



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*

# The Spanish COVID-19 Data Portal.

**Alfonso Valencia**

**ICREA Professor**

**Dir. Dept Life Sciencies BSC**

2021



## Public Consortium

(Spanish + Catalan Government + UPC)

740 scientist and engineers

## Infrastructure

Host MareNostrum 4

RES, PRACE, INB-ELIXIR, EGA, IMPaCT-Data

The third Spanish Institution in EU funding

100 M € from H2020

EOSC-Life ELIXIR- Converge, HealthyCloud, PerMedCoE

## 4 Departments

Earth, Engineering, Computer Science & Life

## Life Sciences

6 Research Group + 5 support Units

130 scientist



# COVID-19 Data Portal



EMBL-EBI



Viral Sequences Host Sequences Expression Proteins Biochemistry Imaging Literature

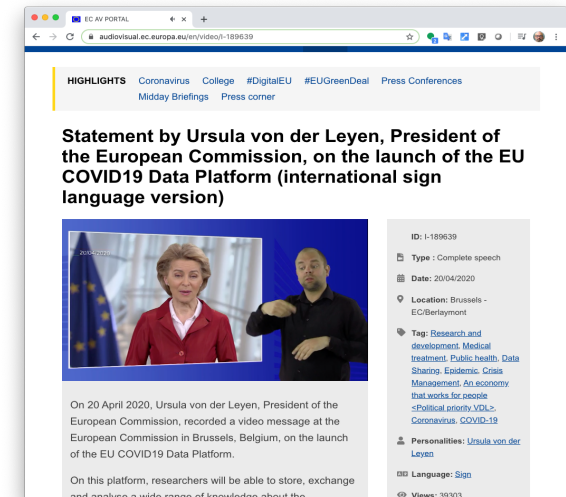
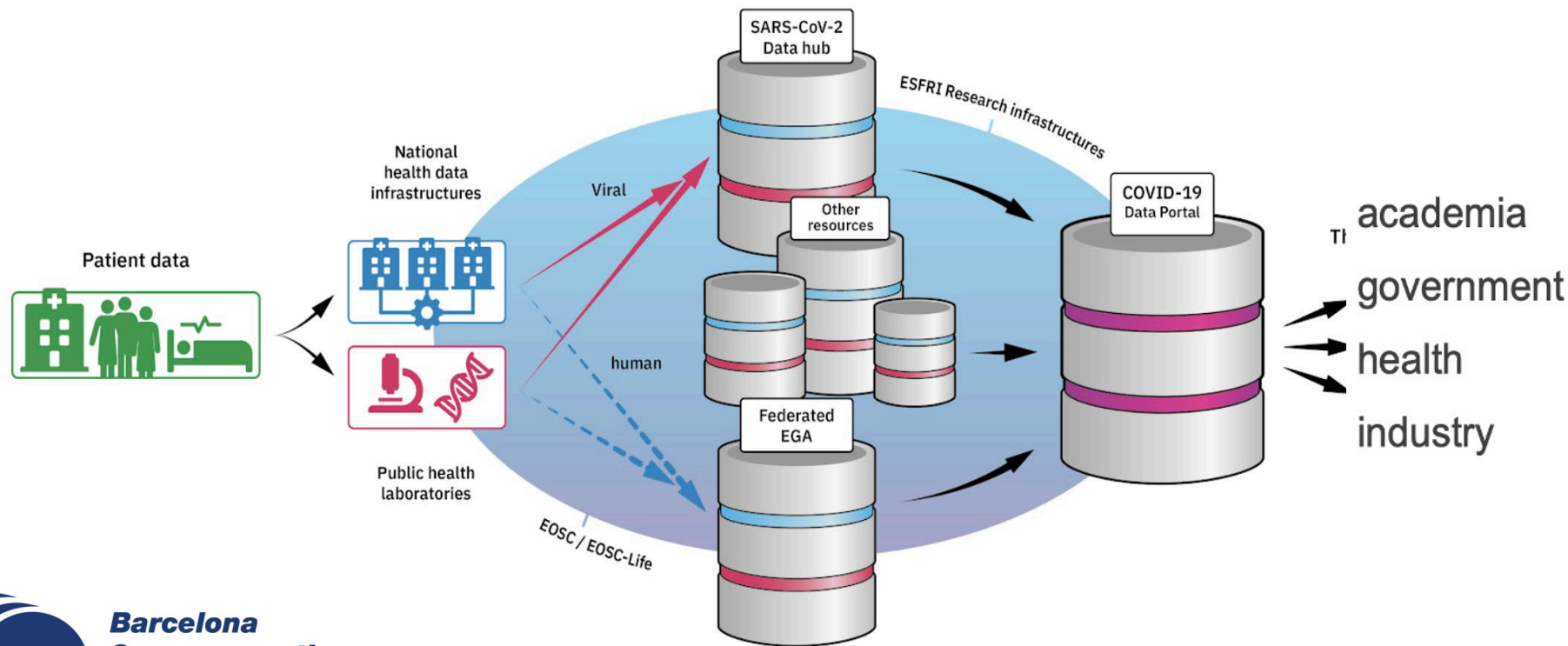
## Accelerating research through data sharing

Read and sign our letter in support of open COVID-19 data >



*"Three months ago, we were facing a pandemic of unknown proportions. But Europe's scientific talent and innovative capacity gave us hope. Through the ERAvsCorona Action Plan, we jointly mobilised European research and innovation to tackle the crisis, and we did it fast. We joined forces, scaled up our response, and it made us stronger."*

Mariya Gabriel, Commissioner for Innovation, Research, Culture, Education and Youth





# COVID-19 Data Portal SPAIN

[Submit](#)[en !\[\]\(3dfb8d66e81160ad61421a3452093d1b\_img.jpg\)](#) | [es !\[\]\(21ece2018b00c7267b3324c50bbed633\_img.jpg\)](#)[Genomics](#)[Resources](#)[Literature](#)[Research projects](#)[Health variables](#)[Working space](#)

## Accelerating research through data sharing

The **Spanish COVID-19 Data Portal** provides information, guidelines, tools and services to support researchers in utilising Spanish and European infrastructures for data sharing, in particular the [European COVID-19 Data Portal](#). For those interested in ongoing large research projects in Spain, we have compiled [a list of projects funded by major funding agencies](#).



Salvador Capella  
Albert Hornos  
Eva Alloza  
Alejandro Asensio

Isabel Cuesta  
Asunción Díaz Franco  
Teresa Gomez Garcia



*Thanks to:*

*EBI - Guy, Amonida & colleagues,*

***ELIXIR-Sweden - Johan Rung,***

*Galaxy Europe - DE: Björn Gruning + BE: Frederick Coppens,  
CRG > ViralBeacon + federated EGA Node > Jordi, Arcadi,  
Laureen Fromont, Marta Ferri, Mauricio Moldes & Fred Haziza.,  
working spaces: [UseGalaxy.ES](#) (Hugo Jiménez, Laia Codo, JL  
Gelpi), VRE (Laia Codo, JM Fernández, JL Gelpi) and others*





## ISC III COVID19

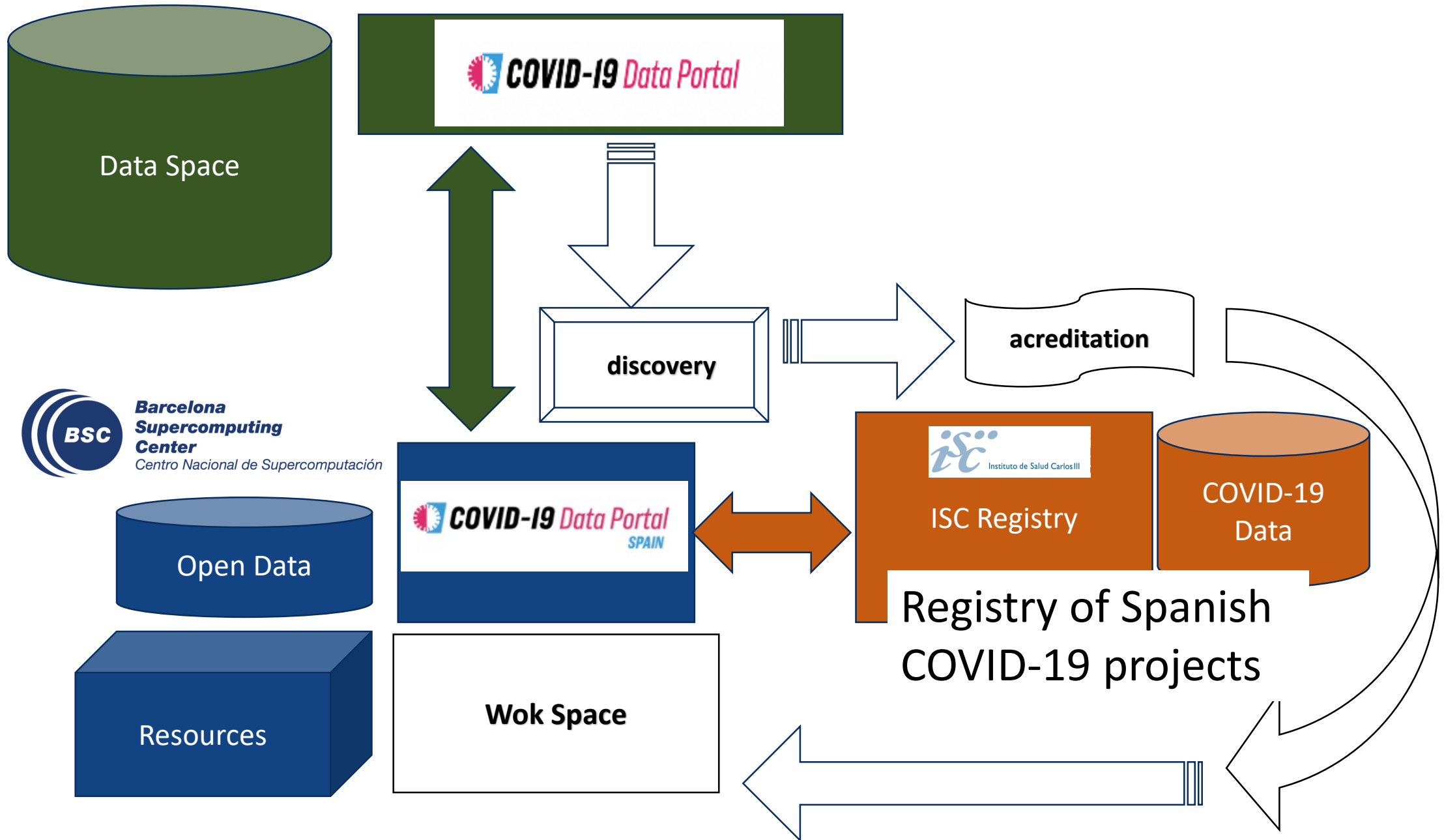
Genomics

king space

Accelerating research through data sharing

### Registry of COVID.19 founded projects









# Working Space

Genomics Res

pace

Accelerating research through data sharing

The **Spanish COVID-19 Data Portal** provides information, guidelines, tools and services to support researchers in utilising Spanish and European infrastructures for data sharing, in particular the [European COVID-19 Data Portal](#). For those interested in ongoing large research projects in Spain, we have compiled [a list of projects funded by major funding agencies](#).

## VRE

Virtual research environments (VREs) are online computational workbenches providing researchers access to relevant datasets, analysis pipelines and visualizers in an integrated manner.



Using [openVRE](#) framework, we are building a cloud-based VRE integrating data repositories and analyses relevant for COVID-19 research. This dedicated

## usegalaxy.es

UseGalaxy.es is the [Spanish National Bioinformatics Institute \(INB\)](#) GALAXY server powered by the [StarLife](#) computational cloud. The instance contains specific tools and workflows to analyse COVID-19 data, COVID-19 genomics and COVID-19 cheminformatics.



[Galaxy](#) is an open, web-based platform for

## COVID-19 Galaxy

Lorem ipsum dolor sit amet consectetur adipiscing elit, interdum erat primis feugiat rhoncus praesent, gravida duis magnis a ultricies sapien. Turpis congue iaculis porta nascetur faucibus mauris eros euismod, pretium eget mi molestie nulla varius.



Lacus aliquam enim eget ac facilisis pharetra luctus, in magna curabitur



# TOOLS

Genomics Resources Liter variables Working space

## Accelerating research through data sharing

The **Spanish COVID-19 Data Portal** provides information, guidelines, tools and services to support researchers in utilising Spanish and European infrastructures for data sharing, in particular the [European COVID-19 Data Portal](#). For those interested in ongoing large research projects in Spain, we have compiled [a list of projects funded by major funding agencies](#).



### Covid19 pathway interpretation and analysis

Web tool implementing a mechanistic model of human signalling for the interpretation of the consequences of the combined changes of gene expression levels and/or genomic mutations in the context of signalling pathways. It includes first versions of affected pathways by Covid-19.

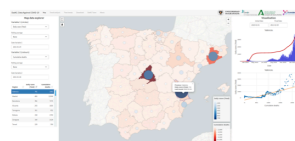
Developed by: *Clinical Bioinformatics Area, FPS*  
— Web-tool



### COVID-19 DisGeNET data collection

The text mining pipeline used in DisGeNET has been customized to scan the literature and identify any mention of genes, diseases and phenotypes, together with mentions of the COVID-19 virus. These mentions are normalized to standard vocabularies. This information is being updated regularly and is publicly available. DisGeNeT is an ELIXIR Recommended Interoperability Resource and part of the ELIXIR COVID-19 resources.

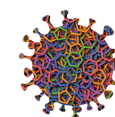
Developed by: *Integrative Biomedical Informatics, GRIB (IMIM-UPF) and MedBioinformatics Solutions*  
— Platform



### Data Against COVID-19

Platform that integrates information from COVID-19 with information on environmental and meteorological factors with temporal space aggregation by Spanish provinces and autonomous communities. Not only does DataAC centralize and integrate this type of data, but also implements different analysis and visual exploration functionalities that allow researchers to analyse and look for patterns among the different data sources.

Developed by: *Bioinformatics Unit, GENyO and Statistics Department, University of Granada*  
— Platform



### Chemical Checker

List of bioactive chemical compounds with potential to be effective against COVID-19. Active collection of suggested drugs from the current COVID-19 literature, with different levels of supporting evidence. [Chemical Checker](#) is used to identify small molecules with similar chemical and bioactivity features to the reported drugs in a universe of 800 thousand bioactive compounds. This resource is automatically updated every day and is intended for research only. Search engine improvement and drug annotations contributions are welcome.

Developed by: *Structural Bioinformatics and Network Biology group, IRB Barcelona in collaboration with the Amazon Search Science and AI group on the NLP tasks.*  
— Web-tool



### COVID-19 Flow-Maps

Flow-Maps is the Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS) system for monitoring COVID-19 outbreaks and mobility-associated risk by integrating health information, population-level mobility patterns into a Geographical Information System.

Developed by: *Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS)*  
— Web-tool



### COVID-19 Viral Beacon

Tool to find SARS-CoV-2 variability at genomic, amino acid and motif level. It has been developed as a branch of the GA4GH Beacon standard, as a special use case for testing and demonstration of new features in Beacon v2 (and implicitly of Beacon v1).

Developed by: *EGA team, CRG*  
— Web-tool



# Health variables

Home / Main Sections / Health variables

## COMMON VARIABLES FROM EHRs -ISARIC LIKE

Clinical variables - ISARIC based

Chronic diseases variables - cohorts based

COVID-19 ISCIII's funded projects variables

Download CSV

Show 10 entries

Search:

Concept	DT	Values	Term	St	
<b>SARS-COV-2</b>	CV	Positive/ Negative/ Equivocal	94315-9  SARS coronavirus 2 E gene [Presence] in Unspecified specimen by NAA with probe detection	LO	
<b>Hemoglobin</b>	PQ	g/dL	718-7 Hemoglobin [Mass/volume] in Blood	LOINC	Laboratory rel
<b>Leukocytes</b>	PQ	x1000/ $\mu$ L	6690-2 Leukocytes [# /volume] in Blood by Automated count	LOINC	Laboratory rel
<b>Lymphocytes</b>	PQ	x1000/ $\mu$ L	731-0 Lymphocytes [# /volume] in Blood by Automated count	LOINC	Laboratory rel
<b>Platelets</b>	PQ	x1000/ $\mu$ L	777-3 Platelets [# /volume] in Blood by Automated count	LOINC	Laboratory related
<b>Neutrophils</b>	PQ	x1000/ $\mu$ L	751-8 Neutrophils [# /volume] in Blood by Automated count	LOINC	Laboratory related
<b>Eosinophils</b>	PQ	x1000/ $\mu$ L	711-2 Eosinophils [# /volume] in Blood by Automated count	LOINC	Laboratory related
<b>Basophils</b>	PQ	x1000/ $\mu$ L	704-7 Basophils [# /volume] in Blood by Automated count	LOINC	Laboratory related
<b>Hematocrit</b>	PQ	%	4544-3 Hematocrit [Volume Fraction] of Blood by Automated count	LOINC	Laboratory related
<b>aPTT</b>	PQ	Sec	3173-2 aPTT in Blood by Coagulation assay	LOINC	Laboratory related

Showing 1 to 10 of 58 entries

Previous

1

2

3

4

5

6

Next



Hospital Universitario  
12 de Octubre

CLÍNICA  
BARCELONA  
Hospital Universitari

Hospital U  
vROCÍO



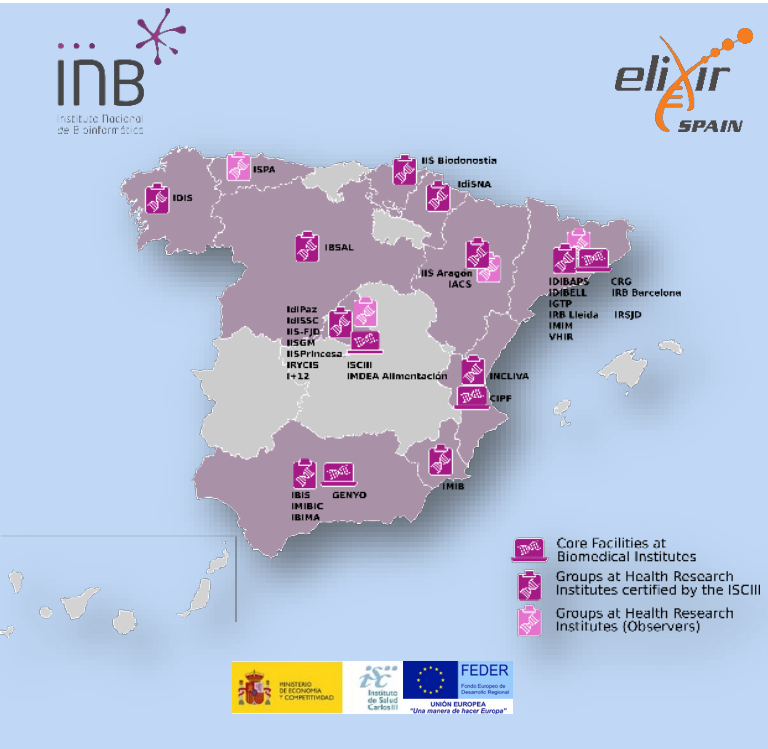
Barcelona  
Supercomputing  
Center


Centro Nacional de Supercomputación

# Collaboration with hospitals and health systems

## IMPaCT

*Building the framework for the National Personalized Medicine initiative by the Health Ministry*



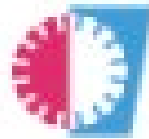
Acronym	
GENYO	
IBIMA	
IBIS	Instituto de Biomedicina de Sevilla
IMIBIC	Instituto Maimónides de Investigación Biomédica de Córdoba
IACS	Instituto Aragonés de Ciencias de la Salud
IIS Aragón	Instituto de Investigación Sanitaria Aragón
ISPA	Instituto de Investigación Sanitaria del Principado de Asturias
	
IRB Barcelona	Institute for Research in Biomedicine Barcelona
IRBLleida	Institut de Recerca Biomèdica de Lleida
IRSJD	Institut de Recerca Sant Joan de Déu
PSMAR-IMIM	Parque de Salud MAR (PSMAR) / Instituto Hospital del Mar de Investigaciones Médicas (IMIM)
VHIR	Vali d'Hebron Institut de Recerca
CIPF	Centro de Investigación Príncipe Felipe
INCLIVA	Institut d'Investigació Sanitària INCLIVA
IDIS / FPGMX	Instituto de Investigación Sanitaria de Santiago de Compostela / Fundación Pública Galega de Medicina Xenómica
IdISBa	Instituto de Investigación Sanitaria Islas Baleares
i+12	Instituto de Investigación Hospital 12 de Octubre
IdiPaz	Instituto de Investigación Hospital Universitario La Paz
IDIPHIM	IIS Instituto de Investigación Sanitaria Puerta de Hierro
IdiSSC	Instituto de Investigación Sanitaria Hospital Clínico San Carlos
IIS-FJD	IIS-Fundación Jiménez Díaz
IISGM	Gregorio Marañón Health Research Institute
IISPrincesa	Instituto de Investigación Sanitaria Hospital de la Princesa
IMDEA	Instituto Madrileño de Estudios Avanzados
IRYCIS	Instituto Ramón y Cajal de Investigación Sanitaria
ISCIII	Instituto de Salud Carlos III
IMIB	Instituto Murciano de Investigación Biosanitaria
IDISNA	Instituto de Investigación Sanitaria de Navarra
IISBiodonostia	Instituto de Investigación Sanitaria BioDonostia
ibs.GRANADA	Instituto de Investigación Biosanitaria de Granada
IDIVAL	Instituto de Investigación Sanitaria Marqués de Valdecillas
CNIC	Centro Nacional de Investigaciones Cardiovasculares
IIB Sant Pau	Institut d'Investigació Biomèdica de Sant Pau



Network of Translational Bioinformatics Groups in Research Institutes of Spanish Hospitals. (INB hosted)

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





# COVID-19 Data Portal

## SPAIN



## Viral Genomes

Genomics Res

space

Accelerating research through data sharing

### Genomics

Genomic resources

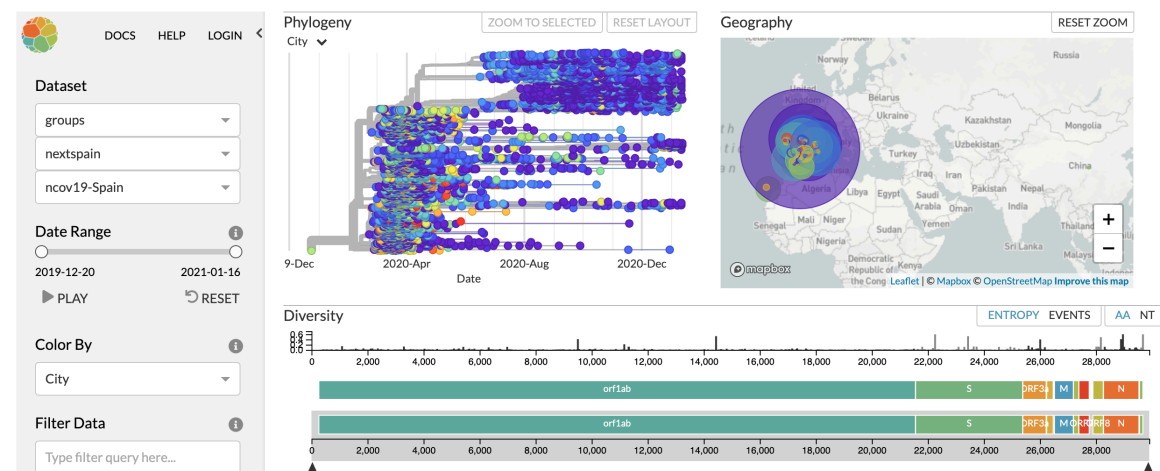
Viral sequences

Host sequences

### Genomic resources

[NextSpain COVID-19 interactive epidemiological map using Nextstrain.](#)

NextSpain COVID-19 shows the results obtained from the analysis of complete genome sequences of SARS-CoV-2 deposited in GISAID or other publicly accessible databases. The analyses are performed with the NextStrain engine to inform the scientific community and the general public about the genomic epidemiology of COVID-19 in Spain: [SeqCOVID Genomic epidemiology of COVID-19 in Spain](#). The epidemiological map is built with this code and it is maintained by [FISABIO Sequencing and Bioinformatics Service](#) and [Tuberculosis Genomics Unit \(IBV-CSIC\)](#).



and services to support researchers in utilising Spanish and European portals. For those interested in ongoing large research projects in Spain, we have

### Genomics

Genomic resources

Viral sequences

Host sequences

### Viral sequences

This section shows the samples sequenced from Severe acute respiratory syndrome coronavirus 2 (taxon: 2697049) in Spain. These results are periodically retrieved from ENA.

Powered by ENA  
European Nucleotide Archive

Showing 1 to 10 of 10,701 entries

Search:  [Download as CSV](#)

Accession	Collection date	Country	Host	Isolate	Description	TAXON
<a href="#">ERS4852070</a>	2020-03-22	Spain	Homo sapiens	COV005837	WGA sequencing of SARS-CoV-2 samples	2697049
<a href="#">ERS4852071</a>	2020-03-20	Spain	Homo sapiens	COV005838	WGA sequencing of SARS-CoV-2 samples	2697049
<a href="#">ERS4852072</a>	2020-04-01	Spain	Homo sapiens	COV005839	WGA sequencing of SARS-CoV-2 samples	2697049

# Access to SARS-Cov-2 genomes is essential to trace new variants and prevent outbreaks

## SCIENTISTS CALL FOR OPEN SHARING OF PANDEMIC GENOME DATA

But others say that certain restrictions encourage faster sharing.

*GISAID lacks: quality control, full genome information and & data reuse.*

Nature | Vol 590 | 11 February 2021 | **195**

In a letter released on 29 January (see [go.nature.com/3rtjgj5](https://go.nature.com/3rtjgj5)), Apweiler and others call for researchers to post their genome data in one of a triad of **databases that don't place any restrictions on data redistribution**: the US GenBank, the EBI's European Nucleotide Archive (ENA) and the DNA Data Bank of Japan, which are collectively known as the International Nucleotide Sequence Database Collaboration (INSDC).

500 scientists, including the 2020 chemistry **Nobel laureate Emmanuelle Charpentier**, and the head of the COVID-19 Genomics UK Consortium, Sharon Peacock.

**PRO** Anyone can anonymously access the INSDC's data and use them as they want, but GISAID requires that users confirm their identity and agree not to republish the site's genomes without permission from the data provider.

Some researchers told Nature that besides arguments about equity and openness, there is an issue with GISAID's differential control over how registered users can download its data.

### AGAINST

***"We really want to share our data, but it is heart-breaking and demotivating when we don't get the credit."***

Senjuti Saha, a microbiologist who works on SARS-CoV-2 genomes at the Child Health Research Foundation in Dhaka, says that she appreciates the call for open data beyond what GISAID offers, but worries that it might further dissuade researchers in low- and middle-income countries (LMICs) from uploading data until they have analysed them.



# COVID-19 as opportunity

- The **European Open Science Cloud and the European Health Data Space** addressing the problems of data access, quality and interoperability – including data protection.
- The **GDPR** was created to regulate and promote the **use of data for research** (secondary use), but some interpretations are very restrictive and harmful to the general interests.
- **Genome data are available at the EGA** (European Genome Phenome Archive). To access them, you must ask permission from an ethical committee that provides access codes.
- **The biggest problem for the advancement of biomedicine is access to medical information.** The **clinical data** stored in medical records (the largest experiment made by humanity is by nature **heterogeneous and irregular**, and largely in free text form, with **access technical, legal and practical difficulties**. A social, ethical, political, scientific and technical challenge.
- **FAIR-X** –beyond FAIR – equilibrated data sets, preventing biases in particular gender biases.





**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*

