Meetings

Covid-19 msc recurrent meeting:

- Zoom: [https://zoom.us/j/94889193537](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fzoom.us%2Fj%2F94889193537&data=05%7C01%7Ck.thalassinos%40ucl.ac.uk%7C1a417abafda34023ca7008da812005ac%7C1faf88fea9984c5b93c9210a11d9a5c2%7C0%7C0%7C637964271815901591%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=uQiTx80jN3QbocjUn%2BiF1zoEp2CEEUy6%2BCMzjHKDkhs%3D&reserved=0)

- [Microsoft Teams](https://eur01.safelinks.protection.outlook.com/ap/t-59584e83/?url=https%3A%2F%2Fteams.microsoft.com%2Fl%2Fmeetup-join%2F19%253ameeting_ZjFjMWI3ZTYtNDgzOC00Mzc5LWIxOTUtYjljYmFlYmNjNmEz%2540thread.v2%2F0%3Fcontext%3D%257b%2522Tid%2522%253a%2522c152cb07-614e-4abb-818a-f035cfa91a77%2522%252c%2522Oid%2522%253a%25225e2e30aa-d5aa-47a0-9abe-014dc7cc030f%2522%257d&data=05%7C01%7Ck.thalassinos%40ucl.ac.uk%7C1a417abafda34023ca7008da812005ac%7C1faf88fea9984c5b93c9210a11d9a5c2%7C0%7C0%7C637964271815901591%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=C%2BLeK%2FnucpHcqiqmPZ7h46mEwRzVOw%2ByTeJnx67CFmU%3D&reserved=0):

***31/08/2022***

1. About databases, in the action plan is mentioned only PRIDE as data repository.

What about the metabolomic and lipidomic data?

It looks like there are at least three databases to access. I am right?

Proteomic - PRIDE

Metabolomic - MetaboLights

Lipodomic - Zenodo (based on the researches published in the website)

OK

Should be possible to collect all the omics data for a single subject's sample



Input

Raw Data **Proteomics**



Access online database

**PRIDE**



Input

Raw Data **Metabolomics**



Access online database

**MetaboLights**



Input

Raw Data **Lipodomics**



Access online database

**Zenodo**

Accession Code **P**

Input Metadata

Accession Code **Z**

Accession Code **M**

Input Metadata

Input Metadata

**Input Record**

**Cov19-MSC Database**



**Cov19-MSC Website**

anonim data

1. ~~How I have to consider EMBL-EBI COVID-19 Data Portal? I do not need to query also this data portal. It is more about PRIDE~~

1.2 It looks like that MetaboLights requires an API Key to work. I did not register but I am wondering if I need to do so or to use a lab account details.

***Please note that you would need an active MetaboLights account and API Key to access the API.***

<https://www.ebi.ac.uk/metabolights/guides/Files/Download>

1. There is a metadata format across all the omics projects? For instance, metadata in PRIDE share something with metadata in Metabolights?
2. For the downloading operations I think we should consider the metadata and the accession code.

I assume that we are interested in fetching raw data using one or both types of input data. I am right?

***06/09/2022***

1. We have to create our metadata?
   1. There is a metadata format across all the omics projects? For instance, metadata in PRIDE share something with metadata in Metabolights?
2. Approve the input-output schema and meta-relationship.
3. Login or not? In case GDPR?

Connect patient ID with Surrey subjects

Metadata are in phosp

Holly Lewis and Mel

Wait for data is coming

***03/10/2022***

1. It will be a website or only a page embedded into the current website?

Up to us

1. How to identify a group of subjects:
   * ID generated by the Cov19 website.

**Experimental Group**

**PRIDE**

**Cov19-MSC Database**

**Experimental Group ID**

**MetaboLights**

**Zenodo**

**Cov19-MSC Website**

**metadata**

**ID**

Go for the Minimum Vaiable Solution

**Asincronous data are not suitable for the ID generation.**

If there are three omics data type that refer to the same sample, metadata could not be considered for the creation of a unique ID.

In the worst case, three researchers could upload the omics data separately at three different moments. But the unique identifier of a subjects group must be generated only the first time, whatever the type of omic data is uploaded. This means that type of data related to the single omics experiment cannot be considered in defining the unique experimental group ID.

1. Define the search options.
   * For wich features/metadata/code must be arrange the search operation?
2. We have to create our metadata?
   * There is a metadata format across all the omics projects? For instance, metadata in PRIDE share something with metadata in Metabolights?
3. Login or not? In case GDPR?

My email

Should we discuss with Melanie the requirements and then communicate the decision to the group?  It looks like she is ready only to show and explain the data as you requested last time.  
Or should we wait until the next meeting and discuss the requirements with the entire group?  
  
Should I have to prepare a couple of slides with the general idea I showed to you last time and then modify the schema under the group's suggestions/requirements?  
Or do you think that the group already knows what should be implemented on the website and in this case, I will follow the project?

It would be better if I would join the lab during the week.

So far 10.30 at Friday

***04/10/2022***

Andy Jones

Says metadata are on Trup

DDa analysing Vs Spectronor

***17/10/2022***

The operations allowed to the user will be **QUERY** and **VISUALISE**.

The submission process will be in charge of the Cov19-MSC team.

Refer to UniProt search webpage.

***18/10/2022***

**PHOSPH Metadata**

*Sample Delivery* - from which research center the data are coming

*Sample Processing* - Refer to Proteomics Metabolomics Lipodomics or SWATH proteomics

*Data Processing* - Which software or method has been used

***21/10/2022*** Drupad - Holly - ~~Eleanor Sinclair~~

Single study ID - **Redcap** online software from Surrey.

Samples could be different but must be standardise

The ID substitution:

protect the patient privacy

do not bind the single group to the Sample Delivery

Surrey ID will be dummy data

Possible searching from Drupad

*Age*

*gender*

*oxygen*

*drugs*

*survived*

Wait for the dummy metadata from Drupad/Holly

***01/11/2022*** MSC meeting

**Sample IDs**

We have all samples.

But we do not know how to link the samples IDs that are storage in different biobanks

The problem is related to the PHOSP and ISARIC IDs that mask the Surrey and Manchester IDs

***15/11/2022*** MSC meeting

Manchester will be interesting to explore by the way.

Upload to PHOSP what?

**DATA CATALOGUE** what sample is taken from where.

The data catalogue has the purpose to identify the cohort of patients over time in any site.

As a connecting entity the catalogue will connect the data from different repositories

At the end, put the data on the public repositories but first of all ....... catalogue !!!

The most updated metadata should be the “Surrey model”. Perhaps, this model will be shared through the group.

Not clear the id issue from ISARIC (PHOSP should include also ISARIC as mentioned by Drupad)

Look for hospitalised and post-hospitalised set of patients. PHOSP is the most important source of data because include the time dimension (patient with follow-up).

At the moment the most searchable columns should be

STUDY ID

age

sex

covid stage

primary site owner

proteosome ???

***06/12/2022*** Drupad

So far, I have considered the subject group as an entity with specific code that represent a group of subjects.

Instead, the group of subjects is not an entity but it is a composite key or a view that is defined combining different variable values at the same times.

This concept modify the meaning of cohort in this database. Since cohort is similar to the subject groups concept I always struggled to think about them in a separate way. With this new update a cohort could be seen as the group of patients that have been collected during the first COVID wave (or the second COVID wave as another cohort).

Always note that Surrey and ISARIC are very similar.

**Primary site owner** is the reference person to contact for the specific site.

We have selected the main tables that will compose the database. These tables have been selected from the grouping activity made through Python and Excel.

I downloaded the labelled version of the REDCap table. I grouped the columns using the column “Complete” as references/delimiters of each group. Then I moved each group in Excel, named each of them as a table and discussed about wich one was to implement in the database model and which not (for the moment). Some tables have been summrised with one value. Some columns became dropdown menu or simply became one column because it was possible to represent them in a more concise way.

***13/12/2022***

Iliyana maybe is working on data comparable with ISARIC or PHOSP.

what happened to someone that has been hospitalised ............

Instrument comparison should be one of the features to have in the db.

Doubt:

The concept of **Technical Replicate** should be considered?

This concept is not referred directly to aliquots. Instead, it is the concept of a sample that has been aliquoted and the single aliquote is analised two times (through the same type of analysis, for instance). These two analysis will be two **‘technical replicates’ of the same aliquote of the same sample\_type**.

No technical replicates

**Clinical data**

From REDCap I can see different clinical data that could describe different aspects of the patient and the sample.

Have you already though about any descriptor?

Perdita mentioned the Covid stage. This could be a descriptor if created combining different columns in REDCap that can define the “Covid situation” of the patient.

We will use the setup of variables that we have selected with Drupad on the 06/12/2022 meeting.

Next meeting 10 january

***20/12/2022*** Drupad

references to db Tables

**Sample**

~~7 - Repeat Instrument : if we need it please explain~~

~~8 - Repeat Instance : if we need it please explain~~

9 - Survey Timestamp : it is not clear why is reffered to the survey\_date. It could be change in sample\_date to record the day when the sample has been taken?

evenet as multiple sample of the same type

**Analysis**

6 - qc\_used : was one of the Perdita requirements - list of names

**Inclusion\_Assessment**

3 - A history of self reported feverishness or measured fever of ≥ 38°C

If I leave a boolean this field just reports if the patient had the feaver before he joined the hospital.

OTHERWISE specify other options

**Comorbidity**

These variables could be represented with a list of options

3 - Chronic cardiac disease, including congenital heart disease (not hypertension)

5 - Chronic pulmonary disease (not asthma)

7 - Moderate or severe liver disease

8 - Mild Liver disease

Instead, I must consider them as simple booleans

YES FOR LISTS

15 - Diabetes complications

if a boolean it just indicates that there are some complications.

OTHERWISE could be another table where more complications could refer to the current patient diabetes stage. Would be one to many relationship Comorbidity --> Diabetes Complications

**Current\_Medications**

I have to consider a relationship one to many Subjects --> Current\_medications

Medication

Subject

Medication

Medication

This table lacks of the medication name for instance.

Maybe in this phase you want to model only one medication?

10 - If it is a boolean for the admission to ICU then it will be in conflict with the table ICU\_Admission

**Interventions**

The first thing I can see in this table is that the interventions span from pharmachological to technichal. I suggest a one to many relationship from Subjects --> Interventions where will be possible to link more than one intervention to the same subjects.

|  |  |
| --- | --- |
| **Pharmachological** | **NON-Pharmachological** |
| Dopamine | Invasive ventilation |
| Dobutamine | High-flow nasal canula |
| Milrinone | ECLS |
| Levosimendan | ECMO |
| Epinephrine/Norepinephrine |  |

Could be possible to have a patient with high-flow nasal cannula that is taking neuromuscolar blocking agents in prone position?

In the opposite case, all the options in this table must be considered mutually exclusive.

2 - urine\_fr: it could be part of the patient clinical assessment

**Remdesevir**

~~6 - Results available for sample taken on the date in section 1 above ?~~

~~7 - Date of assessment: this assessment is for ???~~

**Pathogens - Pathogens Treatments**

**Route** tablets to inhalation -intravenus

We agree to merge the two tables. However they are clearly divided in pathogen identification and pathogen treatment. But there is a potential one to many relationship between Pathogens --> Pathogen\_Treatments

What I have to do?

One doubt is still related to the columns :

|  |  |
| --- | --- |
| 274 | Collection Date |
| 275 | Bio specimen type |
| 276 | Laboratory Test Method |
| 277 | Result |

Why they are considered in the Pathogen Treatment group? I am missing something ?

**ICU\_Admission**

7 - renal\_therapy: *Renal replacement therapy (RRT) or dialysis* and Inotropes / vasopressors look like different features.

Instead, are they related?

references to the db Columns Clustering tag

colNumber - 325 : 332

Are these conditions mutually exclusive? If yes I will create a list of options where only one could be selected.

not at the same time but could change in the timeline consider to represent the timeline

Or could be the case where coexist more than one Oxygen therapy?

e.g. Prone ventilation - Tracheostomy

If yes I need to create another table **Oxygen Therapy** to manage a relationship one\_to\_many with the Subjects table.

colNumber - 335 : 342

if the 325 : 332 are mutually exclusive than I can consider only one **confirm\_duration** reference.

Conversely, if they can be more than one I need another table

Lenovo UK

+44 203 467 4050

<https://www.buy-keys.com/productcategory/software/microsoft-office/>

lenovocareuk@lenovo.com