the first step, the account is created, and the account can log in to the Data Repository, but nothing else. But if we want that user to see any of our data, we have to go ourselves here to the Data Repository. Let's go there. So I'm on the landing page of the node again, and I'm going to go to applications, Data Repository. And I'm going to log in as the administrator. So I'm logged in on both places on the right and on the left, with different accounts, right, right here, I'm the administrator, and then right here, and that, that user that's going to do the analysis that does not have access to anything yet. So as the administrator, and the person that created that project, a few videos ago, there it is, I can go to that project. So as you can see, there's a project right here, and we can see it because we were the ones that created the project. But this user cannot see it yet. So what we can do is click on the project. And then go to the list of tables, which is this second tab here. There are our tables. So if you remember, this is the original table with the original data. And then we created a view over that table that has the harmonised version of some of those variables. This is what we did on the previous video. So now typically, we would want this user to be able to access this data, right? The harmonised data. We could also grant that user access to this data, but typically, what you do is harmonise the data and then grant access to the harmonised data. So how do we grant access to this data or this table? We click on it. Let's click on it and then there's a tab here: “Permissions”. If we click on it, we can see there's already a permission here. For the administrator. This is an implicit permission because it's saying that the administrator can administrate, which of course,

is true because the administrator account is a special kind of account that can basically do everything inside the Data Repository.

So let's now add, we want to add a permission for this specific user, this one. So let's click “Add permission”, “Add user permission”. Okay, so the first thing we have to do is write the name of the account to which we want to grant access. And in our case, it's this. And then the second thing is we have to choose, what level of access will this user have to our harmonised data. So the most basic kind of access is this one, “View dictionaries and summaries”. Let's see what that does. Let's, select that one, and then click Save here. Okay, so there, so this user... remember, we are inside this table with the harmonised data. And we are saying that this user can now view dictionaries and summaries of this table. So if I go back here to the right side, to this user, now if I refresh the page here... so I've refreshed it, and then now we can see it, there is the project. And if I click on it, and list the tables here, there it is, there's the harmonised view. So remember, the permission that we granted was “View dictionaries and summaries”. So let's see what the user is able to do with that. Let's click on that. And he of course, can see the dictionary, here it is, the dictionary, the table and summaries, which means that if I click on a particular variable, let's say this one, I have this tab here, click on “Summary”, and I'll be able to see summary statistics for the variables. So just frequencies, frequencies for categorical variables, because if it's for continuous variables, we can see. Let's see, I think we harmonised this one on the previous video. So we should have data here, yes. So we can also see some extra statistics here for continuous variables.

So basically, what the user is now able to do is, let's go back here, look at the dictionary of the table, and then see summary statistics for each variable. But this user cannot see the data itself, does not have access to individual level data, because as you can see here, the only tabs I have here are “Dictionary” and “Summary”. But if you look here, on the left side, where I'm the administrator, I have more tabs here, including values, right? So this is the tab where you would see the actual data. And this user does not have that, because he doesn't have access to the data, because the level of access that we chose does not allow that.

So let's increase the level of access, let's click on Edit here. Now, instead of this... so we already saw what this does. And now this is the exact same thing. But also, the user will be able to see the values. So let's click on that and save. And then on the right hand side here, I'm going to refresh. And you can see that now, this user is able to see that tab and if he clicks on it, he will be able to see the actual data. So that's the second level of access,

there's a few more, which are the next two are more combinations of the first two.

So this third one will be “Edit dictionary and view summaries”. So it's basically the same as the first one. But instead of “view dictionary”, it's “edit dictionary”, meaning that the user will be able to edit anything that you see here, such as labels and categories, for example, or value types. And then the fourth option is “Edit dictionary and view values”. So it's similar to the previous one. But instead of just the summaries, now the user will be able to also view the values. So it's sort of incremental, right. And then of course, the last one is what we usually call the “god permission”. Meaning that the user will be able to basically do anything on the table. The user will have the same permissions as an administrator would, including... You can see here “Full access to the table, including edition of the dictionary and individual values”, so the user will be able to erase or import more data into the table, and also just remove the table altogether. So that would be like the ultimate level of access.

Setting Up Permissions for Data Analysis [17:47]

So which level of access should we choose then? Well, that depends on what the purpose of the access is. So in the next video, we will see what kinds of analysis you can do on the platform. And it really will depend on... the access that we have to choose here, will depend on the kind of analysis that we want the user to be able to do. But a good rule of thumb is to always grant the minimum level of access that is going to be needed. So we will cover this in the next video. But basically, the analysis... There's two ways to do analysis here. One requires you to have access to the actual data, to individual-level data. And for that, you would have to at least grant this, right? “View dictionary and values”. So, if the user is able to view the values, then that's access to individual-level data.

But there's another sort of analysis, which is distributed, parallel, non-disclosed analysis, which we're going to talk about in the next video. And it's done using a software called DataSHIELD. And that allows you to perform analysis but without having actual access to the real data, to individual-level data. So for that, if you're using DataSHIELD, which we'll see on the next video, this is the only thing that's needed, the minimum permission that you would have to grant a user in order for that user to be able to use DataSHIELD to analyse your data using DataSHIELD.

So that's basically it for permissions. We haven’t we haven't talked about... we just talked about users. But there's also, let's say... so I'm assuming that there's only one person that's going to do the analysis, but there may be

several people. So what you could do instead, instead of creating a user, you could create a group of users. So you just “Add group”. And then... let's say, you have five people that are going to do the analysis, and all of them need access. Instead of creating five users and then granting permission to each individual user, you just create a group, and then add those individuals, those users, to that group. And then on this side, on the Data Repository, you will have to add just a single permission to the group, and then all of the users inside of that group will have that permission. But that's it.

So we're going to talk more about, and test, the different levels of access, depending on the kind of analysis that we want the user to be able to do, on the next video.