ORCHESTRA

Connecting European Cohorts to Increase Common and Effective Response to SARS-CoV-2 Pandemic

WP10_D10.6,

Cohort Builder and Dashboard online

University Hospital Cologne (UHC)

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Project Classification

Project Acronym:	ORCHESTRA
Project Title:	Connecting European Cohorts to Increase Common and Effective Response to SARS- CoV-2 Pandemic
Coordinator:	UNIVR
Grant Agreement Number:	101016167
Funding Scheme:	Horizon 2020
Start:	1st December 2020
Duration:	36 months
Website:	www.orchestra-cohort.eu
Email:	info@orchestra.eu

Document Classification

WP No:	WP10
Deliverable No:	D10.6
Title:	Cohort Builder and Dashboard online
Lead Beneficiary:	UHC
Other Involved Beneficiaries:	UNIVR, SAS, CINECA, RER-ASSR, ISGLOBAL, LMU MUENCHEN, UANTWERPEN, USTUTT, CERMEL, UMCG, CINES, FCRM
Nature:	Website
Dissemination Level:	Public
Due Delivery Date:	31.05.2022
Submission Date:	26.05.2022
Justification of delay:	N/A
Status:	Completed
Version:	1.0
Author(s):	Nick Schulze, Sina Hopff, Marlene Nunnendorf, Jörg Janne Vehreschild

History of Changes

Version	Date	Created/Modified by
1.0	26.05.2022	Nick Schulze, Sina Hopff, Marlene Nunnendorf, Jörg Janne Vehreschild
2.0		

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Table of contents

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Executive summary

This document describes the deliverable number 10.6 (D10.6) of the project “ORCHESTRA cohort: Connecting European Cohorts to Increase Common and Effective response to SARS-CoV-2 Pandemic”. This task aims to create a repository of scripts and tools commonly requested. The ORCHESTRA Dashboard will allow real-time analysis of the public data set (10.4) and partner data sets (10.5). Users will be able to analyse defined risk cohorts for various outcomes and timelines without prior knowledge. The ORCHESTRA Cohort Builder will allow statisticians to select underlying and contributing risk factors and to automatize a selection of matching procedures, including propensity scores, in optimised ratios. They will receive real-time feedback on how selection affects the data set size and quality as well as ratio of matching.

Background

To ensure that the conception and implementation of the ORCHESTRA Builder and Dashboard meets the defined requirements, we started with the implementation of the ORCHESTRA public use file (PUF) which is a simplified and anonymised data set. The latter has been developed on the basis of the GECCO core data set and will be available on the ORCHESTRA portal. The implementation is described in D10.4. We used the contents of the PUF and defined appropriate research questions for the Cohort Builder and Dashboard together with the participating ORCHESTRA members. We defined key questions to the following topics (MS57):

- Influence of SARS-CoV-2 vaccination on the course of the COVID-19 disease
- Particular disease courses of the fragile population and their consequences
- Development of therapy algorithms that foster a positive outcome.

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

Theoretical implementation

Basic principles in the development of the Cohort Builder and Dashboard

To answer the defined key questions, we set up a **Cohort Dashboard** that allows real-time analyses of the data of the ORCHESTRA PUF, especially for users that are not familiar with the structure of the ORCHESTRA data set. For the creation of the Cohort Dashboard, variables are preselected by ORCHESTRA members that are familiar with the data structure by utilizing the **Cohort Builder**.

For the development, we conducted market research and investigated several possible solutions and consulted end users to receive useful insights. As a first result, we detected that no readily available software was fitting, nor was it adaptable enough to our use case, but we were able to use the research results and the user's feedback to create some basic principles:

1. Flexibility for possible future changes in data structure and data retrieval

We expect that a) the underlying data structure (table relations, variable changes, addition or removal of questions and answers etc.) might be changed according to the development of the pandemic and thus the research interests and b) the use cases targeted by this tool could be adapted over time as more knowledge is gained. Therefore, we made sure that the dashboard structure is updateable and editable in the future and these changes do not lead into breaking consistency between different versions.

2. User friendliness

A major drawback of most of the solutions studied is that little to no feedback is received on how the data is affected by the selection. Thus, the user ends up with only one final result. This leads to a huge amount of back tracking and user frustration and is harming to the overall user experience and adoption. We therefore made the data process and the in-between-results transparent and ensured continuous accessibility. This allows a comparison of different sub-

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

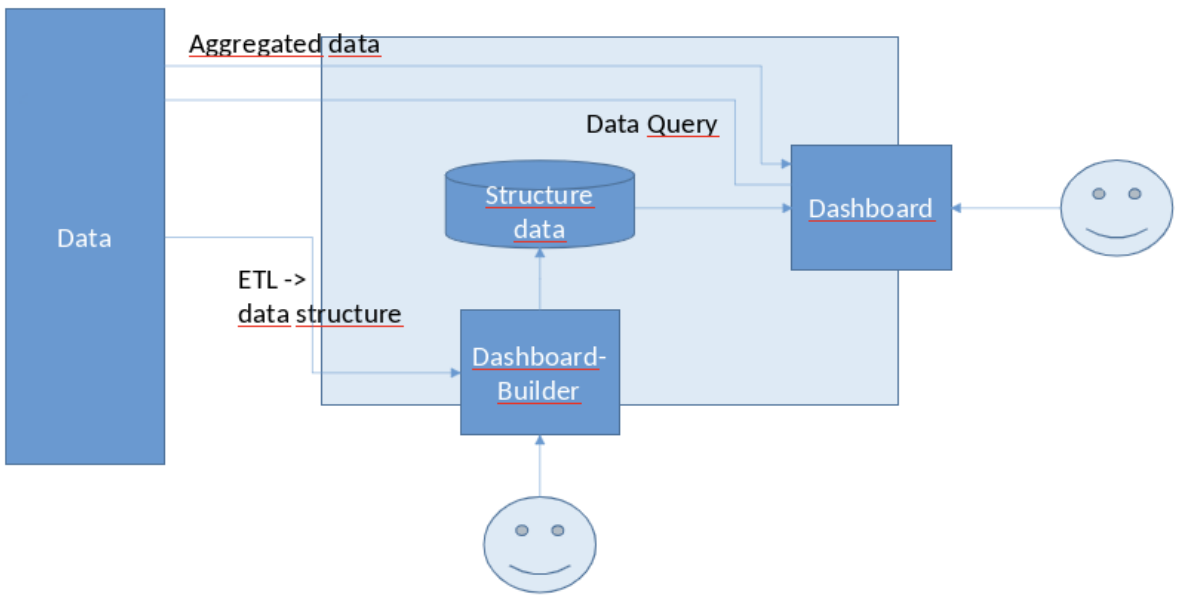
cohorts with regard to the selected criteria. Furthermore, we ensured that the tool will be available online and that an easy to use user interface is designed.

3. Ability to integrate in other projects

Another observed drawback is that a lot of the investigated solutions are (a) not well documented, (b) difficult to integrate without a complicated set-up into our IT infrastructure, (c) not price efficient, and (d) not adjustable for our use case. Therefore, we wanted to ensure that the tooling is dynamic enough to be adjusted to different use cases to make further integration and development easier. In order to achieve that goal, we decided to develop an open-source solution that is building on other open-source technologies and pursues an integrative and modular development strategy.

4. Security of data accessibility

Although direct access to the data of the PUF is possible – as it is a public available data set – most data privacy regulation will not allow an on-client data processing of medical data. Therefore, we decided to use a proxy approach. This means that the client will send a query



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

to the server which in return will send only aggregated data. The server functions as a proxy between the data and the client and the access can be strictly limited.

Design of the Cohort Dashboard

During the conception and development phase, we first drafted multiple prototypes and went through an iterative process. For the first draft of the Cohort Dashboard, we incorporated our insights from the LEOSS dashboard (dashboard.leoss.net) and decided to make the division into sub-cohorts a direct result of the user's selection. Additionally, we opted to give more insights to the users by adding descriptive visualizations like bar charts. This enables the user to make a better-informed decision on which selection he/she wants to do next.

We decided on three main parts for the cohort dashboard:

1. Variable selection

The user can select different variables (e.g. risk factors, information on patient demographics) and integrate them in the tree-node structure (see 2.).

2. Tree-node structure

The root node represents the total amount of selected patients, and the branching nodes represent sub-cohorts that have been selected by the previous variable selection (see 1.)

3. Visualisation of the variables

By selecting a node, informative visualization will be displayed (e.g. distribution of age, gender, risk factors).

ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

In a next step, we will integrate a tool set for data scientists to create and edit the dashboard structure (define new key questions, alter key questions) regarding the underlying data structure.

Technical implementation

We decided to use *dash* - a visualisation framework in python (https://www.python.org/). That framework allows an easy translation into a web application with *d3.js* and *ReactJs* components. A huge base of components are already functional available and future components can be easily integrated.

Cohort Builder

A paper prototype for the Cohort Builder was created with multiple design iteration. The creation of a dashboard starts with the import of the structure of the corresponding data.

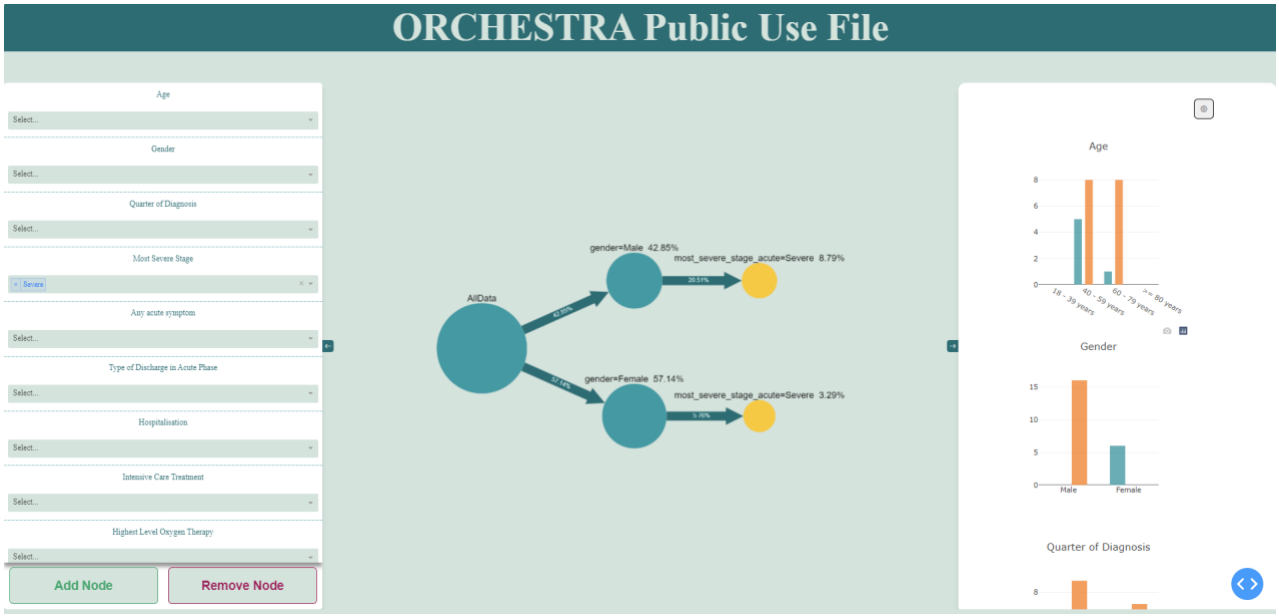
ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Afterwards, the dashboard title and the dashboard description are edited. Then, variables for the dashboard can be selected out of the imported data set columns.

The users can add information about variable names, variable descriptions, the order of selection choices, the variable type, and the visualization option. For the variable type, different types can be selected:

- Text
- Number
- Dates
- Multi-Choice
- Single-Choice

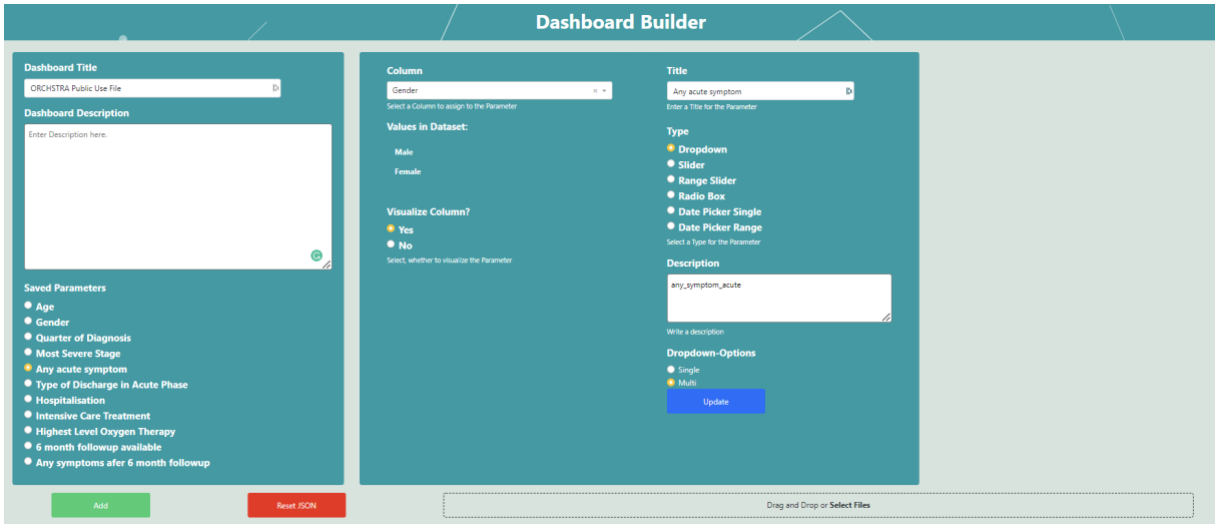
For a better understanding of the data, a preview window is available. Variables can be changed or added subsequently. Upon exiting the editing process, the dashboard is exported in JSON data format and forwarded to the cohort dashboard.



ORCHESTRA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016167

Cohort Dashboard

The Cohort Dashboard imports the JSON file which creates the inputs for the variable selection and pre-configures the variable visualisation.



1. Variable selection

Different preselected variables appear for selecting in the Cohort Dashboard. With the “add node” button, the selection is sent as a data query to the server for processing and sending back the necessary data. This includes information like the total amount of patients for each sub-cohort and distributed information for the corresponding bar charts.

User also have the possibility to select more than one variable at the same time. This works as an OR-connection, so that multiple variables can be conjugated (e.g. different age groups or risk factors).

2. Node-tree structure

After adding a new node, it displays in the node-tree structure as a new branching node. Branching nodes represent sub-cohorts of the previous node. Important information of the selected sub-cohorts displays inside the node, such as the total amount of patients and the percentage of the respective patients with regard to the previous node.

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167

By using the “remove node” button, incorrectly set notes can be removed. To help users to easily undo their selections, an action history is shown. For a friendly usage of the tree-node structure, we added the functionality to move the frame, to zoom in and out and to toggle away the variable selection and visualisation menu. This ensures that the users’ interface is clean and understandable.

3. Visualisation of the variables

The selected variables with the corresponding sub-cohorts are visualized in bar charts to help further understanding of the data. Users can move the mouse pointer over the bars to get further information. After selecting multiple nodes, the bar charts are updated and the selected sub-cohorts are shown in the visualisation. With this multi-node selection, users can immediately compare sub-cohorts based on the available parameters and are able to make more informed decisions on how to adjust deepen the variable selection.

The created graphs can be saved as a .png file via an export function to make them available for further use, for example, for discussion.

Finally, the created node-tree structure - including the created sub-cohorts - can be exported and downloaded in JSON format to import them again later. This allows to exchange the whole cohort building process with other users. It further enables better cooperation and discussion and is the basis of a sharing function.

ORCHESTRA has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 101016167