

INFORME del GTM¹ sobre “Datos e información en la epidemia COVID-19 y propuestas para la evolución digital del sistema de salud”

Estructura del informe:

1. Resumen ejecutivo
2. Introducción y Motivación
3. Cuestiones relacionadas con la epidemia COVID-19 que requieren datos para poder ser contestadas
4. Datos e infraestructuras para el análisis de la COVID-19
5. Hacia un sistema de salud más resistente
6. Industria de la salud basada en datos
7. Recomendaciones

Apéndice 1. Ejemplos de la utilidad de entornos ricos en datos para el desarrollo de sistemas de predicción basados en métodos de Inteligencia Artificial en temas relacionados con COVID-19.

Apéndice 2: Proyectos COVID-19 en varios países

1 Resumen ejecutivo

Este informe analiza exclusivamente las infraestructuras de datos y recursos informáticos necesarios para dar respuesta a cuestiones urgentes relacionadas con la epidemia de COVID-19 sin entrar en aspectos organizativos o legales. En este sentido el informe pretende ofrecer una visión transversal a la asistencial con el propósito de ser útil para impulsar la necesaria transformación digital del Sistema Nacional de Salud.

El documento describe las necesidades y plantea las implementaciones necesarias para resolver las carencias detectadas en la organización, acceso y gestión de los datos biomédicos en la respuesta a aspectos específicos de la epidemia. Finalmente, el informe propone como objetivo la consolidación de esos desarrollos en una

¹ El Grupo de Trabajo Multidisciplinar (GTM) asesora y apoya al Ministerio de Ciencia e Innovación en materias científicas relacionadas con el COVID-19 y sus consecuencias futuras. El [GTM](#) está compuesto por: José M. Ordovás (Presidente), Mariano Esteban, Rocío García-Retamero, Beatriz González López-Valcárcel, Alfonso Gordaliza, Marco Inzitari, Pedro Jordano, Itziar de Lecuona, Laura M. Lechuga, Ramón López de Mántaras, José Molero, Agustín Portela, Diego Puga, José Javier Ramasco, Francisco Sánchez-Madrid y Alfonso Valencia. Enric Banda actúa como observador, y Maria Sol Serrano Alonso como secretaria. Todos los componentes del GTM colaboran de forma desinteresada con el Ministerio de Ciencia e Innovación. Además, en este informe han colaborado de forma desinteresada las siguientes personas: miembros del grupo “acción matemática contra el coronavirus” del Comité Español de Matemáticas (CEMat), Joaquín Dopazo (FPS Junta de Andalucía.es), Pablo Serrano (Hospital 12 Octubre SERMAS), Xavier Pastor (Hospital Clinic), Carlos Parra (Hospital Virgen del Rocío SAS), Salvador Capella y Martin Krallinger (BSC).

infraestructura estable de Medicina Personalizada alineada con los desarrollos europeos.

2 Introducción y Motivación

COVID-19, la enfermedad causada por el virus SARS-CoV-2, es la primera gran pandemia en la era de la conectividad global. Con la epidemia nos enfrentamos a un desafío. En España, tras la situación creada por la primera ola, donde la falta de conocimiento sobre la fisiopatología de la infección, y sin protocolos ni guías clínicas comunes, y con una considerable falta de recursos, las consecuencias han sido catastróficas a nivel sanitario, social y económico. Ahora nos enfrentamos a una segunda ola de la epidemia a cuya gravedad se suman las consecuencias todavía bien visibles de la primera ola.

En esta situación es urgente dar respuesta a una serie de cuestiones directamente relacionadas con la organización del Sistema de Salud, planear las líneas de investigación que a más largo plazo deben responder a los aspectos básicos de esta y otras posibles epidemias, particularmente las relacionadas con coronavirus.

De todo el conjunto de necesidades, este documento se centra en los aspectos relacionados con el acceso a datos biomédicos, incluyendo datos ómicos, socioeconómicos y médicos, junto a los desafíos que el manejo, uso y distribución de esa información representa. Esta es una cuestión que a lo largo de los meses pasados ha demostrado ser clave para la toma de decisiones en respuesta a los desafíos de la COVID-19 y que ahora se encuentra en el foco de una serie concertada de esfuerzos europeos para consolidar la infraestructura de datos y recursos informáticos tan necesaria a medio y largo plazo. Ya en septiembre de 2019 el informe comisionado por la Organización Mundial de la Salud (OMS) y el Banco Mundial realizaba una serie de recomendaciones como el desarrollo de simulacros, mayor esfuerzo en la preparación de evaluaciones de riesgo económico ante una pandemia o el fortalecimiento de los mecanismos de colaboración y coordinación internacional. Nada de esto es factible si no hay un sistema de registro de datos sólido, homogéneo y comparable y un marco común de indicadores, al menos a nivel europeo. Por todo ello, se debe trabajar en favor de la adopción de acuerdos de monitorización de datos y de análisis de situaciones de estrés sanitario, de forma similar, salvando las distancias, a los acuerdos de Basilea existentes para el sector financiero.

Hay que tener muy presente también que todo lo que aprendamos sobre los aspectos epidemiológicos de la COVID-19 será de gran utilidad en futuras pandemias, tanto la forma de recoger, gestionar y tratar los datos, como el control de la epidemia, los protocolos de actuación o la gestión interna de los hospitales cuando sus capacidades se ponen a prueba.

3 Cuestiones relacionadas con la epidemia COVID-19 que requieren datos para poder ser contestadas

Es fundamental responder a las preguntas relacionadas con la susceptibilidad y predicción del riesgo de enfermedad basado en factores pre-mórbidos. Proponer alternativas para la organización de sistemas de alerta clínica y estrategias de vacunación. También es necesario desarrollar sistemas para responder a cuestiones relacionadas con el curso de la enfermedad y las opciones de tratamiento, incluyendo posibles fármacos y protocolos. Asimismo, es importante analizar los recursos sanitarios necesarios y determinar estrategias para su gestión óptima. Finalmente, es también importante determinar el grado de riesgo de distintos colectivos y personas para orientar a las autoridades sanitarias sobre distintas políticas de salud pública.

COVID-19: cuestiones que requieren una respuesta efectiva²

a. Comportamiento epidemiológico y predicción de las consecuencias de las intervenciones poblacionales.

Es necesario desarrollar predictores siguiendo las tipologías de los modelos epidemiológicos actuales: SIR, o basados en agentes con diferentes elementos de Inteligencia Artificial. Los modelos son necesarios para modelar y explorar la viabilidad, aceptabilidad, y potencial adherencia a distintas estrategias para reducir la propagación del SARS-CoV-2.

Los modelos deben tener en cuenta componentes como, patrones de viaje, distribución de la población, situaciones habitacionales y situación socioeconómica en general, áreas sanitarias e incidencia de la epidemia (hospitalizaciones, números en UCI, etc.), así como la información y modelos desarrollados en epidemias previas. Estos datos son fundamentales para enfrentar los principales problemas inmediatos referidos a limitar la transmisión, gestionar la carga hospitalaria, recursos de análisis (PCR, test, imagen médica) y mejorar el manejo clínico de los pacientes, traduciéndolos a términos prácticos de recursos materiales y humanos.

² La Convocatoria de Expresiones de Interés de proyectos de investigación frente al virus SARS-CoV-2 y la enfermedad COVID-19, que gestiona el Instituto de Salud Carlos III, como Agencia de Financiación de Investigación en Biomedicina y dependiente del Ministerio de Ciencia e Innovación (MCIN) (dotada con cerca de 25 Millones Euros) ha financiado 130 proyectos de investigación y otros posteriores de varias CCAA y fundaciones, proyectos que atienden directamente a varios de los puntos sobre necesidades médicas urgentes mencionados en este apartado, incluyendo los puntos a-f y j. Estos proyectos, en especial los financiados por el MCIN fondos/COVID, incluyen la generación de cohortes prospectivas y el Estudio Nacional Sero-epidemiológico (ENE-COVID), y deben proveer una primera línea de respuestas a estas preguntas urgentes, aun siendo conscientes de que muchas de ellas continuarán siendo temas científicos abiertos. Otros puntos mencionados en ese apartado, incluyendo el referido al desarrollo de criterios científicos para la priorización de las vacunas (g), es responsabilidad de los organismos competentes.

b. Identificar los principales factores de riesgo en el momento del diagnóstico³.

Incluyendo la predicción de fenotipos de la enfermedad definidos sistemáticamente (neumonía asintomática, leve y grave, SDRA, tromboembolismo, respuesta inmunoinflamatoria, etc.) y de determinantes de la evolución de la enfermedad (riesgo de readmisión, pautas de gestión y optimización de Unidades de Cuidados Intensivos, etc.) con objeto de contribuir a sistemas de soporte a la decisión clínica que incluyan aspectos como tipo de tratamiento farmacológico, ventilatorio, etc.

La identificación de factores de riesgo está asociada directamente con el desarrollo de sistemas de predicción de curso de la enfermedad como soporte a la organización y respuesta de los sistemas de salud.

c. Identificar los factores predictores de los efectos tardíos de la enfermedad^{4,5}.

Incluyendo efectos secundarios a niveles: función cardiovascular y de vasculatura periférica, coagulopatías, SNC, función renal y hepática entre otros. Con especial atención a trastornos de salud mental y los aspectos sociodemográficos y socioeconómicos relacionados.

d. Encontrar la respuesta terapéutica y resultados clínicos según condiciones anteriores (comorbilidades y tratamientos)^{6,7,8}.

Evaluar la efectividad de los medicamentos y tratamientos, estratificados por niveles de atención (atención primaria y hospitalización), fases de la enfermedad (prehospitalario, pre-UCI e UCI) y poblaciones, para identificar el riesgo/beneficio de las intervenciones terapéuticas. Desarrollo de una estrategia efectiva para la evaluación de los métodos.

e. Determinar los factores de riesgo genético en humanos y procedentes de la genética del virus mediante la aplicación de metodologías genómicas.

Se requieren estudios con suficiente capacidad estadística (tamaño poblacional y representatividad) y la posibilidad de organizar cohortes y muestras, así como la recopilación masiva de genomas virales para el desarrollo de modelos de transmisión viral para trazar la situación epidemiológica. Análisis de variantes y mutaciones del SARS-CoV2.

³ Liang, W., Liang, H., Ou, L., et al. (2020) Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern Med.*

⁴ Wu, Y., et al. (2020) Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain, Behavior, and Immunity.*

⁵ Morlacco, A., et al. (2020) The multifaceted long-term effects of the COVID-19 pandemic on urology. *Nature Reviews Urology.*

⁶ Flaxman, S., Mishra, S., Gandy, A., et al. (2020) Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature.*

⁷ Wu, X., et al. (2020) Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study. *medRxiv.*

⁸ Richardson, S., et al. (2020) Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA.*

f. Determinar Poblaciones Vulnerables.

Abordar preguntas relacionadas con el riesgo de infección, la gravedad de la enfermedad en el momento de la presentación y resultados a largo plazo en poblaciones vulnerables específicas, incluidos ancianos y personas que viven en centros de salud, población con comorbilidades (hipertensión, obesidad, diabetes, EPOC, cáncer, pacientes con trasplantes, etc.), poblaciones inmunodeprimidas, mujeres embarazadas y niños, personas con desventajas socioeconómicas y grupos minoritarios. La identificación de estas últimas poblaciones requiere de muchos datos y mucha información para poder analizar las correlaciones y calibrar modelos específicos.

g. Desarrollar criterios científicos para la priorización de las vacunas.

Estos criterios deberán ser modelados teniendo en cuenta las características de las futuras vacunas (e.g. posible riesgo de inducir neumonitis), junto a los datos derivados de los correspondientes ensayos clínicos en relación con su eficacia a la hora de frenar la transmisión, y en particular, la transmisión entre la población más vulnerable.

- i. Riesgo de complicaciones en personas altamente vulnerables y estratégicas.
- ii. El nivel de seroprevalencia a COVID-19 y otros tipos de virus (es decir, adenovirus) será un factor importante a tener en cuenta en la preparación para una futura campaña de vacunación.
- iii. Los aspectos éticos y procedimientos de laboratorio, la evaluación de los proveedores y las estimaciones económicas, representaran los factores de contorno a considerar.
- iv. La identificación de grupos de individuos que puedan considerarse “altos propagadores” en función de su actividad o movilidad.

Es especialmente relevante tener en cuenta las peculiaridades del sistema inmune de las personas mayores en el diseño de vacunas, ya que las personas mayores no responden tan bien a ellas como los jóvenes. Este aspecto es muy importante para el diseño de vacunas y la identificación de los adyuvantes adecuados para potenciar las respuestas de anticuerpos neutralizantes, así como las respuestas protectoras de linfocitos T CD4 y, sobre todo, de CD8 citotóxicos antivirales. La combinación de varios adyuvantes que actúen sinérgicamente sobre múltiples vías (por ejemplo, varios receptores tipo TOL) y reactiven la inmunidad celular, es esencial en los protocolos de vacunación de la población mayor dada su reducida capacidad de respuesta a una vacuna general para toda la población.

h. Sistemas de seguimiento de contactos y cumplimiento de cuarentenas

- i. Seguimiento manual. Dificultades y tecnología subyacente. Problemas organizativos y requerimientos de personal especializado. Integración con los sistemas de atención primaria y testeo. Límites del sistema, costes y sostenibilidad.
- ii. APPs basadas en teléfonos móviles (bluetooth). Modelos centralizados y descentralizados, solo el sistema centralizado es útil para el sistema de salud

público. Sistemas descentralizados y dependencia tecnológica de las grandes compañías de telefonía. Problemas técnicos y éticos asociados al uso de Apps. Integración con los sistemas de trazado de casos de los sistemas de salud. Aceptación pública y evaluación de la eficacia en distintas condiciones, e.g. densidad de población.

- iii. Determinación de establecimiento y duración de los periodos de cuarentenas tanto en personas sospechosas de COVID como preventivas de los contactos, y en particular en poblaciones esenciales con alto grado de contacto.

i. Predicción de recursos sanitarios, desarrollo de indicadores de riesgo y logística hospitalaria y extra-hospitalaria

Los modelos epidemiológicos referidos en el apartado a han demostrado su utilidad mediante la generación de distintos escenarios, para la evaluación de la eficacia de distintas medidas de distanciamiento social, restricciones de movilidad, aplicación de tests y rastreo de contactos en la expansión del virus. Sin embargo, los servicios de salud necesitan conocer, además de la prevalencia e incidencia de la pandemia en la población, la cantidad de recursos críticos necesarios (EPIs, respiradores, ...) para proporcionar atención médica a los pacientes en el futuro más inmediato. La disposición de tales previsiones facilita la planificación de dichos recursos, como pueden ser las camas hospitalarias, las camas UCI, personal médico y equipamiento médico especializado. También es necesario anticipar las necesidades de personal y material de análisis (reactivos para PCR, diagnóstico por imagen, personal de laboratorio...), personal de rastreo y de atención primaria junto con medios materiales (líneas de centrales telefónicas...). Es en esta labor de previsión de los recursos necesarios donde los modelos de la investigación operativa, basados en el análisis estadístico, la simulación y la optimización, pueden prestar una gran ayuda a los responsables de la planificación y de la logística hospitalaria^{9,10}. Estos modelos permiten obtener indicadores de riesgo de saturación de los recursos sanitarios facilitando la gestión óptima de recursos limitados.

j. Nuevas áreas de investigación a desarrollar y generación de nuevos datos.

Más allá de las cuestiones antes mencionadas, seguimos necesitando una mejor comprensión de la enfermedad. Un conocimiento que será necesario para controlar su progresión y consecuencias, y seguirá demandando una infraestructura de datos adecuada. Entre ellos: Genética viral, determinantes de las manifestaciones clínicas de la infectividad viral y marcadores clínicos de carga viral; Genómica humana, biomarcadores moleculares de la expresión génica y estudios proteómicos y marcadores genómicos (“deep phenotyping”); Marcas epigenéticas, inmunogenética, biomarcadores de la respuesta inmune innata, marcadores de inmunidad específica anti-SARS-CoV-2, tipificación sistemática de HLA en proyectos genómicos humanos;

⁹ Currie, C.S.M., Fowler, J.W., Kotiadis, K., et al. (2020). How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation*, 14(2), 83–97.

¹⁰ Daniel García-Vicuña, Marta Cildoz, Martín Gastón-Romeo, Cristina Azcarate, Laida Esparza and Fermín Mallor (2000). Operations Research Helps Public Health Services Managers Planning Resources in the COVID-19 Crisis. *BEIO*, 127-151.

Interacción de COVID-19 y otras enfermedades, COVID y enfermedades raras, cáncer y otras enfermedades respiratorias (principalmente gripe); hospitalización de pacientes, con tiempos de estancia, tratamientos recibidos y evolución clínica. Todos ellos son temas de investigación que se basarán en datos existentes y que generarán nuevas cantidades de datos masivos (por ejemplo, datos de secuenciación o proteómica).

4 Datos e infraestructuras para el análisis de la COVID-19

La Comisión Europea y muchos países de la UE han lanzado convocatorias de financiación de proyectos de investigación para proporcionar soluciones a las cuestiones planteadas más arriba sobre prevención, protección y tratamiento de COVID-19.

La urgencia de estos problemas hace razonable pensar que deben resolverse en base a cohortes retrospectivas utilizando los datos y conocimiento obtenido durante los últimos meses, más todo el legado de medicina y datos socioeconómicos. También es razonable pensar que el análisis de estos datos requerirá de los entornos computacionales adecuados.

Analizamos el escenario actual en cuanto a la disponibilidad de datos sobre COVID-19 e infraestructuras para su análisis.

4.1 Información Genética y Genómica

a. Datos sobre SARS-CoV-2

- i. Los datos sobre SARS-CoV-2 deben recogerse en un sistema abierto y público, sin restricciones de licencias comerciales. Es fundamental recoger toda la información, tanto de los datos de secuencias en bruto como la meta-información asociada y utilizar protocolos de ensamblado y detección de variantes estandarizados que proporcionen una calidad adecuada.
- ii. GISAD¹¹, provee un sistema popular de recogida, análisis y visualización de datos. Desafortunadamente las restricciones sus licencias que hace que la distribución no sea abierta hacen que, a pesar de su utilidad, su uso sea cuestionable. Además, no proporciona las secuencias originales, ni facilidades para la reconstrucción de los genomas virales, lo que hace inviable estimar la calidad de los datos o calcular las posibles variantes menores. La alternativa ofrecida por la base de datos ENA del EMBL-EBI resuelve estas limitaciones.
- iii. Acceso a datos genómicos virales de proyectos nacionales (habrá aproximadamente 10k secuencias virales en el proyecto del Instituto de Salud Carlos III¹²). Los sistemas de visualización pandémica deben ser centralizados y ampliamente utilizados para permitir una comprensión inmediata de las

¹¹ <https://www.gisaid.org/>

¹² "Addressing unknowns of COVID-19 transmission and infection combining pathogen genomics and epidemiology to inform public health interventions" Proyecto financiado por el ISCIII - fondos COVID <http://seqcovid.csic.es/collaborating-institutions/>

consecuencias de los nuevos hallazgos (por ejemplo, la mutación D614G en la proteína Spike que aumenta la infectividad del SARS-CoV-2¹³).

- iv. La secuenciación de patógenos debe declararse de interés clínico para aumentar los datos del genoma viral disponibles.
- v. Relación entre la gripe y COVID-19. Habrá una interacción inevitable (pero aún desconocida) entre la gripe y COVID-19. Se recomienda la secuenciación de patógenos relacionados (ver desarrollo de estrategias de vacunación). Se requiere nueva experimentación animal en modelos de primates de co-infección SARS-CoV-2 y otros virus como el de la gripe)

b. Datos sobre casos COVID-19 humanos

- i. EGA (European Genome-phenome Archive)¹⁴ representa la principal fuente europea de información genómica. EGA es una base de datos clasificada de esencial (“core” en la clasificación de ELIXIR) en la que se almacenan genomas humanos de importancia estratégica (por ejemplo, los genomas de cáncer derivados del proyecto International Cancer Genome Consortium, ICGC) junto a las anotaciones sobre su origen y procedimientos utilizados para su obtención. El acceso a EGA está regido por un sistema coordinado de comités éticos asociados a cada proyecto y conjunto de datos. EGA incorpora progresivamente un sistema de autenticación único promovido por ELIXIR¹⁵.
- ii. EGA evoluciona hacia un modelo federado en el que cada país o agrupación dispone de un sistema de almacenamiento EGA propio accesible solo a sus afiliados, mientras que puede compartir la información agregada con los otros repositorios EGA federados.
- iii. EGA federada se complementa con un sistema de “beacons” promovido por la Global Alliance for Genomic Health (GA4GH) que permite detectar información genómica en los repositorios EGA federados, posibilitando la creación de cohortes virtuales.
- iv. Sistemas centralizados ya en funcionamiento, siguiendo el modelo de biobanco del Reino Unido¹⁶, representan la alternativa ahora mismo más factible para la realización de estudios poblacionales (GWAS¹⁷). De hecho, los primeros estudios apuntan a un componente genético en la respuesta al virus. Para confirmarse estos estudios requieren de un número mayor de casos con más variabilidad genética de fondo y mejores anotaciones sobre la enfermedad.
- v. Las iniciativas genómicas nacionales¹⁸ (Scourge) y locales (p. ej., Nagen¹⁹, EnoD de CIBERER²⁰, entre otras) deben contribuir con datos genómicos y resultados clínicos a los proyectos globales.

¹³ Korber B, et al. (2020). Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell*.

¹⁴ <https://ega-archive.org>

¹⁵ European AAI authentication system (<https://elixir-europe.org/services/compute/aa1>)

¹⁶ <https://www.ukbiobank.ac.uk>

¹⁷ https://en.wikipedia.org/wiki/Genome-wide_association_study

¹⁸ “Determinantes Genéticos y Biomarcadores genómicos de riesgo en pacientes con infección por coronavirus SARS-COV-2”. Proyecto financiado por el ISCIII - fondos COVID.

¹⁹ <https://www.nagen1000navarra.es/en/home>

²⁰ <https://www.ciberer.es/noticias/el-programa-de-enfermedades-no-dagnosticadas-explicado-en-el-hospital-virgen-de-la-arrixaca-de-murciab>

4.2 Datos de imágenes: colecciones existentes de imágenes covid-19

Está en marcha un número de proyectos para la caracterización de imágenes tanto a nivel local como internacional. En general estos proyectos utilizan variantes de aprendizaje automático o “*Machine Learning*” entrenadas con conjuntos de imágenes de COVID-19 anotadas (informes de radiología).

A nivel europeo²¹ y local se dispone de repositorios centralizados de imágenes (muchas comunidades de Cataluña, Valencia²², y Andalucía tienen repositorios centralizados²³). Estos repositorios son necesarios para entrenar los sistemas de predicción basados en AI/ML. Las imágenes deben estar asociadas a anotaciones adecuadas y de calidad, y encontrarse en entornos computacionales que permitan el costoso entrenamiento de estos sistemas.

El área adolece de una falta de estándares de calidad. Iniciativas Europeas por parte de la plataforma AI4EU y locales (AQUAS en Catalunya²⁴) se encaminan a establecer un marco de evaluación técnica estable, incluyendo criterios de evaluación y conjuntos de prueba. Es necesario fomentar los proyectos que combinen el análisis de imágenes para identificar regiones que se puedan usar como descriptores fenotípicos (imágenes como fenotipos) y vincularlas a datos genómicos y médicos.

4.3 Organización Europea

- i. Europa se ha dotado de una serie de infraestructuras en el área de biomedicina que incluyen: BBMRI biobancos, EATRIS medicina translacional, ECRIN ensayos clínicos y ELIXIR- bases de datos y bioinformática alineados con la European Open Science Cloud (EOSC). EOSC, y especialmente con EOSC-Life, ya está jugando un papel importante en la organización de datos para la investigación biomédica.
- ii. En el área de datos médicos y de digitalización las infraestructuras Europeas incluyen Euhealth Data Space²⁵ y Digital Europe²⁶.
- iii. La comisión ha lanzado una serie de proyectos específicos COVID-19, incluyendo proyectos H2020 y proyectos IMI (Innovative Medicament Initiative²⁷), así como extensiones de proyectos existentes. Junto a una larga lista de proyectos financiados en cada uno de los países (ver apéndice 2).
- iv. La iniciativa europea 1+M Genomes²⁸ propone las bases técnicas, éticas y legales para la colaboración en la generación, anotación y análisis de datos de genomas humanos a nivel europeo, con un paquete de trabajo completo dedicado a la estandarización y normalización de datos clínicos.

²¹ <https://www.eurobioimaging.eu>

²² <https://bimcv.cipf.es>

²³ <http://pibicraproject.com/>

²⁴ <https://aquas.gencat.cat/ca/inici>

²⁵ https://ec.europa.eu/health/ehealth/dataspace_es

²⁶ <https://www.digitaleurope.org/>

²⁷ <https://www.imi.europa.eu>

²⁸ <https://ec.europa.eu/digital-single-market/en/european-1-million-genomes-initiative>

- v. La iniciativa de genética de la COVID-19²⁹ es una iniciativa de colaboración internacional para compartir recursos para la investigación de los determinantes genéticos de la susceptibilidad, la gravedad y los resultados de la COVID-19).
- vi. Epidemiología de la pandemia del coronavirus. El consorcio (COPE) tiene como objetivo involucrar a expertos en el desarrollo de una aplicación personalizada de seguimiento de síntomas COVID-19 que funciona como una plataforma de captura de datos en tiempo real³⁰ que obtuvo más de 2.8 millones de usuarios en unos pocos días.
- vii. Además, datos desglosados por sexo de COVID-19 son recopilados por la Global Health 50/50³¹, una iniciativa ubicada en el University College de Londres que aboga por la equidad de género.

4.4 Datos médicos.

a. Acceso a la información.

Existen problemas logísticos, económicos, éticos y legales que hacen que el manejo de información de EHR (“Electronic Health Records”) dentro y fuera de los hospitales sea complejo, y poco efectivo en la práctica. Estas dificultades inciden negativamente tanto en los aspectos asistenciales como de investigación y sin duda restan valor al conjunto de información disponible en el sistema. La situación que ha generado la Covid-19 ha supuesto un cambio importante. Por ejemplo, Hospitales HM ha hecho públicamente accesible parte de esa información (anonimizada y referida a variables codificadas³²) y se han logrado acuerdos de colaboración entre grandes hospitales y grupos de investigación para analizar información³³ de EHR sobre COVID-19, siempre muy restringido y con cierto grado de discrecionalidad.

b. Estandarización de la información médica.

A las dificultades de acceso a la información se superponen los problemas de normalización de los datos de diversa procedencia que manejan los sistemas de salud. De hecho, muy pocos hospitales disponen de un sistema central de información (“*data lakes*”) y menos aún de un sistema para la extracción de variables normalizado (FIHR - Fast Healthcare Interoperability Resources- o similares) que permita obtener información de modo sistemático e interoperable. Un ejemplo de la situación actual puede ser la selección de pacientes para ensayos clínicos, que se realiza por equipos que buscan en cada EHR pacientes con las características deseadas, en vez de estar asistidos por eficientes sistemas de búsqueda. Es importante tener en cuenta los importantes retos que la incorporación de nuevos tipos de datos representará, incluyendo dispositivos personales conectados (“Internet of Things”) e información en redes sociales y redes de pacientes.

En el caso de la COVID-19 hay pocos hospitales que pueden proveer el conjunto de datos de ISARIC (International Severe Acute Respiratory and Emerging Infection

²⁹ <https://www.covid19hg.org/>

³⁰ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7357669/>

³¹ <http://globalhealth5050.org/covid19/>

³² <https://www.hmhospitales.com/prensa/notas-de-prensa/comunicado-covid-data-save-lives>

³³ <https://www.clinicbarcelona.org/noticias/el-clinic-y-el-bsc-utilizaran-la-inteligencia-artificial-para-predecir-la-evolucion-de-pacientes-de-covid-19>

Consortium³⁴), que se considera el mínimo conjunto de información necesaria y pocos de los biobancos creados disponen de modo automático de la meta-información asociada a las muestras.

Algunos ejemplos de sistemas de información en España incluyen la Base Poblacional de Salud andaluza³⁵ ³⁶ que contiene la información histórica de 12,5 millones de usuarios del sistema de salud andaluz. El Instituto Aragonés de Ciencias de la Salud dispone de una plataforma de explotación de Big data clínico. En Valencia (FISABIO) explotan datos clínicos de toda la población, enlazando varias fuentes (CMBD, prescripciones, consultas extra-hospitalarias, ...) y sistema de información ABUCASIS que conecta los centros de atención primaria. El sistema catalán PADRIS (Programa D'Analítica de Dades per a la Recerca i la Innovació en Salut³⁷) dispone de una base de datos agregada de los EHR de toda la población, incluyendo los informes de atención primaria y dispensación farmacéutica. Sin embargo, estos sistemas tienen un decalaje importante precisamente por la falta de sistemas que en tiempo "cuasi real" puedan poner los datos clínicos el formato adecuado para su análisis y por tanto deben asociarse a sistemas que faciliten su carga en las correspondientes plataformas analíticas. Las alternativas actuales representan una alternativa con soluciones específicas para el acceso directo a los datos relevantes, particularmente EHR (Real World Data y Real World Evidence³⁸), complementadas con aproximaciones computacionales para la extracción de variables a partir de los textos escritos con técnicas de procesamiento de lenguaje natural y minería de datos.

c. Confidencialidad de la información.

Para que el sistema sea adecuado debe partir de una adecuada pseudo-anonimización de los datos con todas las garantías, no sólo técnicas sino también ético-legales, siguiendo las pautas del Reglamento de Protección de Datos con las correspondientes guías tanto éticas como técnicas sobre pseudo-anonimización. La sostenibilidad y garantía del sistema requiere que estos procesos se realicen bajo el control de las autoridades correspondientes, correspondan a una regulación proactiva que introduzca la evaluación del impacto relativa a los tratamientos de datos personales y se base en software libre de uso común. También es necesario realizar un análisis de riesgo de re-identificación previo a aplicar la analítica sobre los datos, incluyendo la incorporación de sistemas de procesamiento del lenguaje para detectar patronímicos y relaciones nominales de personas embebidos en los textos clínicos. Finalmente, la normalización de los datos es imprescindible para permitir la segregación de los datos demográficos de los datos clínicos³⁹.

³⁴ <http://isaric.tghn.org>

³⁵ <https://www.sspa.juntadeandalucia.es/servicioandaluzdesalud/profesionales/sistemas-de-informacion/base-poblacional-de-salud>

³⁶ <https://www.scielosp.org/pdf/gs/2020.v34n2/105-113/es>

³⁷ <https://aquas.gencat.cat/ca/ambits/analitica-dades/padris/>

³⁸ https://en.wikipedia.org/wiki/Real_world_evidence

³⁹ Somolinos R, et al. Sistema anonimizador conforme a la norma UNE-EN ISO 13606. Conference: XXX Congreso Anual de la Sociedad Española de Ingeniería Biomédica. Volume: 1. Project: PITES: Plataforma de Investigación en Tecnologías de la Información para e-Salud. 2012
Cáceres J et al. (2013) La norma UNE EN ISO13606 como soporte a Grupos de Investigación. Informed.

d. Respuesta terapéutica y resultados clínicos.

En España durante la primera oleada hubo unas 120.000 hospitalizaciones. El acceso a esos registros médicos que incluyen las intervenciones y procedimientos durante el curso de la hospitalización es un recurso imprescindible para evaluar los resultados de dichos procedimientos e intervenciones y también para la elaboración de modelos de predicción del curso clínico de la enfermedad.

La disparidad técnica y las limitaciones de los sistemas antes mencionadas, unidas a las difíciles circunstancias durante el pico de la pandemia, hacen necesario un esfuerzo de curación considerable. En la práctica, parece aconsejable comenzar por desarrollar los sistemas en colaboración con los hospitales que están más preparados para generar la información necesaria, con la intención de extenderlos posteriormente al resto del sistema. Los primeros resultados de estas colaboraciones, obtenidos entrenando sistemas basados en Ciencia de Datos y Machine Learning con variables estandarizadas extraídas de los expedientes de pacientes de COVID-19 arrojan resultados prometedores⁴⁰.

Por otra parte, un número que puede aproximarse a 1.4 millones de casos (deducido de la encuesta INE+ISCIII de sero-prevalencia) han pasado la enfermedad atendidos por los servicios de atención primaria. Sería extremadamente interesante poder utilizar también esta información en los modelos de predicción de riesgo y gravedad, así como en los estudios de tratamientos/síntomas.

e. Situación española Proyecto de Historia Clínica Digital del Sistema Nacional de Salud (HCDSNS)

ISARIC ha desarrollado una propuesta de estándar de datos mínimos para COVID-19. Estas variables incluyen observaciones epidemiológicas, síntomas, observaciones clínicas y datos de laboratorio. Distintos hospitales y sistemas de salud están adaptando sus sistemas de información para la generación automática de estas variables a partir de las distintas fuentes de información de cada hospital, aunque desafortunadamente sin los recursos e impulsos que cabría esperar.

El objetivo técnico es desarrollar bases de datos en estándar FIHR de la que sea posible extraer externamente las variables requeridas. Distintos proyectos piloto se proponen resolver los problemas técnicos relacionados con este procedimiento.

Como el nivel de automatización en los hospitales y sistemas de salud es muy desigual, y además los sistemas informáticos son muy distintos, será necesario un esfuerzo considerable de digitalización e integración específico de cada caso para poder proveer ese tipo de base de datos intermedia en formato FIHR.

Existe un esfuerzo europeo por desarrollar un vocabulario común de síntomas traducido a todos los idiomas. España participa en la incorporación de castellano y catalán y en paralelo se están desarrollando la conexión de los términos específicos con sus equivalentes en las ontologías médicas.

Más allá del conjunto de variables normalizadas codificadas en la práctica es necesario obtener información adicional de las distintas fuentes de información médica. Distintos proyectos están en marcha para la extracción de información utilizando técnicas de

⁴⁰ Carrino, et al., Towards predicting the evolution of covid-19 mortality risk: A recurrent neural network approach. BioRxiv

procesamiento del lenguaje natural para este fin⁴¹. Es especialmente relevante el desarrollo de sistemas de comparación de métodos de procesamiento del lenguaje natural) sobre “Electronic Health Records” (Registros de Salud Electrónicos, RSE en español) basados en la disponibilidad de cantidades suficientes de textos anotados por expertos.

En España, el ISCIII coordinará un registro común de una parte importante de los datos generados en el marco de los proyectos financiados por el Fondo COVID-19. Esos proyectos incluyen una cláusula en la que se establece la necesidad de depositar los datos en un registro centralizado gestionado por el ISCIII. Ese registro incluirá un conjunto mínimo de datos, estandarizado y alimentado con la información de los proyectos financiados, bajo oportunas condiciones ético-legales. Será posible que otros proyectos relacionados, puedan adherirse al mismo registro depositando allí la información obtenida.

4.5 Cohortes y biobancos

Junto a la información de EHR las cohortes poblacionales, biobancos y ensayos clínicos^{42 43}, son fuentes importantes de información que deben de hacerse accesible. En particular, los estudios en cohortes poblacionales seguidas en el tiempo (estudios longitudinales) son particularmente relevantes.

Las infraestructuras ELIXIR y BBMRI coordinan la estructura de datos y la capacidad para descubrir datos de interés en los repositorios sin revelar la identidad de los pacientes (sistemas conocidos como “beacons”⁴⁴).

Tanto los nodos nacionales de estas infraestructuras, los CIBERS y los proyectos COVID-19 financiados por el ISCIII trabajan para cubrir las necesidades básicas de información implementando las recomendaciones y estándares europeos de compartición de datos.

4.6 Registro nacional y locales de datos epidemiológicos

El Centro Nacional de Epidemiología, del ISCIII, mantiene bases de datos y análisis con metodología homogénea con organizaciones europeas (Centro Europeo de Control de Enfermedades, ECDC⁴⁵) relacionados con la COVID, como el MoMo (Monitorización de Mortalidad⁴⁶ diaria por todas las causas, capaz de detectar excesos de mortalidad con procedimientos estadísticos) y el Estudio Nacional Sero-epidemiológico⁴⁷ un amplio estudio longitudinal sero-epidemiológico, de base poblacional, cuyos objetivo es estimar la prevalencia y evolución temporal de la infección por SARS-Cov2 en España mediante la determinación de anticuerpos frente al virus.

La base de datos de la Red Nacional de Vigilancia Epidemiológica (RENAVE) contiene información poblacional de interés epidemiológico de todos los casos diagnosticados con COVID-19 en España a partir del inicio de la epidemia.

⁴¹ <https://www.plantl.gob.es/Paginas/index.aspx>

⁴² <https://www.isciii.es/QueHacemos/Servicios/BIOBANCOS/Paginas/RegistroNacionalBiobancos.aspx>

⁴³ <https://redbiobancos.es/actualidad/covid-19/>

⁴⁴ <https://elixir-europe.org/news/elixir-beacons-facilitate-sharing-genomics-data-europe>

⁴⁵ <https://cnecovid.isciii.es>

⁴⁶ <https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/MoMo/Paginas/MoMo.aspx>

⁴⁷ <https://portalcne.isciii.es/enecovid19/>

Asimismo, las comunidades autónomas, a través de las correspondientes consejerías con competencias en sanidad, llevan registros con datos sobre la evolución de la enfermedad en cada uno de los ámbitos regionales. Estos registros se encuentran en muy distintos grados de desarrollo y operatividad y, como se está pudiendo constatar a lo largo de los meses transcurridos por la pandemia, la falta de interoperabilidad y la heterogeneidad de criterios en el diseño de estos sistemas están haciendo imposible la disponibilidad de datos fiables a nivel nacional.

Desde la perspectiva de la Investigación en los Servicios Sanitarios, se han desarrollado proyectos muy potentes como el ATLAS de Variaciones en la Práctica Médica (VPM), el cual permite comparar utilización estandarizada de utilización de distintos tipos de servicios sanitarios, incluyendo los de escaso valor terapéutico, por zonas de salud⁴⁸.

4.7 Datos de hospitalización de pacientes.

El desarrollo de modelos logísticos para la predicción de recursos sanitarios precisa para su calibración de datos asociados a la hospitalización de pacientes COVID-19 tales como fecha de ingreso y de alta tanto en el hospital como en la UCI, intervenciones y tratamientos proporcionados, evolución clínica, etc y características de los pacientes que permitan ser utilizadas como pronósticos de dichas estancias y de los recursos necesarios (edad, sexo, comorbilidad, ...). Toda esta información se recoge de forma ordinaria en los registros electrónicos del hospital. Se puede mencionar la Sanidad Pública de Navarra como caso de éxito, que diariamente ha facilitado estos registros, convenientemente anonimizados, para alimentar modelos de previsión de ocupación de camas y saturación de instalaciones cuyos resultados han sido utilizados durante la primera ola y actualmente en la segunda ola en la toma de decisiones concerniente a la organización de los recursos (preparación de camas para hospitalización y de UCI). También La Rioja está participando en un proyecto similar a través de la Fundación Rioja Salud.

4.8 La situación española con respecto a los conjuntos de datos COVID-19

El Instituto Nacional de Bioinformática⁴⁹ es una plataforma del Instituto de Salud Carlos III constituida por nodos distribuidos en el país y coordinados por el nodo central en el Barcelona Supercomputing Center- Centro Nacional de SuperComputación⁵⁰. El INB actúa como nodo nacional de la infraestructura europea ELIXIR, como coordinador de las actividades de la red de bioinformáticos en Institutos de investigación de los hospitales españoles (TransBioNet⁵¹).

El BSC, como coordinador de la Plataforma INB, participa como principal enlace nacional en la 'European COVID-19 Research Data Platform' el ISCIII participa como 'alternate', así como principal agencia financiadora nacional al respecto. La participación española se coordina mediante un Grupo de Trabajo integrado por BSC, ISCIII y Ministerio de Ciencia e Innovación.

⁴⁸ <https://www.atlasvpm.org>

⁴⁹ <https://inb-elixir.es/>

⁵⁰ <https://www.bsc.es/>

⁵¹ <https://inb-elixir.es/transbionet>

Es importante destacar que estos esfuerzos de coordinación y desarrollo están encaminados a establecer una infraestructura bioinformática sostenible. En este sentido es importante mencionar el soporte que estos desarrollos tienen de la Red Española de Supercomputación (RES) y de la más reciente iniciativa para la gestión de datos de proyectos científicos organizada por la RES⁵². Estas iniciativas deben complementarse con la organización de una nube de datos científicos y biomédicos nacional (y europea) – esta es una carencia crítica que se discute más adelante.

4.9 Acceso, operación e interpretación de la información médica y biológica (genómica) en los casos COVID-19

La Comisión Europea junto con EMBL-EBI y ELIXIR ha lanzado el "portal de datos europeo COVID-19"⁵³ que conecta información genómica de virus y pacientes – aunque de momento no vinculada con el resto de la información médica, si se integra con el resto de las bases de datos biomédicas del EMBL-EBI y con los recursos de cómputo distribuidos (ELIXIR). A este esfuerzo se unen los países que progresivamente van estableciendo sus propios portales COVID-19 coordinados con el portal europeo.

El INB es responsable de la gestión del portal COVID-19 español dentro de esta iniciativa. Este portal debe albergar los datos de los proyectos nacionales sobre COVID-19, conectados como en todas las otras instancias nacionales con el portal europeo.

En el caso español se está realizando un esfuerzo especial en colaboración con la red TransBioNet para completar la información médica (formulario ISARIC mencionado anteriormente) como pieza fundamental para la interpretación de los datos genómicos.

5 Hacia un sistema de salud más resistente

Los hechos han demostrado claramente que el Sistema de Salud y los distintos componentes no están suficientemente preparados para gestionar de forma coordinada e integral la información que almacenan. No ha sido así en el pico de la epidemia en primavera y no es aún así, cuando entramos en la segunda ola⁵⁴.

⁵² <https://www.res.es/>

⁵³ <https://www.covid19dataportal.org/>

⁵⁴ El 15 de Septiembre de 2020 el ISCIII publicó una ambiciosa convocatoria: Infraestructura de Medicina de Precisión asociada a la Ciencia y Tecnología -IMPACT) de la Acción Estratégica en Salud 2017-2020 (<https://www.isciii.es/Noticias/Noticias/Paginas/Noticias/ConvocatoriaIMPACTMedicinaPrecision.aspx>).

Esta convocatoria está organizada en tres programas :

- Medicina Predictiva, está orientado al diseño y puesta en marcha de "una gran cohorte poblacional con datos clínicos, epidemiológicos y biológicos, medidos a nivel individual, que permita representar a la totalidad de la población residente en España, incluyendo la variabilidad étnica y la diversidad geográfica y ambiental".

- Ciencia de Datos, persigue "desarrollar un sistema de recogida, integración y análisis de datos clínicos y moleculares orientado a mejorar la salud de cada paciente individual, y que permita el uso secundario de la información existente en el SNS para el beneficio de la sociedad con objetivos de salud pública, planificación sanitaria e investigación".

- Medicina Genómica, desarrollará "infraestructuras y protocolos de coordinación para llevar a cabo análisis genómicos y otros datos 'ómicos' de forma eficaz, eficiente y equitativamente accesible en todo el territorio nacional, tomando como apoyo grandes centros de investigación españoles que ya disponen de tecnología de secuenciación de última generación y experiencia en su aplicación al diagnóstico de enfermedades humanas".

Los principales cuellos de botella desde el punto de vista de la gestión de la información corresponden a las carencias básicas en la digitalización e interoperación de la información en el sistema de salud. Carencias que la epidemia ha puesto de manifiesto de forma aguda y que deben de superarse progresivamente. Estas incluyen:

- i. Superar la compartimentación en el manejo interno de la información en y entre los sistemas de salud, incluyendo hospitales y centros de atención primaria junto a los sistemas de salud privados, superando la situación actual de atomización y bunkerización. Los sistemas actuales gestionados por empresas privadas, con sucesivos parches superpuestos, hacen muy difícil la gestión de la información y son insuficientes para la construcción de un sistema integrado de agregación e intercambio de información. Es, por tanto, necesario generar nuevos sistemas que, contruidos sobre los ya existentes, superen las limitaciones actuales. Las experiencias en otros países como USA, Noruega, Suiza, Estonia o UK, demuestran la complejidad y dificultad de este tipo de operación. La epidemia COVID-19 representa una oportunidad para comenzar la construcción de un sistema de intercambio de información clínica con unos objetivos bien definidos y un alcance limitado, i.e., proveer el conjunto mínimo de variables (ver capítulo sobre la base de datos Europea de COVID-19), lo que debe permitir ganar experiencia e instrumentos sobre los que construir posteriormente el sistema completo, adaptando la estructura de los correspondientes sistemas de salud para ser capaces de proporcionar la información automáticamente desde cada uno de los sistemas de IT en formatos estandarizados (*ver punto siguiente sobre estándares de información médica*).
- ii. De cara al futuro es fundamental que estos sistemas se construyan en entornos software abierto y que el control del diseño y desarrollo de los correspondientes proyectos permanezca en el dominio público para garantizar tanto la operatividad y conectividad, como la progresividad de la evolución futura. Una organización que a su vez fomentará la participación de las empresas que implementarán y mantendrán los sistemas IT, incrementando su mercado potencial.
- iii. En este contexto es especialmente relevante mencionar la importancia de la conexión de la información entre los distintos sistemas hospitalarios y de atención primaria, que solo en algunas comunidades están conectados. Es así mismo importante conectar la información de los sistemas públicos y privados, evitando las duplicaciones actuales. Incluyendo la mejora de los sistemas de registro automático de la información (e.g. informes desde hospitales y comunidades autónomas), incluyendo el sistema de notificación de enfermedades de declaración obligatoria del Instituto de Salud Carlos III).

Implementación de sistemas para la interoperabilidad y estandarización de vocabularios y ontologías, incluyendo las facilidades para la interoperación entre ellos. En el caso específico de la COVID-19 es fundamental el trabajo en marcha a nivel internacional para la implementación de una ontología multilingüe^{55 56}.

⁵⁵ <https://www.nature.com/articles/s41597-020-0523-6>

⁵⁶ <https://www.termcat.cat/ca/diccionaris-en-linia/286>

- i. Definir y hacer operables los indicadores cuantitativos de acceso a la atención y gestión sanitaria (mortalidad global, lista de espera, intervenciones quirúrgicas, dispensación de medicamentos, pacientes ambulatorios y pruebas canceladas, entre otros)^{57, 58} que permitan el desarrollo de sistemas de ayuda a la toma de decisión basados en indicadores⁵⁹.
- ii. Mejora de los sistemas para la transmisión al sistema de salud de las recomendaciones derivadas de estudios biomédicos. Colaboración con infraestructuras y agencias reguladoras y sistemas de salud con capacidad de convertir los resultados en recomendaciones a los sistemas de salud, creando las líneas de comunicación entre los resultados de investigación, la formalización de estudios clínicos y su recepción por las autoridades sanitarias.

6 Industria de la salud basada en datos

Como se viene señalando en este informe, el sistema de salud está sostenido sobre un conjunto heterogéneo de tecnologías informáticas (IT), en muchos casos instalados de modo bastante anárquico durante años; de hecho, muchos hospitales no disponen de sistemas completamente integrados capaces de gestionar de modo coordinado toda su información. Un ejemplo positivo es el sistema de Salud Andaluz, que integra todo el sistema público, tanto hospitales como centros de primaria y farmacias- dispensación farmacéutica. Aun así, no es fácil extraer información del sistema, incluyendo problemas legales, regulatorios que lo hace lento y complejo para obtener datos mínimos y aún más para compartir datos con otros sistemas donde pueden aparecer nuevos problemas éticos hasta ahora poco analizados.

Es de esperar que sean las compañías en el sector, que han desarrollado los sistemas actualmente en operación, las que sigan encargándose de la instalación y/o operación de los sistemas de IT de los hospitales (aunque es de tener en cuenta que existen alternativas de acceso abierto que deberían considerarse). Dada la complejidad de estos sistemas es, en la práctica, esencial evitar entrar en una reorganización completa del sistema, requiriendo de los hospitales rehacer sus sistemas, algo que sería demasiado complejo y caro. Por el contrario, es más razonable requerir que los sistemas hagan progresivamente accesibles conjuntos de datos en un formato determinado en respuesta a requerimientos específicos (FIHR, Fast Healthcare Interoperability Resources, como estándar actual). En el caso de la COVID-19 estos podrían ser en primer lugar el conjunto de datos mínimos ISARIC recomendado por la Organización Médica Mundial (OMS o WHO). Unos requerimientos progresivos permitirían que distintos sistemas de salud y/o hospitales se incorporaran progresivamente a medida que automatizaran sus sistemas. A su vez, un requerimiento progresivo de compartir más datos aumentaría la demanda a sus sistemas de IT de modo aceptable. La propuesta

⁵⁷ Søreide, K., et al. (2020) Immediate and long-term impact of the COVID-19 pandemic on delivery of surgical services. *Br J Surg*.

⁵⁸ Weinstein, et al. (2020) Delayed Primary and Specialty Care: The Coronavirus Disease-2019 Pandemic Second Wave. *Disaster Med Public Health Prep*.

⁵⁹ Data en biomedicina EIT-health <https://sciencebusiness.net/report/healthy-measures-why-health-outcomes-matter-europes-future>

de una incorporación progresiva al sistema de compartición de datos requeriría inicialmente menos adaptaciones de los sistemas y más un esfuerzo de construcción de las pasarelas para el intercambio de datos (suministro de datos en formato FIHR). Es este contexto es fundamental que los desarrollos sean de código abierto y puedan ser analizados y reutilizados. El sistema de información compartido debe ser la base para la creación de nuevos negocios basados en la explotación de los datos. Un ejemplo es el entrenamiento de los modelos de ciencia de datos y sistemas de inteligencia artificial, que requieren acceso a grandes cantidades de datos, e.g, trayectorias, medicaciones, síntomas -y por tanto los resultados comerciales deberían de explotarse conjuntamente por el sistema público y privado (consorcios público-privados) con el consentimiento de los ciudadanos propietarios de sus datos individuales. Dado el tamaño del sistema de salud español, los sistemas desarrollados podrían tener un alto valor comercial, por ejemplo, sistemas de minería de textos médicos en castellano capaces de extraer información de los textos médicos tienen un enorme mercado potencial.

En este proceso de digitalización e integración es esencial incluir sistemas abiertos de evaluación de la operatividad, transparencia, accesibilidad y calidad de las soluciones de gestión y análisis de datos. Las evaluaciones abiertas sobre un problema determinado son frecuentes en áreas de las tecnologías de la información e Inteligencia Artificial. En ellas se evalúan objetivamente las tecnologías de distintos proveedores, se fijan los niveles de calidad alcanzables y se proponen mejoras específicas de las tecnologías. Este tipo de evaluación es útil para consolidar estándares y evaluar el estado de la tecnología frente a problemas específicos. Cabe destacar que la evaluación abierta de los métodos es aún más necesaria en los desarrollos de métodos de Inteligencia Artificial, dado que su complejidad hace muy difícil una interpretación directa de las razones detrás de las soluciones propuestas por los nuevos métodos de IA. Un sistema de evaluación abierta de métodos debe actuar tanto como potenciador de la participación de compañías, facilitando su acceso a datos y como sistema de certificación de calidad, representando una garantía para las ventas de las compañías y una contribución pública al establecimiento de protocolos y estándares que den seguridad a los usuarios finales. Por tanto, los sistemas de evaluación abiertos deben ser un componente central de la estrategia de acceso e integración de la información biomédica y contribuir a generar las directrices técnicas para las compras públicas, como en los de “Compra Pública Innovadora”.

7 Recomendaciones

La pandemia ofrece una oportunidad para la reorganización del sistema de acceso y uso de la información biomédica y priorizar proyectos que consoliden estos desarrollos en el marco de una apuesta política, social, científica y económica que gira en torno al dato y a los conjuntos de datos para mejorar la toma de decisiones. Construir una infraestructura que lo haga posible es crucial para que la convergencia de tecnologías como la inteligencia artificial y la ciencia de datos puedan tener lugar en el sistema de salud de forma eficiente.

En consecuencia, formulamos las siguientes recomendaciones:

- i. Desarrollar un Proyecto sostenible de integración de información médica mínima estandarizada, con posibles fases según el grado de organización de los hospitales y sistemas de salud. Acompañar el proyecto con un sistema de incentivos para los hospitales y con un canal de participación para las empresas en el sector, pero sin perder el liderazgo público y dentro de un sistema de código abierto y no propietario.
- ii. Integrar la información médica y genética/genómica. Desarrollar un proyecto nacional que se alinee y use los estándares y recursos desarrollados por la iniciativa europea de base de datos COVID-19 y con el 1M+G. Este proyecto se debe desarrollar en entornos de software abiertos, siguiendo las directrices FAIR⁶⁰ para datos y las buenas prácticas en software⁶¹. Este proyecto debe incluir la financiación para implementar EGA federada como el sustrato de manejo de datos genómicos común. Este proyecto debe considerarse como el bloque inicial para la construcción de una infraestructura sostenible que se evalúe periódicamente.
- iii. Promover la integración de las nuevas tecnologías de ciencia de datos (Big data clínico, Inteligencia Artificial, Procesamiento del Lenguaje Natural⁶², Estadística Avanzada, Investigación Operativa, Sistemas Dinámicos, Optimización y Decisión, y distintos tipos de Simulación adaptados a los distintos niveles, desde molecular a poblacional, con proyectos específicos dirigidos a las cuestiones clave y ligados directamente a las fuentes de datos (base de datos nacional Covid-19).
- iv. Las nuevas tecnologías implican la necesidad de un capítulo específico de formación a todos los niveles tanto de usuarios (médicos y otro personal en el sistema asistencial) como de desarrolladores de sistemas de datos y análisis, incluyendo aspectos técnicos, éticos y legales. Sería particularmente importante enlazar esta formación a planes nacionales más amplios en IA y PLN.
- v. Todo este desarrollo debe basarse en una infraestructura computacional adecuada. Es necesario disponer de un sistema nacional de datos y computación en la nube para datos académico y de salud. La Red Española de SuperComputación (RES) ha establecido la primera fase del sistema de almacenamiento y gestión de datos académicos que debe completarse con un sistema de computación en nube asociado a los recursos de la RES. Este sistema

⁶⁰ <https://www.go-fair.org/fair-principles/>

⁶¹ <https://www.rd-alliance.org/plenaries/rda-15th-plenary-meeting-australia/fair-principles-research-software>

⁶² Plan de Tecnologías del Lenguaje - <https://www.plantl.gob.es/>.

de nube y datos debe ser la base para la conexión con los equivalentes europeos, principalmente con la nube académica alemana de.NBI⁶³ y la propuesta europea de nube público-privado para fines de investigación GAIA-X⁶⁴

- vi.** Promover la industria de la salud basada en productos de valor añadido, más allá de la venta de datos o datos sobre casos. Explorar como vehículo de innovación tecnológica la utilización de sistemas de evaluación de métodos abiertos en los que la administración diseña los requisitos y escenarios, basándose en los resultados de las competiciones abiertas, que la industria ejecuta, cerrando un círculo virtuoso público-privado.

⁶³ <https://www.denbi.de/>

⁶⁴ <https://www.data-infrastructure.eu/GAIA/Navigation/EN/Home/home.html>

Apendice 1. Ejemplos de la utilidad de entornos ricos en datos para el desarrollo de sistemas de predicción basados en Ciencia de Datos e Inteligencia Artificial en temas relacionados con COVID-19.

Un ejemplo de uso de tecnología IA en el análisis de datos de COVID-19 es el algoritmo de BlueDot, basado en el procesamiento del lenguaje natural y el aprendizaje automático, para pronosticar el riesgo de propagación a fines de diciembre de 2019 basándose en informes del procesamiento de noticias en periódicos locales y billetes de avión. En todo caso, es bueno poner estas capacidades en el contexto de a otras alternativas más tradicionales. Por ejemplo, Marjorie Pollack, epidemióloga ya había detectado la posibilidad del brote y emitió un informe, mucho más detallado que el de BlueDot, aproximadamente 30 minutos después de la alerta de éste, elevando la categorizando del brote a alto riesgo.

Otros ejemplos de la aplicación de IA en este área incluyen la propuesta de proteínas diana relevantes con el algoritmo de AlphaFold de Deep Mind y la predicción de la estructura de las seis proteínas humanas potencialmente más atacadas por el SARS-Cov-2 junto a la exploración de las redes de interacciones de proteínas para la potencial reutilización de medicamentos. Un número de otras aproximaciones computacionales basadas en IA, se han utilizado para modelar la interacción entre fármacos y proteínas, identificando candidatos como baricitinib u otros antivirales aprobados por la FDA que podrían potencialmente apuntar a las proteínas clave en el desarrollo de COVID-19, o la utilización de aprendizaje por refuerzo para modelar moléculas pequeñas e identificar aquellas que inhiben vías de infección específicas.

La predicción del curso de la enfermedad a partir de datos clínicos y el triaje de pacientes basado en síntomas iniciales y datos clínicos. Identificación de pacientes con mayor riesgo de mortalidad con métodos de clasificación de ciencia de datos junto a las principales características clínicas o moleculares determinantes de las predicciones.

El uso de métodos de análisis de imagen médica para el diagnóstico con métodos de IA incluye, publicaciones como el diseño por Infervision de un algoritmo de visión por computadora para la detección de neumonía por coronavirus en imágenes pulmonares o el uso de Convolutional Neural Networks para aprender automáticamente patrones relacionados con COVID-19 a partir de imágenes, con resultados equivalentes a los análisis por polimerasa con transcripción inversa (RT-PCR).

La utilización de tomografías computarizadas junto a información clínica, se ha utilizado para identificar características fundamentales asociadas al estado de la infección por SARS-CoV-2 (26).

Finalmente, en varias áreas en las que la confidencialidad de los datos es un factor esencial se han desarrollado nuevas aproximaciones denominadas genéricamente “federated learning”, que, aunque son obviamente dependientes de la capacidad para extraer información homogénea e interoperable, son en principio capaces de salvar el grave inconveniente que la falta de suficientes datos representa.



Un aspecto particularmente preocupante es que los sistemas de inteligencia artificial rara vez están dotados de suficiente solidez para responder a los desequilibrios de clase, como el sexo y género. Una situación particularmente preocupante en el caso de COVID-19 (estudio de caso de la Comisión Europea).

El artículo “Mapping the landscape of artificial intelligence applications against COVID-19”, da una visión realista del potencial de la IA respecto a las posibles aplicaciones en la epidemia actual y su potencial para el futuro: “we note that very few of the reviewed systems have operational maturity at this stage. In order to operationalize this research, it is important to define a research road map and a funnel for AI applications to understand how this technology can immediately assist with the response, how it might help later on in the evolution of the current pandemic, and how it can be used to combat future pandemics”.

Apéndice 2: Proyectos COVID-19 en varios países

España

Proyectos de investigación financiados por fondos COVID-19 (ISCIII)

El catálogo completo de proyectos financiados por el Instituto de Saludo Carlos III – Fondos COVID se encuentra en la URL:

https://www.isciii.es/QueHacemos/Servicios/Biblioteca/Paginas/Proyectos_investigacion.aspx

Resumen de hospitales y centros de salud implicados en proyectos ISCIII (algunos representados a través de sus Fundaciones correspondientes):

Hospital de la Santa Creu i Sant Pa, Hospital Universitario Puerta de Hierro, Hospital Clinic de la Comunidad Valenciana, Hospital Clínico San Carlos, Hospital Príncipe de Asturias, Hospital Gregorio Marañón, Hospital General de Elche, Hospital 12 de Octubre, Hospital Val d’Hebron, MÚTUA Terrasa, Hospital La Fe, Hospital La Paz, Fundación de Investigación HM Hospitales, Hospital Clínico Universitario de Valladolid, Hospital del Mar, Hospital Ramon y Cajal, Hospital Universitario Fundación Alcorcón

Fundaciones y centros de Investigación:

CSIC, IACS, CNIO, FFIS, FISABIO, ISGLOBAL, FIMABIS, IDIBAPS, CIBER, IDIVAL, IDIBELL, FFIS, IRBLLEIDA, NAVARRABIOMED, FISEVI, FUNDACION RIOJA SALUD, IDIAP JORDI GOL, IRBLLEIDA, IDIVAL, Fundació Institut de Recerca Biomédica, Fundación Instituto de Investigación Sanitaria Illes Balears, Fundación Pública Andaluza Progreso y Salud, FIISC, Fundación para la Investigación e Innovación Biosanitaria en Asturias – FINBA, IGTP (Instituto de Investigación Germans Trias i Pujol), IMDEA-Nanociencia, Instituto de Investigación Sanitaria Biocruces Bizkaia, FUNDACION INSTITUTO DE INVESTIGACION SANITARIA ILLES BALEARS, FUNDACION INSTITUTO DE INVESTIGACION SANITARIA DE SANTIAGO DE COMPOSTELA, ASOCIACION INSTITUTO BIODONOSTIA, ASOCIACION INSTITUTO DE INVESTIGACION SANITARIA BIOARABA, FUNDACION INSTITUTO DE INVESTIGACION SANITARIA DE NAVARRA, FUNDACIÓN BIOMÉDICA GALICIA SUR, Centre de Recerca Matemàtica (CRM), Instituto de Bioingeniería de Cataluña,

Universidades:

UNIVERSIDAD DE SEVILLA, FUNDACION PARA LA GESTION DE LA INVESTIGACION BIOMEDICA DE CADIZ, BSC, UNIVERSIDAD DE ZARAGOZA, , UNIVERSITAT ROVIRA I VIRGILI, UNIVERSIDAD DE CÓRDOBA, Universidad Pompeu Fabra, Universidad de Santiago de Compostela, Universidad de Extremadura, UNIVERSIDAD DE MALAGA, FUNDACIÓ PARC TAULÍ, UNIVERSIDAD RAMON LLULL, Universidad Politecnica de Catalunya, Universidad de Oviedo, Universitat de València, Universidad Carlos III de Madrid, Universidad del País Vasco, UNED, Universidad de A Coruña, Universidad Complutense de Madrid

Ensayos/estudios clínicos relacionado con fármacos (reposicionamiento, uso terapéutico, efectos clínicos,..):

Hidroxiclороquina, Azitromicina y Tocilizumab, terapia con plasma de convalecientes, mefloquina, sevoflurano, antivirales para SARA basados en polifarmacología, fármacos



que incrementan la expresión de ACE2, reposicionamiento terapéutico dirigido a inhibición de expresión de ACE, angiotensina, efecto de los AINE's y los IECA'S, inhibidores de la proteasa Mpro, perfiles de citoquinas capaces de predecir el desarrollo de distress respiratorio, fármacos y repuesta inmune, tratamientos inmunosupresores

Desarrollo y estudio de sensores, mecanismos de diagnóstico análisis genético:

sensores colorimétricos, diagnóstico basado en nanosensores, diagnóstico basado en aislamiento de exosomas, diagnóstico por phi29 polymerasas, identificación mediante balizas moleculares, tests serológicos, estudios NGS, marcadores inmunológicos, estudio de carga viral en vías respiratorias y sangre, determinantes Genéticos y biomarcadores genómicos, seroprevalencia, muestras ambientales, perfil inmunológico, ozono para la reutilización de equipos, contaminación de superficies, perfil transcripcómico

Vacunas:

vacuna mediante micro/nanoesferas de muNS-Mi, protección de BCG y MTBVAC, , vacuna, MVA-COVID-19, vacuna de mRNA basada en nanopartículas poliméricas , terapia contra el basada en ARN circulares, mRNA/MVAvaccines, Peptide Vector for mRNA Vaccination

Mecanismos de transmisión:

Transmisión por el aire, cartografías de peligrosidad de transmisión, mecanismos de transmisión también en otros estudios mas relacionados con comorbilidades (ver abajo)

Registros y estudios epidemiológicos/Comorbilidades:

Embarazo y gestación, efectos en el sistema cardiovascular, pacientes con VIH, enfermos UCI, afectación neurológica, pacientes con enfermedades inflamatorias, pacientes con artritis inflamatoria crónica, pacientes receptores de trasplantes de órganos, pacientes con insuficiencia cardiaca y cardiopatías familiares, pacientes con hemoglobinopatías, personas > 60 años, pacientes con esclerosis múltiple, pacientes con enfermedades inflamatorias y autoinmunes, pacientes con errores congénitos de la inmunidad

Inmunidad, microbioma, etc:

Microbioma, inmunosenescencia (células T y B), células NK linfocitosis B monoclonal, hiperactivación inmunológica, perfil inflamatorio y estatus redox

Modelado y simulaciones y IA:

inhibición, interacciones con huésped, modelos epidemiológicos físico-matemáticos, modelo mecanístico con IA para la reutilización de fármaco, propagación, y predicción de incidencia, Dinámicas de Proteínas, series temporales para optimizar la gestión de recurso, dinámica de transmisión, estructural de las proteínas, cálculo dinámico de la cápside para su destrucción por resonancia

Otros:

Impacto en la salud mental, transporte a mediados por microtubules, estudio del potencial impacto del COVID19 en mascotas y lince, escala pronóstica en neumonía,

data collection, fibrosis pulmonar residual y capacidad pulmonar, epitelio respiratorio human, coste hospitalario, modelos pre-clínico, Síndrome Post Cuidados Intensivos.

Otros proyectos financiados en España incluyen los del CDTI (<https://www.cdti.es/index.asp?MP=100&MS=917&MN=2>) y los financiados por fondos Supera COVID-19 (Banco Santander, Universidades Consejo Superior de Investigaciones Científicas (CSIC) (<https://www.icscyl.com/investigacion-convocatorias/2020-05-05/fondo-supera-covid-19/>)).

Alemania:

1. Reposicionamiento de fármacos y ensayos clínicos en fase I, I/II y II
2. Control de COVID-19 mediante nuevos enfoques diagnósticos y terapéuticos
3. Enfoques relacionados con medidas de control de la infección o medidas relacionadas con los mecanismos de transmisión.
4. Caracterización de los mecanismos de expansión del virus
5. Condiciones relacionadas con la expansión del virus
6. Desarrollo de modelos clínicos en animales y celulares relacionados con procesos tales como la infección pulmonar
7. Mecanismos de protección de personal asistencial y sanitarios
8. Investigación sobre aspectos éticos, legales y socioeconómicos en relación con los brotes

<https://www.gesundheitsforschung-bmbf.de/de/10592.php>

Suiza:

1. Use protein-based membranes for the purification of air contaminated by COVID-19
2. Examine and manage the economic and social impact of collective emotions, emotional contagions, and pandemic-induced fear during the COVID-19 pandemic. We expect to see that observed differences in cross-regional and cross-national COVID-19 mitigation compliance are partially attributable to emotional contagions
3. Use gaseous acetic acid (CH₃COOH) in room air, as an easy-to-apply measure, in order to lower the pH of exhalation aerosol, expectorated droplets and contaminated contact surfaces, and thereby to effectively mitigate the risk of transmission of SARS-CoV-2
4. Official/institutional health communication from authorities to the public during public health emergencies. This communication involves the presentation of risks under uncertainty. Risk communication. Creation of a toolkit composed of a theoretical framework and operational guidelines on how to best deliver health information during public health emergencies and in contexts characterized by misinformation
5. Identifying and providing strategies to influence individuals' perceptions and behaviors, focusing on how to combat misinformation. The overall aim of this project is to analyze and integrate these different data sources as quickly as possible, thereby creating a novel interdisciplinary surveillance system for SARS-CoV-2/COVID-19 in Switzerland. Key parameters for the parameterization of the

mathematical model, will be extracted automatically or by hand search from the identified full text articles. As a possible extension we will also analyse social media data (e.g. from Twitter, Facebook and Reddit) in real-time which may help to understand how behaviour of individuals changes as the epidemic evolves. The second subproject concerns the analysis and comparison of existing Swiss surveillance data. Data include i) a sentinel surveillance system of 10 Swiss hospitals (project being finalized; partly funded by the Federal Office of Public Health FOPH; O Keiser is the principle investigator (PI)); ii) Sentinella (Influenza/COVID surveillance by general practitioners; available by FOPH); Grippenet (app to report influenza-like symptoms voluntarily; PI A Flahault). In addition, we aim to integrate whole genome sequence data as soon as they become available from the national reference lab for emerging viruses in Geneva (PI I Eckerle). Combining the mathematical model with the other sub-projects will give us a deeper understanding of the COVID-19 epidemic, and allow us to evaluate the effectiveness of interventions (e.g. by identifying risk factors, by improving the management of hospital beds; by focusing interventions on areas of intense viral circulation, and by detecting more or less virulent strains. To this end we will use ribosome profiling and RNAseq in a high-resolution infection time course to quantitatively determine the translational response of the host cell preventive measures: institutions with a governmental mandate, prison system, therapy facilities, as well as care facilities for children and young people PubliCo - an experimental online platform for COVID-19 related public perception

6. The fine characterization of the immune signatures associated with intensive care unit (ICU) and non-ICU hospitalized SARS-CoV-2 infected individuals at the time of admission/diagnosis and longitudinally may help to improve the care of infected patients Of particular importance during the time of the ongoing pandemic is to follow SARS-CoV-2 evolution and to use our established synthetic genomics platform and the established primary human airway epithelium (hAEC) culture system to link genotypes to phenotypes. On a global scale it will be important to carefully link genomic analyses with patient metadata in order to immediately recognize if SARS-CoV-2 variants that arise during the pandemic may have undergone a phenotypic change. We will bridge clinical data, patient samples, coronavirus reverse genetics, and primary human airway epithelial systems, to translate observed virus evolution directly to functional studies. Determine the effectiveness of a desynchronization strategy to protect healthcare workers. A desynchronization strategy was installed in one large unit in which the entire staff was dichotomized. Each of the two teams is working alternating for one week with the goal to reduce contact between the teams and decrease infection rates of the workforce. By comparing health status and productivity of the workforce to other departments will determine the relevance of such organizational distancing in healthcare workers
7. Impact of Sex and Gender on COVID-19 outcomes: Role of ACE-2, TMPRSS2, and gender-specific risk factors
8. Phenotypic characterization of a SARS-CoV-2 clinical strain bio-bank and risk assessment of evolving virus variants
9. EXSCALATE4CoV - EXAScale smArt pLatform Against paThogEns for Corona Virus

<http://www.snf.ch/de/fokusForschung/newsroom/Seiten/news-200518-uebersicht-schweizer-forschung-zu-covid-19.aspx>

Francia

1. Vaccine candidates
2. Knowledge of the virus and its pathogenesis
3. Development of research tools
4. Epidemiological research
5. Therapeutic research
6. Development of diagnostic tests
7. Epidemiological research, social epidemiology
8. Prophylaxis and candidate vaccines
9. Development of diagnostic tools

1. Vaccine candidates

Development of an animal model and accelerated development of a DNA vaccine candidate Spike glycoprotein, lentiviral vectors and B/T-cell vaccine

2. Knowledge of the virus and its pathogenesis

- a. Evolution of SARS-CoV-2 and the antibody immune response in humans during infection
- b. A technique to identify key cellular factors for SARS-CoV-2 infection
- c. SARS-CoV-2 fusion, replication and host responses
- d. Profiling SARS-CoV-2 intra-viral and virus-host interactions on the basis of innate immunity signaling
- e. Working towards controlled detection of SARS-CoV-2 by the innate immune system
- f. Origins, natural reservoirs and interspecies transmission of SARS-CoV-2 and other SARS-like coronaviruses
- g. Monitoring the origins and transmission/spread of SARS-CoV-2 in Laos and Vietnam: searching for SARS-like viruses and detecting antibodies in vertebrates
- h. Characterization of interaction between SARS-CoV-2 Nsp3 protein and non-canonical nucleic acid structures (G4) present in infected cell
- i. A combination of cutting-edge proteomic approaches for improving understanding of SARS-CoV-2 infection mechanisms
- j. Characterization of variable systemic and mucosal immunity during SARS-CoV-2 infection and recovery
- k. Nicotinic modulation of SARS-CoV-2 infection in the brain (NicoSars Project)
- l. Mechanisms of intercellular spreading of SARS-CoV-2: assessing the role of TNTs to escape immune surveillance and expand tropism and pathogenicity (COVID-Spreads project)
- m. Evaluating the determinants of coronavirus tropism in 3D models of human pulmonary epithelia (TROPICORO project - TROPism of CORONaviruses)
- n. Neurological implications of SARS-COV-2 infection – a transversal study (Neurological project)

- o. Restriction of SARS-CoV-2 replication by coding and non-coding genes (ROSA project)
 - p. Deciphering the virus interactions with the entry receptors in human cells, and their potential inhibition (DSAC project)
 - q. Investigating the role of apico-basal proteins during SARS-Cov2 infection (EPIC project)
3. Development of research tools
- a. Development of cell lines for research
 - b. Antibody profiling in recovering patients and development of a serological test for an epidemiological survey in individuals exposed to SARS-CoV-2
 - c. A SARS-CoV-2 genome sequencing technique to analyze low-quality clinical samples
 - d. Development of simple Covid-19 serological tools and targeted serosurvey of at-risk individuals
 - e. Generation of animal models sensitive to SARS-CoV-2 to test vaccines or drugs
 - f. Development of a test to identify inhibitors of SARS-CoV-2 proteases
 - g. Serological surveillance of SARS-CoV-2 and seasonal coronaviruses
 - h. The use of Artificial Intelligence Methods to Discriminate Covid-19 from other Community Acquired Pneumopathy using Chest X-Ray and CT images: Diagnostic of SARS-CoV-2 Coronavirus caused-disease (Covid-19) is a bottleneck in the immediate follow-up of patients. We propose automated detection methods based on Artificial Intelligence (AI) Algorithms (Machine Learning and Form Recognition), on chest Computed tomography (CT) or X-Ray images, to discriminate the Covid-19 disease from other Community Acquired Pneumonia (CAP). The methods will be used as complement to the reverse-transcription polymerase chain reaction (RT-PCR) tests.
 - i. Developing a rapid diagnostic test to detect SARS-CoV2 virus in various environment
 - j. Sequence the complete genome of Covid-19 directly and at low cost from prepared RNA samples from patients (CAPTURE-COV project)
 - k. Air monitoring risk area (ARISE project)
 - l. A zebrafish model to study SARS-CoV-2 neuroinvasion (NeuroCoVid project)
 - m. rBCG vaccine secreting SARS-CoV-2 antigens (rBCG vaccine project)
4. Epidemiological research
- a. Building a collection of human biological samples in France
5. Therapeutic research
- a. Defective viral genomes (DVGs), potential antiviral inhibitors of SARS-CoV-2
 - b. Isolating and characterizing the human antibodies that neutralize SARS-CoV-2
 - c. Find potent antiviral drugs against SARS-CoV2 by targeting well identified proteins essential to the viral cycle.
 - d. Identification and isolation of potent human neutralizing antibodies against SARS-CoV-2
 - e. Accelerating the identification & design of SARS-CoV2 antiviral drugs in combination with the “DrugDesign_SARS2” Project

- f. Urgent Discovery of Drug Candidates That Target Coronavirus SARS-CoV-2
 - g. Drug Repurposing & Design to inhibit the SARS-CoV2 proteases (CoV2-Anti-protease project)
6. Development of diagnostic tests
 - a. Production of recombinant antigens from two SARS-CoV-2 proteins and generation of nano-antibodies against these proteins for diagnostic and therapeutic applications
 7. Epidemiological research, social epidemiology
 - a. Red Cross health workers and volunteers and continuity of care with COVID-19: an operational approach using mixed methods
 - b. Household transmission investigation study for SARS in overseas territories
 - c. Covid-19 risk evaluation among the household contacts of the first cases in Africa
 - d. Evaluation of clinical presentation and evolution of SARS -CoV-2 coronavirus infection in Senegal
 - e. Protocol for assessment of potential risk factors for 2019-novel coronavirus (2019-nCoV) infection among health care workers in a health care setting.
 - f. Engineering an ACE2 derived polypeptide for label-free, real-time detection on a smartphone device and therapeutic intervention
 - g. Evaluation of SARS-CoV-2 infection spread in various populations groups in Cameroon
 8. Prophylaxis and candidate vaccines
 - a. Randomized trial of Covid-19 chemoprophylaxis in healthcare professionals
 - b. Development of novel nano-based multi-epitope vaccine against coronavirus SARS-CoV-2 - Generation of Humanized NSG transfer mouse model for vaccine testing
 - c. Investigation of the first cases of coronavirus (COVID-19) infections and their contacts in Antananarivo, Madagascar
 9. Development of diagnostic tools
 - a. Developing a low-cost lab-on-a-chip for diagnosing SARS-CoV-2

<https://www.pasteur.fr/en/all-covid-19-institut-pasteur/research-projects/coronavirus-covid-19-research-projects>

The Netherlands

The research programme COVID-19 focuses on research aimed at the effects of the coronavirus crisis and the measures taken against the coronavirus pandemic. There are three focus areas:

- i. Predictive diagnostics and treatment
- ii. Care and prevention
- iii. Societal dynamics

The programme is being realised by the Netherlands Organisation for Health Research

and Development (ZonMw) and NWO on behalf of the Ministry of Health, Welfare and Sport. The programme is being funded by the Ministry of Health, Welfare and Sport, the Ministry of Education, Culture and Science and by NWO.

Bottom-up calls for proposals have been published for all three focus areas. A call for proposals *COVID-19 Science for professional practice* for practice boosts and policy questions has also been published for small projects and studies with a maximum budget of € 25,000. In view of the need to act urgently, the deadline for submitting (bottom-up) project ideas for focus areas 1 and 2 is Thursday 14 May 2020 at 14.00 hours CEST, and for focus area 3 it is Monday 25 May 2020 at 12.00 hours CEST. For *COVID-19 Science for professional practice* the deadline is Monday 25 May 2020 at 12.00 hours CEST.

Focus area 1 - Predictive diagnostics and treatment

In this focus area, the emphasis is on urgently needed research for the prevention and treatment of COVID-19-related complaints in the early, acute and recovery phases. This concerns research into new or existing therapies and their working mechanisms and obtaining insights into, amongst other things, the microbiome, immunity, predictive parameters and individualised treatment.

This call for proposals has four themes:

- i. Treatment
- ii. Diagnostics of infection
- iii. Risk analysis and prognostics
1. Virus, immunity, immune response and pathogenesis

The deadline for submitting project ideas is 14 May 2020 at 14.00 hours CEST. The deadline for submitting a full proposal is 15 June 2020 at 14.00 hours CEST.

- Call for proposals Focus area 1 - Predictive diagnostics and treatment

Focus area 2 - Care and prevention

This focus area concerns research aimed at the organisation of care and vulnerable citizens. In addition, there is specific attention for care providers, in the approach for preventing infection and in the (psychosocial) supervision during and after crisis situations in care institutions. The focus lies on both the impact of behaviour and behavioural changes on the spread of the virus, as well as the consequences of the measures for the individual or for specific vulnerable groups.

This call for proposals has three themes:

- i. Organisation of care and prevention
- ii. Care and prevention for vulnerable citizens
- iii. Transmission and epidemiology

The deadline for submitting project ideas is 14 May 2020 at 14.00 hours CEST. The deadline for submitting a full proposal is 15 June 2020 at 14.00 hours CEST.

- Call for proposals Focus area 2 - Care and prevention

Focus area 3 - Societal dynamics

This call for proposals focuses on generating knowledge about the (global) societal dynamics during and after this health crisis. This concerns the effectiveness and impact of measures in response to the coronavirus process as well as the resilience of the

society and the economy. The societal dynamics is therefore not limited to within the Netherlands.

It focuses on broad, societal issues in which several scientific disciplines can be involved. With this call for proposals, consortia, research groups and individual researchers are invited to submit ideas for research projects from which knowledge can be generated about the Dutch and worldwide impact of the coronavirus crisis and the measures taken against this. There are three priority themes within this area:

- i. Research into the effectiveness and impact of measures/strategies in response to the coronavirus crisis
- ii. Research into the resilience of society
- iii. Research into the economic resilience

The deadline for submitting project ideas is 25 May 2020 at 14.00 hours CEST. The deadline for submitting a full proposal is 29 June 2020 at 14.00 hours CEST.

- Call for proposals Focus area 3 - Societal dynamics

COVID-19 – Science for professional Practice

The call for proposals is intended for brief (action) research realised by collaborations between scientific organisations (researchers) and businesses, public organisations, administrative bodies or (local) governments. The scientific organisation submits the application on behalf of the collaboration. This call for proposals focuses on the three focus areas of the COVID-19 Programme, namely:

- i. Predictive diagnostics and treatment
- ii. Care and prevention
- iii. Societal dynamics

The deadline for submitting applications is 25 May 2020 at 12.00 hours CEST.

- Call for proposals COVID-19 – Science for professional Practice

<https://www.zonmw.nl/nl/over-zonmw/coronavirus/programmas/programma-detail/covid-19-programma/projecten/>

Algunos proyectos específicos son:

1. Escenarios para después del encierro en hogares de ancianos
2. GLOBAL: Máscaras de nariz y boca: alternativas a la máscara de nariz y boca quirúrgica
3. El efecto de Covid-19 sobre la innovación y las actividades de emprendimiento corporativo de las organizaciones
4. El modelo de capacidad de transporte público COVID-19: una herramienta de apoyo de políticas para optimizar la capacidad de corona del transporte público
5. El proyecto COFIT-20: estimular la actividad física a través de lecciones grupales en línea en vivo para personas con una enfermedad mental grave durante COVID-19
6. El curso del infarto agudo de miocardio en el momento del bloqueo de COVID-19 en regiones endémicas altas y bajas de COVID-19 en los Países Bajos
7. Terapia profesional en línea como terapeuta profesional.
8. Optimización de los flujos de pacientes y horarios de citas en la clínica.

9. Personas vulnerables en movimiento: adaptando la gama de actividades a la 'nueva normalidad'
10. eCOFIT +: salud en línea para apoyar la rehabilitación geriátrica ambulatoria
11. Logística sanitaria en la sociedad de 1,5 metros: una herramienta de planificación para hospitales holandeses
12. Ver este proyecto ¿Qué significa COVID-19 para la empleabilidad y la capacidad de recuperación de los empleados vulnerables en los servicios de las instalaciones?
13. Una herramienta de comunicación digital para la admisión a cuidados intensivos
14. Monitoreo en tiempo real en la rehabilitación domiciliar de pacientes con infección por COVID-19 después del ingreso hospitalario (REACH +)
15. Percepciones y necesidades prácticas de los ancianos con respecto a las medidas (relajadas) de Covid-19
16. Nuevas aplicaciones de tratamiento de trauma cuando los tratamientos cara a cara no son posibles como resultado de COVID-19
17. Portal COVID-19 de primera línea: desarrollo de mejores prácticas para la rehabilitación COVID-19 de primera línea basada en intercambio de conocimiento digital.
18. Adaptación y efectos de la realidad virtual como herramienta para superar la distancia social en la educación superior
19. investigación de acción sobre el trabajo sexual en el corazón de Brabante durante la época de la corona.
20. Asociación educativa en VVE remoto: principios de trabajo para proporcionar apoyo virtual a los padres de estudiantes en riesgo
21. Condiciones para soluciones tecnológicas en una estrategia de salida de Covid-19, con un enfoque particular en las condiciones legales y sociales
22. Difusión aerogénica del SARS-CoV-2 en situaciones sociales y durante procedimientos médicos.
23. Monitoreo y control efectivo basado en evidencia de Covid-19 después del brote inicial

Bélgica

Areas of research

1. National or international translational multicenter clinical research into the development and efficacy of novel or existing antiviral (of immunosuppressive) drugs, (repurposed or off-label) therapeutics, vaccines, vaccination schedules and/or the reliability of immunological serology-based diagnostic tests and their validation in preclinical models.
2. Clinical-pathological analysis of patient samples for a better understanding of COVID-19 immunity, identification of early biomarkers, infectivity, variability of the virus, comorbidities in COVID-19 patients during or after the acute phase.
3. Mathematical models and translational data-driven epidemiological research with direct added value in support of policy decisions (e.g. studies into the spread of coronaviruses; how to phase out lockdown and other restrictive measures; who to vaccinate first; developing a “post-corona start-up” strategy).

4. Technology development projects that solve a well-defined verifiable bottleneck (e.g. masks, therapeutic agents, ventilation equipment, vaccine production technology, and the like)

Proyectos COVID-19 en la European Research Area (ERA) corona platform

URL:

<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/covid-19?tabId=3>

Proyectos financiados bajo el call:

“Advancing knowledge for the clinical and public health response to the 2019-nCoV epidemic (sc1-phe-coronavirus-2020)”

1) Proyectos con participación española:

- i. **EXSCALATE4CoV** (€ 2 970 875)
Plataforma EXSCALATE contra patógenos del coronavirus
Participa: BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION
- ii. **CoNVat** (€ 2 547 152,48)
Plataformas avanzadas de nanobiosensores para el diagnóstico y la vigilancia global de Point-of-care (POC)
Participa: UNIVERSITAT DE BARCELONA
- iii. **Solnatide** (€ 1 626 025)
Exploración de la seguridad, tolerabilidad y eficacia clínica de Solnatide IMP en pacientes infectados con el nuevo coronavirus de 2019
Participa: BCN PEPTIDES SA
- iv. **MANCO** (€ 3 034 581,25)
Anticuerpos monoclonales contra el nuevo coronavirus 2019
Participa: AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS
- v. **I-MOVE-COVID-19** (€ 2 601 928,32)
Red europea multidisciplinar de investigación, prevención y control de la pandemia COVID-19
Participa: INSTITUTO DE SALUD CARLOS II
Participa: ORGANISMO AUTONOMO INSTITUTO DE SALUD PÚBLICA Y LABORAL DE NAVARRA
- vi. **RiPCoN** (€ 1 230 635)
Creación de perfiles de interacción rápida de 2019-nCoV para el aprendizaje profundo de reutilización de fármacos (deep drug-repurpose learning - DDRL) basado en redes

2) Proyectos sin participación española:

- i. **ATAC** (€ 3 020 040)
Terapia de anticuerpos contra coronavirus (COVID-19)
- ii. **CORESMA** (€ 2 759 195)
COVID-19-Outbreak Response combinando estrategias de e-health, serolomics, modelado y inteligencia artificial
- iii. **HG nCoV19 test** (€ 931 604,75)
Desarrollo y validación de una prueba de diagnóstico molecular rápido para nCoV19
- iv. **EpiPose** (€ 2 933 912,50)
Inteligencia epidémica para minimizar el impacto social, económico y de salud pública de 2019-nCoV en Europa
- v. **Fight-nCoV** (€ 2 807 026,25)
Lucha contra el coronavirus (SARS-CoV-2) mediante antivirales de amplio espectro: uso de modelo animal para ‘viral challenge’
- vi. **HERoS** (€ 2 855 410)
Respuesta a emergencias de salud en sistemas interconectados
- vii. **CoroNAb** (€ 2 771 657,50)
Nanocuerpos y anticuerpos contra 2019-nCoV
- viii. **Prevent-nCoV** (€ 3 399 387,50)
Prevención de la infección por SARS-CoV-2 mediante el desarrollo y las pruebas clínicas de una nueva vacuna de partículas similares a virus (VLP)
- ix. **RECoVER** (€ 5 590 820)
Respuesta rápida a la investigación de emergencia europea SARS-CoV-2
- x. **OPENCORONA** (€ 3 000 000)
Desarrollo rápido de terapias a través de Open nCoronavirus Vaccine Platform
- xi. **SCORE** (€ 2 627 551,25)
Respuesta rápida terapéutica de Coronavirus
- xii. **CORONADX** (€ 2 970 208,76)
Tres pruebas de diagnóstico rápido (Point-of-Care) para Coronavirus, mejorando la preparación para epidemias, la salud pública y los beneficios socioeconómicos

Proyectos IMI relevantes para COVID-19

IMI recibió 144 propuestas de proyectos para sobre COVID-19.

<https://www.imi.europa.eu/news-events/newsroom/8-new-covid-19-projects-announced>

De los cuales 8 proyectos iniciales fueron seleccionados en un proceso de “fast track” o evaluación rápida. El presupuesto del IMI para COVID-19 fue incrementados 45 millones iniciales aun total de 72 millones de Euros.

Los 8 proyectos ‘fast track’ son de dos tipos: relacionados con diagnóstico y relacionados con tratamientos.

Resumen proyectos relacionados con diagnóstico:

- i. RAPID-COVID (con participación española): desarrollo de una prueba de diagnóstico que pueda detectar simultáneamente el SARS-CoV-2 y otras 30 bacterias y virus respiratorios comunes.
- ii. DECISION (con participación española): desarrollo de un sistema de diagnóstico molecular desechable, miniaturizado y bajo costo que permite evaluar pacientes con una fiabilidad comparable a pruebas de laboratorio prácticamente en cualquier lugar pero dando los resultados en cuestión de minutos.
- iii. DRAGON: aplicará inteligencia artificial y aprendizaje automático para desarrollar un sistema de apoyo a la toma de decisiones para un diagnóstico y pronóstico mejorado y más rápido.
- iv. KRONO: desarrollo de una prueba simple que se puede utilizar en la consulta de un médico o en el hogar del paciente, generando resultados en solo 40 minutos.
- v. COVID-RED: uso de epidemiología clínica y dispositivos digitales (como dispositivos portátiles y aplicaciones móviles) para la detección rápida de potenciales casos priorizarlos para las pruebas tipo test.

Resumen proyectos relacionados con terapias:

- i. CARE (con participación española): detección de candidatos entre fármacos existentes que podrían ser efectivos como tratamientos para la pandemia COVID-19 (reposicionamiento/reutilización de fármacos) así como desarrollo de nuevos fármacos especialmente diseñados para combatir el virus SARS-CoV-2. Después de exhaustivas pruebas en el laboratorio, se llevará los fármacos candidato más prometedoras a ensayos clínicos en humanos.

- ii. MAD-CoV 2 (con participación española): caracterización y análisis de aspectos moleculares de SARS-CoV-2 para el desarrollo de nuevos tratamientos. Para lograr esto, se usarán técnicas de ingeniería de tejidos humanos (human engineering tissue) para probar nuevos tratamientos en el laboratorio, especialmente en relación con el receptor ACE2.

- iii. Impentri: Muchos pacientes con infección grave por COVID-19 experimentan una acumulación de líquido en los pulmones, lo que dificulta la respiración y, en el peor de los casos, contribuye al fallecimiento de los afectados. La respuesta inmune de los pacientes está implicada en parte en este proceso de acumulación de líquido. Hay indicios de que imatinib podría usarse para este escenario clínico. Se planea realizar un ensayo clínico para probar adecuadamente la eficacia y seguridad de este fármaco como tratamiento para pacientes con COVID-19 con inflamación pulmonar.