BY-COVID POLICY BRIEF

Open data to support European pandemic preparedness

BY-COVID learnings and policy recommendations after 18 months of cross-disciplinary collaboration

Authors: Niklas Blomberg (ELIXIR) ORCID: 0000-0003-4155-5910, Marieke Willems (ELIXIR) ORCID: 0000-0001-9664-6225, Corinne S. Martin (ELIXIR) ORCID: 0000-0002-5428-2766, Andy Smith (ELIXIR) ORCID: 0000-0002-6025-9271, Elaine Harrison (ELIXIR) ORCID: 0000-0003-1149-2242, Carazo, Jose-Maria (CNB-CSIC) ORCID: 0000-0003-0788-8447, Enrique Bernal-Delgado (IACS) ORCID: 0000-0002-0961-3298, Iris van Dam (ScienSano) ORCID: 0000-0002-3965-6184, Romain David (ERINHA) ORCID: 0000-0003-4073-7456, Patricia M. Palagi (SIB) ORCID: 0000-0001-9062-6303, Martina Draščić Capar (CESSDA) ORCID: 0000-0001-8848-3008

May 2023

Contact: by-covid-admin@elixir-europe.org

Funded by the European Union

As seen during the COVID-19 pandemic, and other infectious disease outbreaks, researchers, health care professionals and "citizens (in terms of consent to share)" need to store, document, share, access, analyse, link and process research and clinical data across disciplines and national borders in a coordinated response. Like other infectious disease outbreaks, such as haemorrhagic fevers (e.g. cholera), COVID-19 will remain a societal challenge beyond the immediate outbreak, considering its destructive and disruptive impact on healthcare systems and the global economy. In addition to SARS-CoV-2, the pathogen at the source of COVID-19, the risk from other emerging pathogens also persists, which will require similar concerted action to identify and characterise infections with pandemic potential, and enable rapid public health action to mitigate health and societal impacts.

Provision of comprehensive open data on infectious agents and related diseases during outbreaks supports evidence-based decision-making across scientific, medical, public health and policy domains and promotes reproducibility of research outcomes.

European readiness for future pandemics is of utmost importance, and whilst preparedness for such eventualities requires provisions for e.g. rapid vaccine production and public procurement for personal protective equipment - far outside the scope of BY-COVID - the 'open data' aspects is a key component that should be addressed to ensure the preparedness of infrastructure as part of existing frameworks such as the European Open Science Cloud (EOSC). Contributing to enhance data sharing and utility for streamlined local to global public health decision-making and action, as defined by the World Health Organisation (WHO) as a key objective in the Global genomic surveillance strategy **(1)**.

In pandemic times, the mobilisation of raw viral sequences and the identification and monitoring the spread of SARS-CoV-2 variants is particularly important. The BeYond-COVID (BY-COVID) **(2,3)** project was funded by the European Union under the call "FAIR and open data sharing in support to European preparedness for COVID-19 and other infectious diseases" (HORIZON-INFRA-2021-EMERGENCY-01) **(4)** and will run until 2024, to make COVID-19 data accessible to research scientists and others such as medical staff in hospitals or government officials. The world has generated vast amounts of data in response to the COVID-19 pandemic, and is still generating more. This data comes from many different sources, and identifying, connecting and integrating it for effective analysis is challenging on many fronts.

This Policy Brief presents preliminary results from the BY-COVID project as part of its comprehensive, sustainable and evidence-informed plan to effectively promote and improve FAIR (Findability, Accessibility, Interoperability, and Reusability) and open data sharing in support for European preparedness for COVID-19 and other infectious diseases. The Brief also places the project and its results in the context of the upcoming EOSC Partnership, the development of the European Health Data Space (EHDS) and the European Health Emergency Preparedness and Response Authority (HERA).

Promoting open data sharing for pandemic preparedness

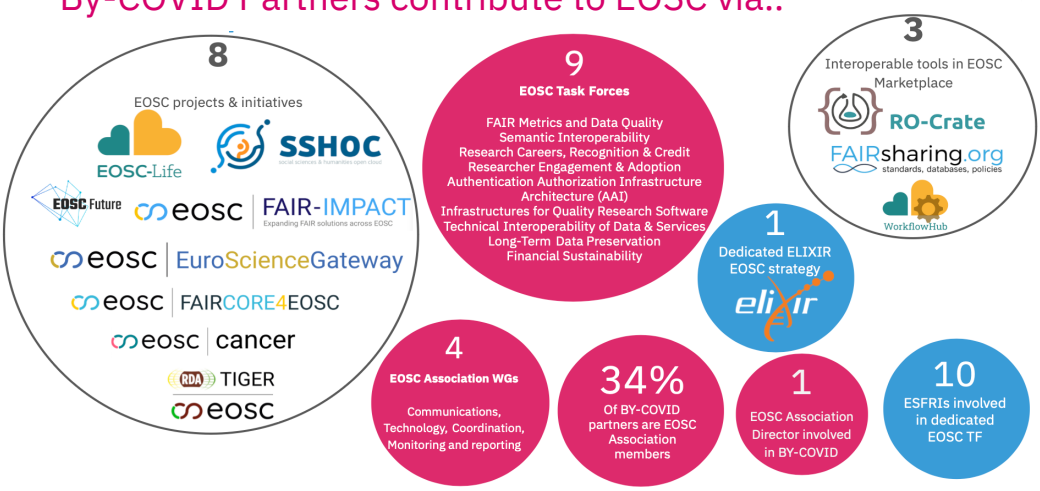
BY-COVID mobilises, connects, standardises, exposes and analyses COVID-19 data from many different sources to support European pandemic preparedness

To boost FAIR and open data sharing in support for European preparedness for COVID-19 and other infectious diseases, BY-COVID works to:

- Enable researchers, healthcare professionals and citizens (in terms of consent to share) fighting the spread of infectious diseases **to store, document, share, access, analyse, link and process research and clinical data across disciplines and national borders;**
- **Federate research and clinical data (human and viral) through national and international centres**, so as to enable pan-European and global sharing and hence research advances for better preparedness;
- **Develop the necessary digital tools and data analytics**, including for the identification and tracking of variants of concern, in support of public health action;
- **To improve linkages of FAIR data (and associated metadata) on pathogens, on their diseases and on their socio-economic consequences, considering a range of research fields**, such as omics, clinical, and epidemiological research, social sciences and humanities, so as to take a holistic approach to preparedness and response.

The BY-COVID project will also work to ensure that the above efforts are usefully **contributing to the European Open Science Cloud (EOSC) Partnership and the European Health Data Space (EHDS).** In the second half of the project, BY-COVID will publish a follow-on policy brief on its contribution to EOSC, providing lessons learned and future recommendations based on the EOSC monitoring framework sourced from BY-COVID partners involvement in EOSC projects, initiatives and dedicated task forces.

By-COVID Partners contribute to EOSC via..

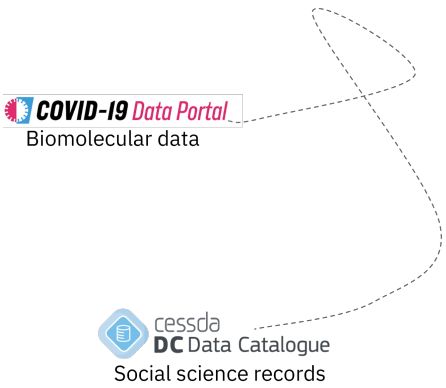


BY-COVID's contributions to pandemic preparedness

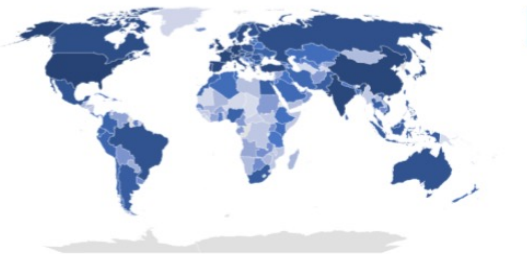
The COVID-19 pandemic has shown the importance of Open Science - from understanding the genetic basis of the virus and monitoring new mutations to tracking the spread of the disease across populations, open data has played a critical role in enabling new scientific knowledge and facilitating public health interventions.

The BY-COVID project is one of the Horizon Europe projects that supports the operation of the European **COVID-19 Data Platform (5)**, a critical resource providing access to open literature and data on both the virus and the disease, enabling researchers in their efforts towards e.g. viral characterisation, the development of vaccines and treatments, and variant analysis. For instance, the COVID-19 Data Portal, a component of the Platform, contains 16.9 million viral sequences and 950 000 open scientific articles related to COVID-19. The Portal has also been deployed to support the more recent mpox outbreak and is now forming part of the pandemic preparedness toolkit to address future pathogen outbreaks **(6)**. The Portal has proved to be incredibly popular with researchers across the globe.

There are a great number of national and EU-funded research projects relating to SARS-CoV-2 and COVID-19, each generating new knowledge and valuable data that must be publicly available and reused. BY-COVID provides the infrastructure that allows this to happen in the long-term. For example, an **indexing system has been developed as part of the project, allowing over 400 social science records (7) to be accessible through the European COVID-19 Data Platform (8)**.



COVID-19 Data Portal Statistics
March 2020-March 2023



	Visits	Visitors	Pages
TOTAL	8,862,599	470,368	4,394,136
Average monthly	239,530	12,713	118,760

This includes data that assess the impact of the pandemic on school education through public perceptions and acceptance of lockdown rules. This is a major achievement of the project, as well as a concrete implementation, through European funding, of a cross-cutting (social sciences and humanities in a project with strong life science focus). The BY-COVID project continues to index further data resources to support interdisciplinary research.

Project participants have collaboratively developed a new resource **- the Infectious Disease Toolkit (9)** - which helps scientists find the specific tools and guidelines they need for storing, sharing, accessing, analysing, linking and processing infectious disease data, allowing them to respond quickly to future disease outbreaks. A close and mutually-beneficial collaboration has been established with the ISIDORe **(10)** project (also funded through a COVID emergency response topic) to guide FAIR data management and provide metadata harmonisation, training and support to transnational user access to research infrastructures and enable publication of user data in the COVID-19 Data Portal. The two project has submitted a joint paper describing how to address FAIR data management in multidisciplinary projects, in particular addressing the challenge of managing data from transnational user access **(11)** .

Fostering long-term sustainability and wide uptake of its results, BY-COVID works on integrating them into EOSC and the EHDS, via targeted engagement, collaboration and communication mechanisms. BY-COVID also makes use of, and further develops, EOSC tools and resources e.g. those developed through the EOSC-Life project. For instance, the **BY-COVID project provides a framework and technologies such as FAIRsharing (12) , RO-Crate (13) & Workflow-Hub (14**) , for making data from other infectious diseases open and accessible to everyone. Finally, BY-COVID is closely collaborating with the EU-funded ELIXIR-CONVERGE project, which received an uplift destined to enhance human and viral data sharing through the ELIXIR Nodes and support the development of the European COVID-19 Data Platform) **(15)**.

BY-COVID's recommendations

To boost FAIR and open data sharing in support for European preparedness for COVID-19 and other infectious diseases, those developing policy and planning future activities for EOSC, EHDS and HERA (including the Pandemic Preparedness Partnership) should consider promoting the use of the following:

1. The European COVID-19 Data Platform

The European COVID-19 Data Platform consists of three components, the COVID-19 Data Portal, the SARS-Cov-2 Data Hubs, and the Federated EGA (European Genome Phenome Archive). The infrastructure built for the Platform is easily repurposed to cover other pathogens (e.g. mpox) **(16),** which makes it very suited for future pandemic preparedness. The main interface for this is the Pathogens Portal **(17)**. Also critical to this infrastructure are the Pathogens Data Hubs, which offer a complete service to data users, especially public health agencies and other scientific groups responsible for generating pathogen sequencing data. Services include data submission tools, analysis infrastructure and workflows, presentation tools, and visualisation services **(18)**. Key to the service is the integration and linking of data across data types (from molecular to clinical-epidemiological, cohort studies, and social sciences data). Deposition of all types of data is encouraged via the Data Hubs and the Portals promoting FAIR, open sharing and usage for the benefit of the global scientific community.

Encouraging researchers to use the Platform is an effective way to ensure that others benefit from open data. This can be done through mentions of the Platform in funding calls, e.g. the Horizon Europe topic 'HORIZON-HLTH-2023-DISEASE-03-07' encourages applicants to establish contact with the Platform in order to ensure that data is made available for others to use. Reference of the Platform can also be added to guides **(19)** for grantees, to ensure that data producers are aware of suitable places to deposit their data. National and EU funding agencies that fund research projects on infectious disease could do similar and reference the COVID-19 Data Platform in their national research agendas and funding programmes.

2. Infectious Disease Toolkit

The IDToolkit is built on the RDM Toolkit technology, developed as part of the EU-funded ELIXIR-CONVERGE project. It sources and shares best practices for rapidly managing data cross-domains and captures the pan-European knowledge from the infectious disease community brought together in BY-COVID, ISIDORe and EOSC4Cancer projects for future preparedness. Community members are encouraged to contribute to the IDToolkit via events such as a dedicated content-a-thon and content quality is overseen by an editorial board.

3. BY-COVID technologies across domains

FAIRsharing, RO-Crate and the WorkflowHub tools were developed in the EOSC-Life project **(20)**, funded under H2020 INFRA EOSC call, and adopted in the EOSC framework and its recommendations (see infographic). These tools are now taken forward in the BY-COVID project, funded in the INFRA EMERGENCY call, to be embedded in the European infectious disease preparedness as well as in other EOSC and applied projects. FAIRsharing.org allows the rapid construction of catalogues that capture data resources, standards and guidelines for a field while maintaining links across the domain oriented catalogues. It hosts guidelines for many journals and funders and has been used to build the data catalogue for BY-COVID. RO-Crate and the WorkflowHub.eu are essential to capture open, reproducible analysis workflows and make these portable and reusable across countries, institutes and compute centres.

The collaboration across domains within the BY-COVID project, particularly involving social science perspectives, allows for access to valuable insights into human behaviour, vulnerable populations, and the impact of different policies. This inclusive approach contributes to enhancing public health communication. By incorporating these insights into policymaking, there is an opportunity to develop more effective strategies for preventing and controlling infectious diseases.

To policymakers

1. Place Open Data at the heart of national pandemic preparedness plans aligned with ERA vs Corona Action Plan **(21)** and WHO Global genomic surveillance strategy.
2. Encourage consortia and projects generating data to deposit their data in the public domain, through RIs & databases such as the European COVID-19 Data Platform.
3. Provide adequate support for national data hubs and other infrastructures at the local level.
4. Ensure investments across life science domains to maintain & reuse community-driven knowledge sharing platforms, informing future projects & foster links with RI facilities.
5. Enable RI-led investments in technologies and their on-going adaptation to specific requirements for making infectious diseases data open and accessible to everyone.
6. Provide adequate training to foster the adoption of dedicated research tools, data sharing practices, and FAIR & metadata creation.

References

1. Global genomic surveillance strategy for pathogens with pandemic and epidemic potential, 2022–2032 https://www.who.int/publications/i/item/9789240046979
2. BY-COVID CORDIS page: https://cordis.europa.eu/project/id/101046203
3. BY-COVID website: https://by-covid.org/
4. HORIZON INFRA 2021 EMERGENCY 01 call: : https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-infra-2021-emergency-01
5. COVID-19 Data portal: https://www.covid19dataportal.org/
6. Pathogens Portal: https://www.ebi.ac.uk/ena/pathogens/v2/
7. Social Science Data in COVID-19 Data Portal: https://www.covid19dataportal.org/search/social-sciences
8. BY-COVID News: https://by-covid.org/news-events/by-covid-data-indexing-system/
9. Infectious Disease Toolkit: https://www.infectious-diseases-toolkit.org/index
10. ISIDORe project: https://isidore-project.eu/
11. David R, Richard AS, Connellan C, Lauer KB, Chiusano ML, Goble C et al. Umbrella Data Management Plans to integrate FAIR data: Lessons from the ISIDORe and BY-COVID consortia for pandemic preparedness. Data Science Journal. 2023 Jan 9;(submitted). doi: 10.5281/zenodo.7520086
12. FAIRsharing: https://fairsharing.org/3656
13. RO-Crate: https://www.researchobject.org/ro-crate/1.1/introduction.html
14. WorkFlowHub: https://workflowhub.eu/
15. https://elixir-europe.org/news/elixir-converge-receives-second-uplift-covid-19-variant-surveillance ; https://elixir-europe.org/about-us/how-funded/eu-projects/converge/wp9
16. Monkeypox pipeline news on BY-COVID website: https://by-covid.org/news-events/monkeypox-pipeline/
17. Pathogens Portal: https://www.ebi.ac.uk/ena/pathogens/v2/
18. Rahman, N. et. al. bioRxiv 2023.04.19.537514; doi: https://doi.org/10.1101/2023.04.19.537514
19. EOSC-Life project: https://www.eosc-life.eu/
20. EC grantee guide on COVID-19 Data Portal use: https://by-covid.org/pdf/Horizon-Europe-Data-Platform-flyer-March-2023.pdf
21. ERA vs. Corona Action Plan: https://www.esfri.eu/latest-esfri-news-covid-19-news/era-vs-corona-action-plan

by-covid.org

@BYCOVID_eu

/company/by-covid/

admin@elixir-europe.org

Funded by the European Union (GA 101046203). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

BY-COVID