BHP LIMS - Feature #4842
Create export functionality for Senaite lab data - TD & Karabo

04/10/2019 18:37 - Justine Legbedze

Status:	Resolved		**Start date:**	04/10/2019
Priority:	Normal		**Due date:**	
Assignee:	Ame Diphoko		**% Done:**	0%
Category:			**Estimated time:**	0.00 hour
Target version:			**Spent time:**	0.00 hour

Description

Adding ticket to system for this critical milestone.

History

#1 - 09/10/2019 15:17 - Coulson Kgathi

- Status changed from New to In Progress

#2 - 16/10/2019 15:10 - Coulson Kgathi

- Status changed from In Progress to Resolved

#3 - 29/10/2019 15:53 - Justine Legbedze

- Status changed from Resolved to In Progress

As per 10/16 weekly call, Coulson to update team once export functionality has been tested by lab and is ready for study team to test. Changing status back to "In Progress".

#4 - 06/11/2019 03:35 - Coulson Kgathi

- Status changed from In Progress to New

#5 - 06/11/2019 04:04 - Coulson Kgathi

- Status changed from New to Resolved

#6 - 11/11/2019 17:29 - Keolebogile Nicholas Mmasa

- Status changed from Resolved to New

1. Can you separate date and time field
2.Both at "samples" and "analysis" links/tabs, can we have option too download ALL that was input into senaite over a given period? in addition to the already added individual sample state. e.g.: If ii want a csv of all Lab reqs that we keyyed/captured into senaite between 01-jul and 21-Jul, what will be the right "sample/analyss state" to use?

#7 - 11/11/2019 17:29 - Keolebogile Nicholas Mmasa

- Status changed from New to Feedback

#8 - 11/11/2019 20:05 - Coulson Kgathi

- Status changed from Feedback to Resolved

Hi Nicholas

1. Sorry about separating date and time, we cant do it at least not any time soon at all.

2. This is already there just do not select any filtering then it will pull all.

Confirm if i answer all your questions

#9 - 11/11/2019 21:15 - Justine Legbedze
- File Senaite Exports UAT_Sample vs Analysis State_2019.11.11.pdf added
- Status changed from Resolved to Feedback

1. In response to the date/time issue, while the date and time cannot be separated at this time, the format of datetime should not be a string character variable (as it currently is), but rather formatted as datetimes (see below suggestions). Otherwise, end users (TD team and lab) will not have the capability to manipulate the files as needed. I am unable to change the formats in Excel, and it'd take some coding in SAS to try to manipulate these fields.

Samples Excel File
Variables with both date and time: Date Sampled, Date Received, Date Assayed, Date Published - **reformat to m/d/yyyy h:mm**
Variables with date only: DateOfBirth - **reformat to m/d/yyyy**

Analyses Excel File
Variables with both date and time: Date Created, Date Sampled, Date Received, Date Captured, Date Assay, Date Verified, Date Published - **reformat to m/d/yyyy h:mm**
Variables with date only: DateOfBirth - **reformat to m/d/yyyy**

2. To follow up on Nicholas' "sample state" vs. "analysis state" question, shouldn't there be a variable in the CSVs that shows the current status of each sample? I also don't really understand all of the state types (see attached). And I'm not so sure that I understand how to view each file. In the previous LIS system export, we had 3 files: received, stored, and resulted. Ultimately, I'd like to be able to export files for specimens that were received by the lab, specimens that were aliquoted and placed in storage, and specimens that were assayed along with their results. I guess we can discuss on tomorrow's call on how to achieve this.

3. "Sample Type" variable - how is this field derived? For some sample IDs (i.e. - 08504AAD64), it appears like it is a combination of the aliquot type (whole blood) and container type (SST), while others (i.e. - 08506AAD6402) is just the aliquot type (serum). There should be 2 separate variables for aliquot type and container type, no?

4. "Sample Template" variable - how is this field derived? Is this the same as the panel name captured in the EDC? Why is it missing for some sample IDs and not others?

5. Could the lab please share a list of all aliquot type codes? It took some time to understand how the sample-id is constructed, but it seems that the 4th and 5th digits are unique to sample type (i.e. 02 is whole blood, 31 is pbmc).

6. Minor issue but noteworthy - there should be a separate naming system for samples files vs. analyses files. Right now, both are named "LIMS-Export-date".

7. Another minor issue - Is it possible to code that any values that have leading blanks to remove those leading blanks? For example, if a PID is " 085-40990165-1-10" to be "085-40990165-1-10".

#10 - 13/11/2019 00:23 - Justine Legbedze

As per today's DMC-lab call, I am updating next steps for the feedback that I left on 11/11:

1. **Pending**: DMC to fix formatting for all dates and datetime fields.

2. **Pending**: DMC to add a 3rd export function on Senaite "Data Export" screen that will allow us to export a Excel file for specimens based on the following states: received, stored, and results published.

3. **Pending**: DMC to provide document that lists the different "Sample Types".

4. **Pending**: DMC to provide document that lists the different "Sample Templates".

5. **Closed**: DMC provided screenshot of aliquot codes from Senaite.

6. **Pending**: DMC to update naming of files for "samples" vs. "analyses" so that they are not named the same when exporting from the system.

7. **Pending**: Justine to send data cleaning query to TD team of all client-sample-id's and participant IDs that need to be manually fixed in Senaite. DMC to also look into validation coding on the back end that will not allow the clinicians to enter these IDs starting with blanks. For example, if a PID is " 085-40990165-1-10", it should be "085-40990165-1-10".

#11 - 20/11/2019 01:01 - Justine Legbedze

As per DMC update on 11/15 (Ame Diphoko):
The test server has been updated with the changes as per request on the last td, lis call. We have since added the functionality where you will be able to prepare export data with the option to apply state filters [Received, Stored, Published], "Received" grabbing every state except [cancelled, invalid, rejected, to_be_preserved, sample_shipped]; The system now also allows for you to apply a filter on sample type.
Justine's Feedback on 11/19:
On the Data Export screen, I see the new section for "Received, stored, published" but I'm still not clear on why the same report options are listed under Samples vs. Analyses. I ran 3 export files (received, stored, published) from both Samples and Analyses.
  - In my first comparison (received), the Samples file has n=3827 observations, while the Analyses file has n=4687 observations. I'm not sure which file provides a list of samples received. I'm not sure which file reflects the list of samples received.
  - In my 2nd comparison (stored), the Samples file has n=2142 observations, while the Analyses file has n=1781 observations. I'm not sure which file reflects the list of samples stored.
  - In my 3rd comparison (published), the Samples file has n=670 observations, while the Analyses file has n=777 observations. I'm not sure which file reflects the list of samples published.
Ultimately, I'd like to be able to export 3 files:
   1. Specimens that were received by the lab (received)
   2. Specimens that were stored (stored)
   3. Specimens that were assayed along with their results (published)

Also, the date and datetime fields are still being shown as string character formats and not as date formats.

#12 - 20/11/2019 14:55 - Coulson Kgathi

Hi Justine the difference between samples and analysis was given in the last call we had with the lab, i can try explain them again in the call today, but maybe the lab are the best people to explain this but all in all they are both samples but differ in the states they can take.

#13 - 20/11/2019 14:57 - Coulson Kgathi

Note that we may not necessarily give you data exactly as it was from the old LIS as they are 2 different systems and also tthe schema in which this 2 systems store data is different. But i think what is important is that you get all variable that you need for all TD samples.

#14 - 20/11/2019 15:28 - Justine Legbedze

Coulson - Perhaps on our call today, it would be helpful for someone to walk me through exactly how to run a report if I want to know for example the samples that were received by the lab. Because the way that the current module is in place, I can run the "received" report under both Samples and Analyses, but the # of observations in each output file is not the same.

#15 - 26/02/2020 01:43 - Justine Legbedze

Requested meeting this week with DMC and lab to discuss data needs from Senaite.

#16 - 05/03/2020 18:30 - Justine Legbedze
- File TD-Lab-DMC Senaite Meeting Agenda_2020.03.03.docx added
- Subject changed from Create export functionality for Senaite lab data to Create export functionality for Senaite lab data - TD & Karabo

I have attached the agenda from March 3rd call with DMC (Coulson, Ame) and lab (Terence, Comfort) teams for reference.

The following next steps are to occur:
1. DMC and lab will create a schema diagram that outlines the data flow steps for sample processing in Senaite (similar to other studies), as per Dr. Moyo's suggestion.

2. DMC and lab will provide a data dictionary that clearly defines each variable and associated values (i.e. - sample type, sample state, analysis state).

3. DMC will address the issue between "Samples" and "Analyses" export files (with the assistance of the consultant).
-For example, see sample_id 08502ADR84. On the Samples file, there is 1 observation for the parent tube for the panel pbmc/plasma storage (with a sample_state of "stored"). On the Analyses file, there are duplicate sample_ids of the parent tube along with unique IDs PBMC/plasma aliquots.

4. DMC will create a new ticket on GitHub and update regularly based on feedback provided in this Redmine ticket.

Coulson - we will check in on next week's DMC call (3/11) to determine progress made with these tasks and when we should schedule the next Senaite call.

#17 - 05/05/2020 18:35 - Justine Legbedze
- Project changed from Bhp085:-TshiloDikotla to BHP LIMS

Updating to LIMS Redmine Project.

#18 - 05/05/2020 18:37 - Justine Legbedze
- Assignee changed from Coulson Kgathi to Ame Diphoko

#19 - 05/05/2020 18:53 - Justine Legbedze
- File LIS Workflow.docx added
- File Sample Identifier Breakdown_SM_Jordi--Karabo.docx added
- Status changed from Feedback to Resolved

Updating ticket to Resolved status per Ame's email on 19-Mar:

"The test server has been updated with the latest code changes concerned with fixing the data export function for "Samples" and "Analysis" files. Please test it thoroughly and give feedback to DMC team."

Also attaching LIS workflow and state transitions and definitions for sample processing documents from DMC.

Files

Senaite Exports UAT_Sample vs Analysis State_2019.11.11.pdf	304 KB	11/11/2019	Justine Legbedze
TD-Lab-DMC Senaite Meeting Agenda_2020.03.03.docx	225 KB	05/03/2020	Justine Legbedze
Sample Identifier Breakdown_SM_Jordi--Karabo.docx	33.3 KB	05/05/2020	Justine Legbedze
LIS Workflow.docx	102 KB	05/05/2020	Justine Legbedze