## Leveraging Clinical Data to Advance on the Clinical Management of COVID-19: Predictive Medicine - ***Sanitas* and *Grupo HM Hospitales* COVID Challenge**

The unprecedented COVID pandemic crisis imposed a new order and highlights the need for quick, innovative solutions. COVID-19 has served as a catalyst for many recent innovations in healthcare, particularly the use of real-world data, a lesson that will no doubt carry through to life after the pandemic. Researchers and leaders have to look beyond time-intensive, expensive randomized controlled trials. In order to find the answers to a question that is always evolving, the industry will need to examine information that is similarly ever-changing: Real-world data.

Real world data, defined by the FDA as data relating to patient health status or the delivery of healthcare routinely collected from a range of sources is increasingly being used in clinical decision-making. Organizations are leveraging EHR data, patient registries, and mobile device information to better understand trends and outcomes, leading to improved care delivery. During the current pandemic, this information has become all the more valuable. Real-world data can help inform leaders about anything from high-risk patient populations to the impact of measures like social distancing. Through advanced analytics techniques, visualization tools, and other data-driven methods, organizations can begin to get the most out of real-world data sources, facilitating improved care now and in the future.

Give the current state of affairs, we propose to provide the open data sets containing anonymized EHR data from thousands of COVID-19 patients from Sanitas and Grupo HM Hospitales, to provide a research platform to advance our knowledge, prediction, treatment and overall understanding of COVID-19 with the clear objective of implementing all resulting tools to come out of this challenge back into the three Hospital groups and

## The Data Sets:

**SANITAS DATA4GOOD**

Sanitas OpenDataCovid dataset contains information on treatments and medications given and data on vital signs, as well as laboratory and diagnostic test results. In addition, this information is complemented by data on patients' pre-existing conditions or illnesses that doctors have noted during admission, as these are clinically relevant when treating a patient with COVID-19. The profile of patients affected by the virus is complemented by demographic data.

All the files provided contain anonymised information on patients who have been treated at Sanitas centres and have tested positive for COVID-19. The information obtained is only associated with treatment of the virus, i.e. no previous or subsequent data from the patient's medical records is provided.

The dataset is divided into five files (demographic data, medication, vital signs, laboratory information and diagnosis) in a table structure. Each table will have an information structure that corresponds to:

**Table 1. COVID-19 patients**

This table provides data on the socio-demographic profile of the patient, together with general information about their COVID-19 treatment process, such as dates of admission to the hospital, dates of admission to the accident and emergency department, initial diagnoses, medical vital signs on admission, etc.

**Table 2. Medication**

This table provides data on all medication given to treated patients.

**Table 3. Vital signs information on the ward**

This table provides data on all vital signs collected from patients during their stay at our hospitals (ONLY ICU)

**Table 4. Laboratory information**

In the laboratory table you will find the results of the tests carried out on each patient during their stay and in A&E prior to this.

**Table 5. ISCD10 patient classification**

This table provides data on the diagnoses, procedures and neoplasms associated with the time the patient spends in our hospitals. In particular, it provides the historical data during the first 20 days of their stay.

**COVID DATA SAVE LIVES, GRUPO HM HOSPITALES**

HM Hospitales makes an anonymous dataset freely available to the international medical and scientific community with all the available clinical information on patients treated in our hospital centers for the SARS-CoV-2 virus. This clinical dataset collects the different interactions in the COVID-19 treatment process, including detailed information on diagnoses, treatments, admissions, ICU admissions, diagnostic imaging tests, laboratory results, discharge or death, among many other records.

The information in this data set comes from the HM Hospitales EHR system. It contains the anonymized records of over 3000 patients, admitted with a diagnosis of COVID POSITIVE or COVID PENDING, since the beginning of the epidemic to date. The information is organized in tables according to their content, all of them linked by a unique admission identifier. This identifier is the de-anonymization key, explicitly created for this purpose, and has nothing to do with the actual identifier of each admission.

The main table includes data on the admission and the patient (age and sex), data on the previous emergency if there has been one (2,226 records), data on their stay in the ICU if there has been one and records of the first and last set of emergency constants.

The medication table shows all the medication administered to each patient during admission (more than 60,000 records), with the dates corresponding to the first and last administration of each drug, identified by their brand name and classification in the ATC5/ATC7.

In the table of vital signs, there are all the basic records of constants (54,000 records so far) collected during admission with their date and time of registration.

The laboratory table contains the results of the determinations (398,884 records) of all the requests made to each patient during admission and in the previous emergency, if any.

And finally, the ICD10 coding tables show the records of diagnostic and procedural information coded according to the international ICD10 classification in its latest distributed version (does not include COVID), for the patients referred, both for episodes of hospital admission (more than 1,600) and for the emergency (more than 1,900) prior to those episodes, if any.

## The Challenge

With the opening of this combined dataset, we intend to take the first step and serve as an example for other institutions to be encouraged to share their information and thus, together, be able to offer the medical and scientific community clinical data with which to obtain predictive models of evolution, epidemiological models, information on the response to the various treatments applied, knowledge of virus for the creation of a vaccine, and sociodemographic data on the impact on the population of the virus.