The website Covid 19

Mass spectrometry Coalition

<https://covid19-msc.org/>

Working Setup

In person lab time

10.30 -Thursday

10.30 - Friday

Recurrent meetings

**Konstantinos :**

- [Zoom](https://ucl.zoom.us/j/7219642208?pwd=SjR5WVVFUGg2MU9lVitEdnRlU2ZaZz09)

**Covid-19 msc:**

- [Zoom](https://zoom.us/j/94889193537)

- [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)

The Idea

Take samples with different range of severity.

Collecting data:

proteomics

metabolomics

lipodomics

We must make the repository with all data and link the data with the metadata.

Create a dashboard where is possible to collect again the raw data and reprocess everything.

Work package 3.

Flask for the database in the backend.

We want to connect pride - metabolomic

From <https://covid19-msc.org/action-plan/> I considered the point 5:

5.      Open data repositories

Critical to our collective effort is how we manage the data. We suggest that for omics data you use a data repository (ideally PRIDE for proteomics datasets, as part of ProteomeXchange), and that you tell us the accession number. PRIDE enables the inclusion of the submitted datasets in the EMBL-EBI COVID-19 data portal (<https://www.covid19dataportal.org/>). After deposition we will record these along with the requisite metadata to identify the datasets. We will soon be updating the website with a dedicated page where you can submit this information. The website will also be updated regularly so that many researchers can access the processed data. It is through this that we will gain the most and most quickly.  For native data on recombinant viral antigens get in touch as we will make the data repository and we will curate that data for all to examine and to compare to data from patient samples.

What the best way to connect these data (metadata are necessary)

Storyboard

The website should manage two main operation:

* Upload record
* Download record

**Upload a new record**

A researcher just completed the collection of his data.

The researcher should submit his data on one of the repository specific for each omics data type:

|  |  |
| --- | --- |
| **Omics Data Type** | **Repository** |
| Proteomics | PRIDE |
| Metabolomics | MetaboLights |
| Lipodomics | Zenodo |

The **accession code** to these data,whatever is the repository where they have been uploaded, will be part of the patients record uploaded on the Covid19-msc website.

The user should knows if its experimental group of patient has already generated other omics data from other researches. This information is essential for create new group codes or not.

* *YES he knows*
  + In this case the researcher must know the Covid19-msc's accession code for the specific group of patients already present in the Covid19-msc database.
  + **He must use this code to fetch the specific record.**
  + Finally the record must be updated with the accession code related to the specific repository where the omics data have been uploaded.
* *NO he does not know*

In this case the website must generate a new code for a new entry

Download a record

A researcher wants to access the data for .........

Therefore she/he goes to the website and start the search on the search line

It could also refine the search providing additional information like metadata

For which

META-RELATIONSHIP

V1



Input

Raw Data **Proteomics**



Access online database

**PRIDE**



Input

Raw Data **Metabolomics**



Access online database

**MetaboLights**

**Research Subjects**



Input

Raw Data **Lipodomics**



Access online database

**Zenodo**

Input Metadata

Accession Code **Z**

Accession Code **M**

Input Metadata

Input Metadata

**Input Record**

**Cov19-MSC Database**



Accession Code **P**

**Cov19-MSC Website**

V2



Input

Raw Data **Lipodomics**



Access online database

**Zenodo**



Input

Raw Data **Metabolomics**



Access online database

**MetaboLights**



Input

Raw Data **Proteomics**



Access online database

**PRIDE**

Accession Code **Z**

Accession Code **M**

Accession Code **P**

Input Metadata

Input Metadata

Input Metadata

**Subjects**

**Group**



**Cov19-MSC Database**

**Input Record**

**Cov19-MSC Website**

**Subjects Group Descriptor**

**Projects**

INPUT - OTPUT

single omics data type

<https://www.breezetree.com/articles/excel-flowchart-shapes/>

**LOADING DATA v1**



EMBL-EBI COVID-19 Data Portal



Access online databases

PRIDE



Input

Raw Data

Accession Code

Input Metadata

**Cov19-MSC Database**



**Input Record**

**Cov19-MSC Website**

**LOADING DATA v2**

**Create Input Record**

**Generate Accession Code**

**START**

**Search for Existing Code**

**New Entry?**

**yes**

**no**

**Cov19-msc**

**DB**

**STOP**

**Update Exsisting Record**

**DOWNLOADING DATA v1**

Searching



Access online databases

PRIDE

**Cov19-MSC Database**



Metadata

**Accession Code**

Accession Code

**Cov19-MSC Website**



Output

Raw Data

**DOWNLOADING DATA v2**

**Cov19-msc**

**DB**

**Input Search Criteria**

**START**

**Return list of**

**file links**

**STOP**

Visualization Record

The most important fields to visualise in the table

meeting 18/10/2022

The subject group point of view is confirmed. In term of searching operation, what will be more relevant between subject group and Project ?

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

*Sample Processing* -

Proteomics

Metabolomics

Lipodomics

SWATH proteomics

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

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group ID | Project Title | Researcher Name | Institution | Description | SampleProcessing | Sample Delivery | Protocol | Data Processing | Repositories | Links |

accession code

project title

researcher

sample delivery

description

sample processing

protocol

data processing

repository name

**Project**

group ID

group size

group description

**Subject Group**

0..3

1..\*

researcher name

researcher institution

**Researchers**

1..\*

Institution name

**Sample Delivery**

1..\*

type

**Sample Processing**

1..\*

type

**Protocol**

1..\*

type

**Data Processing**

1..\*

repos accession code

repository name

accession code

**Repositories**

1..\*

type

**Data Processing**

1..\*

type

**Protocol**

1..\*

type

**Sample Processing**

1..\*

Institution name

**Sample Delivery**

1..\*

1..\*

researcher name

researcher institution

**Researchers**

1..\*

repos accession code

repository name

accession code

**Repositories**

1..\*

project\_ID

project title

researcher

sample delivery

description

sample processing

protocol

data processing

repository name

**Project**

group ID

group size

group description

**Subject Group**

21/10/2022

Drupad gave his suggestion

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group ID | Project Title | Researcher Name | Institution | Description | SampleProcessing | Sample Delivery | Protocol | Data Processing | Repositories | Links |
| Age |  |  |  |  |  |  |  |  |  |  |
| Gender |  |  |  |  |  |  |  |  |  |  |
| oxygen |  |  |  |  |  |  |  |  |  |  |
| drugs |  |  |  |  |  |  |  |  |  |  |
| survival rate |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

group ID

group size

age

institution

survival rate

drugs

oxygen

**Subject Group**

group ID

group size

**Subject Group**

gender type

survival rate

**Gender**

age min

age max

mean age

**Age**

**Oxygen**

**Drug**

Fisical schema tool

<https://dbdiffo.com/dbdiffo.php>

Page Layouts - Omics

Navigation Bar

Data Base Rows / Search Results

Footer

Logo

**Search**

SEARCH RESULT

Navigation Bar

Footer

Logo

**Search**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Project Title | Researcher Name | Description | SampleProcessing | Sample Delivery | Protocol | Data Processing |
| *#MSC001* |  |  |  | P - M - L |  |  |  |
| *#MSC002* |  |  |  | P - M - L |  |  |  |
| *#MSC003* |  |  |  | P - L |  |  |  |
| *#MSC004* |  |  |  | P |  |  |  |
| *#MSC005* |  |  |  | M - L |  |  |  |
| *#MSC006* |  |  |  | P - M - L |  |  |  |
| *#MSC007* |  |  |  | L |  |  |  |
| *#MSC008* |  |  |  | M |  |  |  |

RECORD DETAILS

- Sample Processing Level

Navigation Bar

|  |  |  |
| --- | --- | --- |
| **Group ID** #MSC001 | | |
|  | | |
| *Proteomics* | Publication 1 | PRIDE |
| *Metabolomics* | Publication 2 | MetaboLight |
| *Lipodomics* | Publication 3 | Zenodo |

Footer

Logo

**Search**

RECORD DETAILS

- Sample Processing Level

- Raw Data Links Level

Navigation Bar

|  |  |  |
| --- | --- | --- |
| **Pride Accession Code** #PXD036103 | | |
| *File Type 1* | link 1 | file size 1 |
| *File Type 2* | link 2 | file size 2 |
| *File Type 3* | link 3 | file size 3 |
| *File Type 4* | link 4 | file size 4 |
| *File Type 5* | link 5 | file size 5 |

Footer

Logo

**Search**

Page Layouts - Group Descriptor

Navigation Bar

Data Base Rows / Search Results

Footer

Logo

**Search**

SEARCH RESULT

Navigation Bar

Footer

Logo

**Search**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Group ID* | Group Size | Female | Male | Survival Rate | Oxygen | Drug | Institution | Sample Processing |
| *#MSC001* | 1234 | 678 | 123 | 90% |  |  | yyyyyyy | P - M - L |
| *#MSC002* | 2355 | 135 | 234 | 95% |  |  | xxxxxx | P - M - L |
| *#MSC003* | 2956 | 754 | 234 | 91% |  |  | fffffffff | P - L |
| *#MSC004* | 1946 | 123 | 658 | 90.5% |  |  | wwwww | P |
| *#MSC005* | 1612 | 345 | 446 | 92.7% |  |  | aaaaaa | M - L |
| *#MSC006* | 367 | 244 | 123 | 93% |  |  | eeeeeee | SP - P - M - L |
| *#MSC007* | 1532 | 354 | 243 | 96% |  |  | yyyyyyy | L |
| *#MSC008* | 3528 | 234 | 768 | 91.2% |  |  | mmmm | M |
| *#MSC009* | 1946 | 123 | 658 | 90.5% |  |  | wwwww | P |
| *#MSC010* | 3528 | 234 | 768 | 91.2% |  |  | mmmm | M |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

RECORD DETAILS

- Sample Processing Level

Navigation Bar

Footer

Logo

**Search**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group ID** #MSC006 | | | | | | | | |
| Group Size | | Survival Rate | | Oxygen | Drug |  | | |
| 367 | | 93% | |  |  |
| Female | Male | Female | Male |  |  |
| 244 | 123 | 60% | 40% |  |  |
| Sample Processing | | Sample Delivery | | Data Processing | Protocol | Raw data files | Project Title | Researcher |
| Proteomics | | eeeeeee | | software 1 |  | 5 |  |  |
| SWATH Proteomics | | eeeeeee | | software 2 |  | 8 |  |  |
| Metabolomics | | eeeeeee | | software 3 |  | 6 |  |  |
| Lipodomics | | eeeeeee | | software 4 |  | 9 |  |  |

Navigation Bar

Footer

Logo

**Search**

|  |  |  |
| --- | --- | --- |
| **Group ID** #MSC006 | | |
| **PRIDE Accession Code** | #PXD036103 | |
| **Sample Processing** | Proteomics | |
| **Raw Data Files** | 5 | |
| File name 1 | 1.5 Gb | link 1 |
| File name 2 | 1 Gb | link 2 |
| File name 3 | 1.3Gb | link 3 |
| File name 4 | 1 Gb | link 4 |
| File name 5 | 1.2 Gb | link 5 |

Metadata

Metada are generated in order to link different types of project datafiles.

PRIDE

In PRIDE there is a standard for metadata, the Proteomics Sample Metadata Format.

*The Proteomics Sample Metadata Project aims to standardize the way ProteomeXchange partners and the proteomics community capture the relation between the* samples *and the* data *generated within a PX submission.*

The standard derives from the [MAGE-TAB](https://www.fged.org/projects/mage-tab/) extended and repurposed for Proteomics data.

<https://github.com/bigbio/proteomics-metadata-standard>

<https://www.fged.org/projects/mage-tab/>

API

PRIDE API

<https://www.ebi.ac.uk/pride/ws/archive/v2/swagger-ui.html#/>

Flask API

<https://anderfernandez.com/en/blog/how-to-create-api-python/>

Python API

<https://realpython.com/api-integration-in-python/#rest-architecture>

**ENDPOINTS - DOWNLOADING DATA**

Apparently the most important endpoint for query PRIDE are:

[**/projects/{accession}**](https://www.ebi.ac.uk/pride/ws/archive/v2/swagger-ui.html#!/projects/getProject)

<https://www.ebi.ac.uk/pride/ws/archive/v2/swagger-ui.html#!/projects/getProject>

[**/projects/{accession}/files**](https://www.ebi.ac.uk/pride/ws/archive/v2/swagger-ui.html#!/projects/getFilesByProject)

<https://www.ebi.ac.uk/pride/ws/archive/v2/swagger-ui.html#!/projects/getFilesByProject>

Web Development

Bootstrap

How to Edit, Customize, and Override Bootstrap CSS to Suit Your Brand

<https://blog.hubspot.com/website/how-to-override-bootstrap-css>

Adding CSS styling to your website

https://pythonhow.com/python-tutorial/flask/Adding-CSS-styling-to-your-website/

SVG

SVG Tutorial: How to Code SVG Icons by Hand

<https://www.aleksandrhovhannisyan.com/blog/svg-tutorial-how-to-code-svg-icons-by-hand/>

Retool

<https://retool.com/>

SQLalchemy

Data Management With Python, SQLite, and SQLAlchemy

https://realpython.com/python-sqlite-sqlalchemy/#working-with-sqlalchemy-and-python-objects

SQLite (for database connection)

https://docs.sqlalchemy.org/en/14/dialects/sqlite.html

sqlite3

https://docs.python.org/3/library/sqlite3.html

TKinter -SQLalchemy

<https://github.com/EmaSMach/tkinter-form>

<https://github.com/EmaSMach/tkinter-form/commit/6c3400f47f5bb01e654177cba01db3ece49674d3#diff-b01eeed0ec6fc4f4799e44c7b4084e649d6eee4f1aca52ac935553ff82108398>

Ajax

<https://en.wikipedia.org/wiki/Ajax_%28programming%29>

AJAX with jQuery

https://flask.palletsprojects.com/en/1.1.x/patterns/jquery/

[Html auto refresh table](https://stackoverflow.com/questions/6521968/html-auto-refresh-table)

https://stackoverflow.com/questions/6521968/html-auto-refresh-table

References

Messner et al., **Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection**, July 22, 2020, *Cell Systems 11*, 11–24, doi: <https://doi.org/10.1016/j.cels.2020.05.012>

Perez-Riverol et al., **The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences**

7 January 2022, Nucleic Acids Research, Volume 50, Issue D1, Pages D543–D552, doi: <https://doi.org/10.1093/nar/gkab1038>