Deliverable D8.2

Data Management Plan

Project Title (grant agreement No)	Beyond COVID Grant Agreement 101046203		
Project Acronym (EC Call)	BY-COVID		
WP No & Title	WP8: Coordination, Project Management and Ethical, Legal and Social implications		
WP Leaders	Andrea Troncoso, ELIXIR; Juan Arenas, ELIXIR		
Deliverable Lead Beneficiary	7 BBMRI		
Contractual delivery date	31/03/2022	**Actual Delivery date**	30/03/2022
Delayed	No		
Partner(s) contributing to this deliverable	ELIXIR empirica		
Authors	Eva Garcia Alvarez, Michaela Th. Mayrhofer, Peter Holub - BBMRI		
Contributors	—----		
Acknowledgements (not grant participants)	—---		
Reviewers	Katharina Lauer, ELIXIR Nina Van Goethem, Sciensano Ramon Launa Garces, IACS		

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

Log of changes

Date	Mvm	Who	Description
10/02/2022	V0	Eva Garcia Alvarez, Michaela Th. Mayrhofer, Peter Holub	
17/02/2022	V0/1	Katharina Lauer	
23/03/2022	V0/2	Nina Van Goethem, Ramon Launa Garces	
29/03/2022	V1	Eva Garcia Alvarez, Michaela Th. Mayrhofer, Peter Holub	

Table of contents

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

1. Executive Summary

The Data Management Plan (hereinafter “the project DMP”) is a mandatory deliverable by the European Commission (EC) that has to be tailored to the specific project activities and which should document the relevant project repositories and how they are managed. The Data Management Plan for BY-COVID is developed by WP8 as an integral part of the governance structure, and also oversees the quality assurance of the project outcomes. BY-COVID will not generate new sensitive patient level data but integrate and link existing resources. Hence, responsibility remains on data access committees and processes for each dataset. The project DMP considers:

- Overview of the project data flow
- Compliance with Open Science and FAIR sharing
- Ethics Requirements as required according to the Grant Agreement (WP9)

This document is BY-COVID Data Management Plan (DMP) version 0.1 delivered in month 06. It outlines the data types that will be collected or generated and how these will be handled, processed and shared. It describes the standards that will be incorporated and the related methodology for data collection, processing and data sharing. This deliverable is based on the template and the guidelines provided by the European Commission for Horizon Europe.

The project DMP is a live document that will be reviewed and updated periodically to ensure it remains up to date.

Funded by the European Union

BY-COVID is funded by the European Union’s Horizon Europe research and innovation programme under grant agreement number 101046203.

2. Contribution towards project objectives

The project handbook defines the project process that provides the framework to accomplish all projects objectives within the scope, budget and the required level of quality. This deliverable contributes to all objectives as listed below:

	Key Result No and description	Contributed
Objective 1 Enable storage, sharing, access, analysis and processing of research data and other digital research objects from outbreak research.	1. A research data management practice in European research infrastructures practice that drives discovery, access and reuse of outbreak data and directly links experimental data from HORIZON-INFRA-2021-EMERGENCY-02 transnational access projects into the COVID-19 Data Portal.	Yes
	2. Workflows and processing pipelines that integrate transparent quality management and provenance and are openly shared.	Yes
	3. Research infrastructures on-target training so that users can exploit the platform.	Yes
	4. Engagement so that stakeholders (RI, national centres, policy makers, intergovernmental organisations, funders and end-users) incorporate FAIR and open data in infectious disease guidelines and forward planning.	Yes
Objective 2 Mobilise and expose viral and human infectious disease data from national centres.	1. A comprehensive registry of available data with established procedures to collate data governance models, metadata descriptions and access mechanisms in a pandemic scenario.	Yes
	2. Mechanisms for the initial discovery across data sources based on available metadata at the reference collection.	Yes
	3. Demonstrated transnational linking of real-world data from national surveillance, healthcare, registries and social science data that allow the assessment of variants to serve the research needs of epidemiology and public health.	Yes

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

	4. Demonstrated assessment of emerging SARS-CoV-2 variants against data generated in the on-going European VACCELERATE clinical trials project to investigate vaccine efficacy.	Yes
Objective 3 Link FAIR data and metadata on SARS-CoV-2 and COVID-19	1. A platform that links normative pathogen genomes and variant representations to research cohorts and mechanistic studies to understand the biomolecular determinants of variant response on patient susceptibility, and disease pathways.	Yes
	2. An open and extensible metadata framework adopted cross-domain that supports comprehensive indexing of the infectious disease resources based on mappings across resources and research domains.	Yes
	3. A provenance framework for researchers and policy-makers that enables trust in results and credit to data submitters, workflow contributors and participant resources.	Yes
Objective 4 Develop digital tools and data analytics for pandemic and outbreak preparedness, including tracking genomics variations of SARS-CoV-2 and identifying new variants of concern.	1. Broad uptake of viral *Data Hubs* across Europe deliver an order-of-magnitude increase in open viral variant detection and sharing.	Yes
	2. Infrastructure and quality workflows mobilised and shared to produce open, normative variant data that is incorporated into national and regional data systems and decision making.	Yes

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

Objective 5 Contribute to the Horizon Europe European Open Science Cloud (EOSC) Partnership and European Health Data Space (EHDS).	1. Guidelines and procedures for FAIR data management and access will be established, building on work of other guideline producing consortia such as the Global Alliance for Genomics and Health (GA4GH), the 1Mio Genomes Initiative (1MG) and the Beyond One Million Genomes project (B1MG).	Yes
	2. Services, software, protocols, guidelines and other research objects that are openly accessible for reuse by the EOSC Association and the community at large as a foundation for European preparedness for infectious diseases, leveraging developments in EOSC-Life, SSHOC, EOSC-Future, EGI-ACE and other EOSC projects.	Yes
	3. Alignment (both policy and implementation routes) will have been achieved between the data governance strategies for routinely collected health data in the EHDS initiative, including the TEHDAS Joint Action and future EHDS Pilot Actions.	Yes
	4. To empower national centres to build capacity and train platform users and data providers (e.g., from life, social or health sciences), and with experts from across partner institutions collaborating to create training materials for the identified gaps, and to exchange experiences and knowledge.	Yes

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

3. Methods

The integrated ecosystem that BY-COVID will deliver is based on established services and resources, ensuring sustainability beyond the lifetime of the project. The adoption of open source platforms on commonly used cloud-based infrastructures and connection of the components through open standards supports long-term sustainability in a distributed landscape.

The Ethics Requirements (D9.1-D9.4) have been assessed as required according to the Grant Agreement (WP9) in a survey per Work Package.

The respective data access committees and processes for each dataset remain accountable form the basis for project operations. Data security remains the responsibility of the databases and archives holding data (e.g., the EGA has a defined security process and follows best practice guidelines aligned with the GA4GH Security Working Group, the BBMRI-ERIC Policy for Access to and Sharing of Biological Samples and Data and follows OECD Council Recommendations on Health Data Governance). IP background and existing data licences for the project might be diverse. IP rights will be managed in the consortium agreement. Where code, software or ontologies are developed, a permissive licence will be applied and an open approach will be taken by partners in line with their institutional policies. While access rights and licences predating the BY-COVID project will be respected, participating organisations will be encouraged to review existing licences to comply with Open Access requirements. Data sharing and reuse will follow FAIR principles and make use of licences such as the Creative Commons or Open Data Commons. Where open access policies can not be applied instantly, BY-COVID will follow established embargo principles (e.g., in the European Nucleotide Archive, ENA; release after 6 or 12 months).

4. Description of work accomplished

4.1 Project overview

BY-COVID aims to identify, connect and integrate data for the effective study of the COVID-19 disease and causative agent as well as other infectious diseases. Infectious diseases are complex and their analysis requires data from different disciplines. Thus, BY-COVID will link established and emerging research infrastructures and data resources from biomolecular research, public health, clinical research and social science, using standards fully aligned with the European Open Science Cloud (EOSC).

Three main pillars can be identified in the project methodology:

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

Mobilisation - Mobilised data will be indexed and organised in the COVID-19 Data Platform (https://www.covid19dataportal.org/), where data is incorporated through a flexible, tiered system for data integration. Data incorporated into the COVID-19 Data Platform will be embedded in the wider EOSC data ecosystem, establishing guidelines and procedures for FAIR data management and ensuring long term, rapid open access.

BY-COVID will allow mobilisation - meaning access and transfer - of data by using a trialled system of SARS-CoV-2 Data Hubs and other existing infrastructures such as the European Nucleotide Archive (ENA), CESSDA social science archives, and biobank catalogues in a "federation of federations", following community practised standards. Hence, ultimate data responsibility belongs to the data providers, reflected in their own DMP or data best practises guidance.

Connect and expose - A tiered indexing system will be developed, protecting truth and privacy, specially in the case of sensitive data. Guidelines for ensuring data interoperability will be established by implementing community-driven standards, together with offering support to the integration of COVID-19 data resources in the COVID-19 Data Portal.

Use & analyse data - BY-COVID will integrate standardised data management and analysis methods and protocols to ensure FAIR (Findability, Accessibility, Interoperability, and Reusability) and FAIR-Health are an integral part of the process.

4.2 Data flow

The data flow is distributed among the different technical Work Packages (WP1-WP2-WP3-WP4-WP5), flowing in a domain specific manner.
WP1 will establish and improve SARS-CoV-2 Data Hubs and handle centralised data; namely 1) non-identifying biomolecular data, which are not sensitive and 2) non-identifying non-biomolecular patient data without dedicated data resources to date, which are not sensitive after de-identification. Both types of data will be managed in SARS-CoV-2 Data Hubs, enriching and extending them for data sharing functions and other non-sensitive biomolecular data types.

WP2 will harmonise domain specific data. Thus, these data are subjected to different levels of sensitivity. Non-patient related data is non-sensitive and can be openly shared. Simplification of the access to sensitive human/patient bio-molecular data will be possible through Federated EGA, BBMRI infrastructures, and Estonian and Dutch cohorts. To further streamline this access, WP2 will determine how data governance procedures can be improved. Interoperability is essential to expedite this process, and it will be addressed

Funded by the European Union
BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

using standards such as Global Alliance for Genomics and Health (GA4GH) Data Use Ontology or Researcher Passports.

Sensitive human/patient clinical and health data will be managed in national research infrastructures, including Federated EGA, PHIRI, ECRIN, BBMRI. A specific task (Task 2.3) will address governance and access mechanisms, meta-data descriptions, and procedures to solve technical, legal and organisational barriers. This task will address the secondary use of data for research purposes by (i) identifying key contextual information on how data is generated, curated and managed; (ii) providing guidance on standards for subsequent de identification purposes, e.g. 2-level pseudonymisation and, (iii) providing guidance on standard pipelines for data curation and quality assurance, including semantic interoperability.

Socio-economics data sources used in BY-COVID can be sensitive, such as contact tracing data. These will be de-identified at source before reused and will be held securely by Research Infrastructure data resources, including Federated EGA, PHIRI and BBMRI. Sensitivity will be evaluated for population-level data that may or may not be sensitive and will be managed in national or research infrastructures. Finally, non-sensitive aggregated or ecological data will be managed in national or regional collections and in research infrastructures (e.g., EIRENE or BBMRI as part of environmental biobanking).

An initial data harmonisation step at domain level across jurisdictions of the data and the metadata will be done by WP2. The level of harmonisation will depend on the public availability of the data. For data under access control, the harmonisation will be done at the metadata level while for publicly available data sources the harmonisation can include the data as well. All elements and access mechanisms to enable semantic interoperability will be collected and provided to the reference catalogue at WP3.

WP3 will set the framework for interoperability across domains and define common metadata formats. The FAIRsharing resource will assist with the selection of broad standards, such as Bioschemas. In addition, domain- and infrastructure-specific standards from WP2 and WP4, existing mappings, work in previous initiatives (EOSC-Enhance, SSHOC, FAIRsFAIR) and the collaborations in EOSC-Future will be leveraged.

This Work Package will support findability by indexing into common databases, a crucial process for a successful connection of different data types. The indexing will consist of different tiers. Tier 1, the most detailed, will capture granular record-level identifiers of non-sensitive data, metadata and attributes. Notably, sensitive data will stay at source and will be only indexed at metadata level. Tier 2 will provide metadata and attributes at a dataset or study level. Tier 3 will provide resource level metadata based on a curated FAIRsharing Collection, which will also be a key element in the EOSC integration through collaboration with openAIRE.

Funded by the European Union
BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

The indexing system will be built based on the EMBL-EBI Search, with a satellite implementation by BBMRI/UMCG in the BBMRI Directory based on Molgenis technology. The web services provided by this task will be consumed by the COVID-19 Data Portal for interactive users, but also openly accessible by external users, for example allowing external installations/consumers to discover additional local data in the context of the central COVID-19 Data Portal.

WP4 will collect reproducible and open access analysis methods and protocols and make them openly available to all researchers via an Infectious Diseases Toolkit. Workflows for data quality control support and COVID/19 data analysis will be accessible through the WorkflowHub.eu. Specifically, tools for evaluating quality of data will be determined or developed, and existing services such as CQL for HL7 FHIR or the QC in the Data Hubs will be used to define quality assurance levels. Regarding analysis, integration and visualisation of COVID-19 data, open workflows based on the Galaxy Project will be established, using the indexing developed by WP3. Federated analysis strategies will be implemented to preprocess or analyse sensitive data on national infrastructure and the EOSC-Life provenance model will be used as a basis for an infectious disease provenance model. Provenance will be captured at a sufficient level to preserve the data derivation chain back to original sources; identifiers will interlink provenance of data to the source of biological samples (WP2) and parts of provenance traces will be kept confidential or anonymised in a distributed model (ISO 23494 standard series under development).

Suitable repositories (e.g., Zenodo, Biostudies, etc) will provide access to RO-Crate, a model for data-centric research objects used by EOSC-Life that implements FAIR Digital Objects using established open standards, that will be used to package data, provenance and workflows. The provenance, WorkflowHub and RO-Crates contribute to the metadata and indexing framework of WP3 including the Bioschemas metadata framework and data and workflow citation.

WP5 will interrogate data from different domains to demonstrate the usability of the BY-COVID FAIR data ecosystem. Data from WP1, WP2, indexed by WP3, will be reused in order to address specific research questions. In addition, previously available data from already established sources (e.g., Sciensano, IACS, PHIRI Data Hubs or COVID-19 Portal entries) will be analysed as part of the BY-COVID use cases. The respective data access (ethical review) committees and/or other processes for each dataset will be completed prior to accessing the data to execute the analytical pipelines on those sources.
The processing of the individual level data will mainly run federated at the institutions hosting and/or having easy access to linked individual level data (Data Hubs), while the visualisation, network analysis, modelling and expression analysis of open access non sensitive/aggregated data will be done in a centralised way, in collaboration with WP4. The

Funded by the European Union
BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

analyses will be performed with open access tools and develop new solutions when needed. As a result, the use-cases will produce non-sensitive data, such as causal models, Common Data Models, synthetic data sets to guide implementation of analytical algorithms, matrices with matching weights (e.g., inverse probability weights), estimands of effect (e.g., average treatment effect), interpretation of omics visualisation or mechanistic modelling of cellular signalling routes.

BY-COVID partners will provide the computational and storage capacities, together with supplementary resources from an agreement with the EGI-ACE project.

4.3 Open Science and FAIR sharing

BY-COVID partners have broad experience in Open Science practises:

- Open data access:
  - European Nucleotide Archive
  - omicsDI
  - MDR
- Open standards and FAIR practises:
  - FAIRsharing
  - RDMkit
  - FAIR Cookbook
  - Bioschemas
  - GA4GH
- Open software and technologies:
  - WorkflowHub.eu
  - Galaxy project
  - RO-Crate
  - Common Workflow Language
- Open community-driven collaborative practises:
  - ELIXIR Biohackathons
  - Galaxy smorgasbord

From these Open Science Practices, suitable repositories (e.g., WorkflowHub, RO-crate) and persistent and unique identifiers (e.g., digital object identifiers) will ensure findability and discoverability. These procedures will also allow effective sharing, together with the use of standards such as Bioschemas and Common Workflow Language.

Furthermore, EOSC practises (e.g., EOSC Interoperability framework, EOSC Enhance D4.3 “Analysis of existing research data cataloguing efforts towards integrated discovery”) and

Funded by the European Union

BY-COVID is funded by the European Union’s Horizon Europe research and innovation programme under grant agreement number 101046203.

RDA recommendations (consortium partners were co-leading RDA COVID-19 working group) will be followed using open standards and metadata in WP2 and WP3.

Accessibility will be ensured by the European COVID-19 Data Platform, where data will be as open as possible, considering data protection requirements. Non-sensitive data (e.g., pathogen genome, variant data) will be linked with research infrastructure data in WP2. Metadata of sensitive data will be openly accessible and connected via open standards for cataloguing and indexing (WP3). Notably, data sources are responsible for monitoring correct execution of access rights and adherence to ethical and legal requirements.

This project will drive data use and re-use by linking FAIR open data to workflow environments and providing access to analysis and visualisation tools building trust and reproducibility with provenance and quality assurance mechanisms. Thus, the project won't generate any new sensitive patient level data and each participant institution is responsible for accounting for national legislations, the administrative provisions and the implemented data access procedures. Nonetheless, if any new data is generated (specially from WP5 use cases) it will be handled in the next version of the DMP.

5. Conclusions

Overall, during BY-COVID, data access committees and processes for each dataset remain accountable for the basis of the project's operations. Currently, no active participant recruitment/engagement or AI systems are used. The current DMP, including the Ethics Requirements (Annex 2, status 23.03.2022), will be reviewed to ensure it continues to be fit for purpose and that any changes introduced in the repositories or in the process to handle personal data (contact details) are incorporated into the document.

6. Next steps

Review before the mid term review or earlier if needed / requested.

7. Impact

A sound and established Data Management Plan constitutes the foundation for a fruitful collaboration across WPs, with External Experts and Stakeholders.

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

Annex 1

The Project Management repositories do not store research data, but project data required to manage, monitor and control the project execution. ELIXIR intends to use these repositories to monitor and control BY-COVID project administrative, legal, financial and technical aspects providing a collaborative workspace to generate project deliverables, milestones and manage internal and external communication activities. At the same time, it aims to foster collaboration across project participants.
The BY-COVID Project Management Repository Structure is presented in Table 1.

Table1. BY-COVID Project Management Repository Structure

Top level folder	Content
1. Project MASTER File	Master document with details from the project: contact lists, effort distribution, lists of deliverables & milestones, GANTT chart etc Project monitoring spreadsheet - contains details of all actions, events, deliverables and milestones, along with due dates and templates - monthly reports are generated from this for all partners/WPs. It includes the email address of the partners participants that are relevant for the monitoring and control activities: PIs, Deputies, administrative, financial and legal contacts
2. Legal Documents	Legal documents pertaining to the project: Consortium Agreement, Grant Agreement, Description of Action, amendments, contracts etc
3. WPs	Working folders for the WPs
4. Deliverables & Milestones	Repository for project deliverables and milestones (templates, working drafts and submitted documents)
5. Project meetings, Events & TCs	Details of all project meetings and events (agendas, minutes slide presentations etc)
6. Periodic Reporting	Financial and periodic reports from all project partners

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

7. Guidance and Templates	Project handbook, containing guidance pertaining to all aspects of the project (processes, communications, management structure and responsibilities etc) Project templates: deliverables, milestones, agendas, minutes, slide presentations etc
8. Project Communications and Outreach Materials	Project branding & style guidelines, press releases, articles, presentations, newsletters, graphic resources, etc
9. Grant Agreement Preparation	Documentation related to the grant agreement preparation phase.

Access: All project participants that have registered and consented get editorial access to the project repository hosted and managed by the ELIXIR Project Management Unit Team. External experts involved in the project by the WP leaders can get commenting rights to the project Google Drive. This form is called “Registration of Participation in BY-COVID (EC HE 101046203)” and it is accessible here: https://docs.google.com/forms/d/e/1FAIpQLScCmcPkWJcj3A_PPln7E4-iLf2wmDlb5r9B_IUpd3ypwe0Xiw/viewform

Funded by the European Union

BY-COVID is funded by the European Union’s Horizon Europe research and innovation programme under grant agreement number 101046203.

Annex 2 (click link for a better view of the spreadsheet): WP9 Ethics Requirements D9.1-D9.4 (status 23.03.2022)

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
D9.1 H - Requirement 1	Do you actively recruit/involve human participants?	YES/NO reply	NO		NO		NO	NA		NA
	If yes, specify briefly which procedures and criteria (including stakeholder research activities) that involve human participants are in place (incl. links to procedures, policies, etc)	Open-Ended Response					N/A			
	Are informed consent procedures implemented for the participation of humans in all research activities?	YES/NO reply					N/A			
	If yes, please provide a description of the procedure (or link to relevant document)	Open-Ended Response					N/A			

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
	If yes, note that the templates of the informed consent and information sheets (in a language and terms intelligeble to the participants) must be kept on file. The templates must include all relevant information regarding the protection of data. This is in the responsibility of the WP and/or partner using the IC. 1. Please provide here a contact (name, email) in case the templates are requested by the EC. 2. If public information, please provide a link to the IC templates and participant information	Open-Ended Response					N/A			
	Are children involved?	YES/NO reply					N/A			
	If yes, please provide details on how the consent of the legal representative (and assent of the research participant, when applicabble) will be obtained (ideally link if public, if not pdf)	Open-Ended Response					N/A			

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
	Copies of opinions/approvals by ethics committees and/or competent authorities for all activities involving humans must be obtained before the start of that activity and must be kept on file. 1. Please provide a contact (name, email) in case the templates are requested by the EC.	Open-Ended Response					N/A			
	Detailed information about the research activities performed in the demonstrator projects in WP5 must be submitted as a deliverable, including clarification as to whether they involve research with human participants.	Open-Ended Response	N/A	N/A	N/A	N/A	N/A This use case uses linked population & health care data (resuse of existing datasets from several data hubs). No research with human participants takes place	N/A	N/A	N/A
D9.2 - HCT - Req 2	In case **human biosamples** are obtained **within the project or from another project**, details on the type of biosamples must be kept on file. To be obtained before the start of that activity and must be kept on file.	Open-Ended Response					N/A	NA		NA
	For the **human biosamples obtained from a biobank**, details on the biosamples types and on the biobank and access to it must be kept on file.	Open-Ended Response					N/A			

Funded by the European Union

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
	If applicable, copies of relevant documents for using, producing or collecting human biosamples (e.g., ethics approval, import licence, accreditation/designation/authorisation/licensing) must be kept on file.	Open-Ended Response					N/A			
	If yes, please provide a contact (name, email) and where the details are kept on file (e.g biobank xy).	Open-Ended Response					N/A			
D9.3 - POPD	Are special derogations pertaining to the rights of data subjects or the processing of genetic, biometric and/or health data have been established under the national legislation of the country where the research takes place relevant for the activities in the respective WP?	YES/NO reply					YES	NA		NA

Funded by the European Union

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
	If yes, a declaration of compliance with respective national legal framework(s) must be provided here.	Open-Ended Response					IACS: Data controller in BIGAN (IACS) invokes derogation (j) of Article 9 (2) in line with the purpose of this processing by the data hubs and in accordance with the national legislation for the processing of the population and healthcare data needed Sciensano: Linkage of individual-level data within the LINK-VACC project was approved by the medical ethics committee UZ Brussels – VUB on 03/02/2021 (reference number 2020/523) and obtained authorization from the ISC Social Security and Health (reference number IVC/KSZG/21/034).			
	Please provide the contact details of the Data Protection Officer (DPO) per organisation hosting data (if already kept on file in respective WP, pls add shortcut to documentation)	Open-Ended Response					DPO IACS: protecciondatos.iacs@aragon.es DPO Sciensano (roel.heijlen@sciensano.be)			
	The **contact details of the DPO** must be made available to all data subjects involved in the project. For Host Institutions not required to appoint a DPO under the General Data Protection Regulation, a detailed data protection policy for the project must be provided.	Open-Ended Response					N/A			

Funded by the European Union

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
	Why is it relevant to process the data that you process and how is the principle of 'data minimisation' ensured?	Open-Ended Response					Linking real world population and health care data to SARS-CoV-2 variants: vaccination->infection & variants->severity; common data model requires the minimum data to reply to the research questions			
	Please provide a description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants. In particular, **a detailed description of the COVID-19 Data Platform (WP3) system architecture and of its data flow,** including connections with local centres must be provided.	Open-Ended Response			The portal will include metadata and open data on available services, workflows and emerging research data. It will not include personal data on human participants.		Movement of scripts and aggregated results between orquestrator (IACS) and Data Hubs (IACS, Sciensano, THL, UT, GOG, RIVM). Data results received by the orquestrator (IACS) are either: a) aggregated, thus it's impossible to trace back to individuals whose data is concerned b) individual anonymous data in the case of the matrix with matching weights (e.g., inverse probability weights)			

Funded by the European Union

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
	Please provide a detailed description of the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing (describe status of development or provide link to respective policy)	Open-Ended Response					(Security measures implemented for in-country analyses at each data hub to produce comparable aggregate and summary data that can be pooled and integrated by the orquestrator (IACS) into the federated infrastructure model are explain by each hub) IACS: Security measures section at: https://www.iacs.es/actividad-tratamiento-bigan/ (In Spanish) Sciensano: The selected variables from the individual data sources within the LINKVACC project are kept in the pseudonymized environment of healthdata.be and under a project mandate. Access rights to the pseudonymized data in the healthdata.be data warehouse are granted ad nominatum for the scientists involved in the surveillance activities at Sciensano.			
	Please provide a description of the anonymysation/pseudonymisation techniques that will be implemented (brief description or link to respective document/policy)	YES/NO reply					(Anonymysation/pseudonymisation techniques implemented are described by each data hub) IACS: triple anonymisation ; https://www.iacs.es/actividad-tratamiento-bigan/ (In Spanish) Sciensano: A link between the individual data takes place thanks to the use of a pseudonymized national reference number managed by healthdata.be.			

Funded by the European Union

WP9 - Ethics Requirements			WP1 - suppport for virological analyses in emerging disease outbreaks	WP2 - Accessing heterogeneous data across domains and jurisdictions for enabling the downstream processing of COVID-19 and future pandemic episodes data	WP3 - COVID-19 integration platform	WP4 - Connecting the COVID-19 data platform to analysis tools and local portals	WP5 - A continuously evolving demonstrator project feeding the changing research questions that surface during an on-going pandemic to solutions WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP6 - engage, train and build capacity with national stakeholders	WP7 - outreach, communication, industry and public engagement	WP8 - Coordination, PM & ELSI
	Is personal data transferred from the EU to a non-EU country or international organisation, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679?	YES/NO reply					NO			
	If yes, note that confirmation must be kept on file. Please provide contact (name, email) and/or where to find information.	Open-Ended Response					N/A			
	Are personal data transferred from a non-EU country to the EU (or another third state)?	YES/NO reply					N/A			
9.4 - AI	Please provide a detailed description of all the AI systems, techniques, and activities involved in the project (link to document, policy or similar)						N/A	NA		NA
	If any AI system, technique, or activity is involved in the project, provide a detailed explanation on how respect of fundamental human rights and freedoms will be ensured (eg. policy)						N/A			

Funded by the European Union

WP9 - Ethics Requirements			WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP5.4.1 (domain-specific use case on epidemic hotspots: Wastewater Based Epidemiology) UNIMIB	WP5.4.1 (domain-specific use case on epidemic hotspots: Alert_Cov) UNIMIB	WP5.4.2 (domain-specific use case on disease maps) UNILU
D9.1 H - Requirement 2	Do you actively recruit/involve human participants?	YES/NO reply	NO	NO	NO	NO
	If yes, specify briefly which procedures and criteria (including stakeholder research activities) that involve human participants are in place (incl. links to procedures, policies, etc)	Open-Ended Response	N/A	N/A	N/A	NA
	Are informed consent procedures implemented for the participation of humans in all research activities?	YES/NO reply	N/A	NO	NO	NO
	If yes, please provide a description of the procedure (or link to relevant document)	Open-Ended Response	N/A	N/A	N/A	NA
	If yes, note that the templates of the informed consent and information sheets (in a language and terms intelligeble to the participants) must be kept on file. The templates must include all relevant information regarding the protection of data. This is in the responsibility of the WP and/or partner using the IC. 1. Please provide here a contact (name, email) in case the templates are requested by the EC. 2. If public information, please provide a link to the IC templates and participant information	Open-Ended Response	N/A	N/A	N/A	NA
	Are children involved?	YES/NO reply	N/A	NO	NO	NO
	If yes, please provide details on how the consent of the legal representative (and assent of the research participant, when applicabble) will be obtained (ideally link if public, if not pdf)	Open-Ended Response	N/A	N/A	N/A	NA
	Copies of opinions/approvals by ethics committees and/or competent authorities for all activities involving humans must be obtained before the start of that activity and must be kept on file. 1. Please provide a contact (name, email) in case the templates are requested by the EC.	Open-Ended Response	N/A	N/A	Guarrantor of Privacy: Opinion on the draft Directive of the president of Istat on "Identification of the processing of personal data referred to in Articles 9 and 10 of Regulation (EU) 2016/679 in the context of the alert-cov statistical work - June 24, 2021 [9681795] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9681795	NA

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

WP9 - Ethics Requirements			WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP5.4.1 (domain-specific use case on epidemic hotspots: Wastewater Based Epidemiology) UNIMIB	WP5.4.1 (domain-specific use case on epidemic hotspots: Alert_Cov) UNIMIB	WP5.4.2 (domain-specific use case on disease maps) UNILU
	Detailed information about the research activities performed in the demonstrator projects in WP5 must be submitted as a deliverable, including clarification as to whether they involve research with human participants.	Open-Ended Response	N/A This use case uses linked population & health care data (resuse of existing datasets from several data hubs). No research with human participants takes place	N/A This use case collect quantitative and qualitative virl data, Not human data	Epidemiological surveillance is the normal process of collecting, analysing and disseminating health data for public health purposes. One of the purposes of communicable disease surveillance is to detect epidemic hotspots in order to start timely interventions. The importance of spatial hotspots (or clusters) in infectious disease epidemiology has been increasingly recognized, since targeting hotspots is an important component of outbreak-control strategies. In the case of airborne infections, interventions may include infection containment measures through individual and collective isolation, vaccination campaigns, prophylactic treatment of subjects at risk, timely expansion of hospital receptivity. The accurate and timely prediction of epidemic hotspots results in the reduction of severe clinical outcomes (e.g., mechanical ventilation or death) and in avoiding the use of strict containment measures with the consequent social and economic implications. According to these premises, the Alert_CoV is an ongoing epidemiological research project aimed to develop a system for early detection of COVID-19 epidemic hotspots in Italy. The project is coordinated by the University of Milano-Bicocca, and currently involves the Italian National Institute of Statistics (ISTAT), the Italian National Health Institute (Istituto Superiore di Sanità, ISS) and six Italian Regions (i.e., the project covers almost one third of the entire Italian population). In order to identify potential predictors of epidemic hotspots, several early markers will be considered, among which, for example, chest x-rays, paracetamol prescriptions, calls at the hospital emergency room, social media and Google search for keywords related to COVID-19 disease and its symptoms. The data relating to the early markers will be considered in aggregate form for each of the more than 100 thousand census sections (each including, on average, about 150 resident individuals) included in the current study. The analytical process consists in monitoring and counting the early markers of interest until a situation suggestive of the onset of an epidemic hotspot is highlighted, generating, thus, a potential alarm signal.	NA
D9.2 - HCT - Req 2	In case **human biosamples** are obtained **within the project or from another project**, details on the type of biosamples must be kept on file. To be obtained before the start of that activity and must be kept on file.	Open-Ended Response	N/A	N/A	N/A	NA
	For the **human biosamples obtained from a biobank**, details on the biosamples types and on the biobank and access to it must be kept on file.	Open-Ended Response	N/A	N/A	N/A	NA
	If applicable, copies of relevant documents for using, producing or collecting human biosamples (e.g., ethics approval, import licence, accreditation/designation/authorisation/licensing) must be kept on file.	Open-Ended Response	N/A	N/A	N/A	NA
	If yes, please provide a contact (name, email) and where the details are kept on file (e.g biobank xy).	Open-Ended Response	N/A	N/A	N/A	NA

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

WP9 - Ethics Requirements			WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP5.4.1 (domain-specific use case on epidemic hotspots: Wastewater Based Epidemiology) UNIMIB	WP5.4.1 (domain-specific use case on epidemic hotspots: Alert_Cov) UNIMIB	WP5.4.2 (domain-specific use case on disease maps) UNILU
D9.3 - POPD	Are special derogations pertaining to the rights of data subjects or the processing of genetic, biometric and/or health data have been established under the national legislation of the country where the research takes place relevant for the activities in the respective WP?	YES/NO reply	YES	NO	YES	NO
	If yes, a declaration of compliance with respective national legal framework(s) must be provided here.	Open-Ended Response	(For data hubs to reply)	N/A	see https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9681795 The Privacy Guarantor, in consideration of the epidemiological emergency from COVID-19 and the need and urgency to have timely, reliable and complete official statistics on epidemiological phenomena to support interventions to combat the health emergency and those aimed at managing the recovery phase, has authorized Istat to process, possibly in joint ownership with other subjects that are part or parcel of the National Statistical System, personal data also inherent to particular categories of data. 1. pursuant to art. 58, par. 3, lett. b) of the Regulations and art. 13, paragraph 2 of the d.l. 19 May 2020, n. 34, converted into law n. 77, expresses a favorable opinion on the draft directive of "Identification of personal data processing referred to in Articles 9 and 10 of Regulation (EU) 2016/679 as part of the alert-cov statistical work carried out pursuant to art. 13 of the law decree n. 34/2020 converted with amendments by law 17 July 2020, n. 77 "	NA
	Please provide the contact details of the Data Protection Officer (DPO) per organisation hosting data (if already kept on file in respective WP, pls add shortcut to documentation)	Open-Ended Response	(For data hubs to reply)	responsabile.protezionedati@iss.it	responsabile.protezionedati@iss.it	NA
	The **contact details of the DPO** must be made available to all data subjects involved in the project. For Host Institutions not required to appoint a DPO under the General Data Protection Regulation, a detailed data protection policy for the project must be provided.	Open-Ended Response	N/A	N/A		DPO of the University of Luxembourg: Sandrine Munoz.dpo@uni.lu
	Why is it relevant to process the data that you process and how is the principle of 'data minimisation' ensured?	Open-Ended Response	Linking real world population and health care data to SARS-CoV-2 variants: vaccination->infection & variants->severity; common data model requires the minimum data to reply to the research questions	Genomic SARS-CoV2 data (quantitative data and qualitative data) from wastwater surveilance are important for monitoring and predicting epidemic hotspots. No human data collected.	The importance of spatial hotspots (or clusters) in infectious disease epidemiology has been increasingly recognized, since targeting hotspots is an important component of outbreak-control strategies.	We process aggregated (anonymised) molecular data on expression of biomolecules in COVID-19 affected patients; we ensure data minimisation by processing only the data on: molecular identifiers, group effect size (log 2 fold change of expression in case-control setup) and adjusted p value.

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.

WP9 - Ethics Requirements			WP5.2.1 (Baseline use case on vaccination and infection) and WP5.2.2 (Baseline use case on variants and severity)	WP5.4.1 (domain-specific use case on epidemic hotspots: Wastewater Based Epidemiology) UNIMIB	WP5.4.1 (domain-specific use case on epidemic hotspots: Alert_Cov) UNIMIB	WP5.4.2 (domain-specific use case on disease maps) UNILU
	Please provide a description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants. In particular, **a detailed description of the COVID-19 Data Platform (WP3) system architecture and of its data flow,** including connections with local centres must be provided.	Open-Ended Response	Movement of scripts and aggregated results between orquestrator (IACS) and Data Hubs (IACS, Sciensano, THL, UT, GOG, RIVM). Data results received by the orquestrator (IACS) are either: a) aggregated, thus it's impossible to trace back to individuals whose data is concerned b) individual anonymous data in the case of the matrix with matching weights (e.g., inverse probability weights)	The Data Owner or Data Provider can impose different accessibility regimes to different User Groups. 2. The exceptional accessibility regimes consists of the following: a. Providing access to aggregated data only; b. Providing access to filtered single measurement data only; c. Providing access to generalised single measurement data only; d. Restricting access to data for a limited time period (i.e. embargo period); e. Providing metadata only; f. Providing access to single measurement data, subject to conditions, obligations and safeguards to comply with data protection legislation. g. Using a combination of the above (i.e. combination of (a) to (e). Even if exceptional accessibility regimes are imposed, metadata shall always remain retrievable through the EU4S-DEEP platform for all User Groups. Access to wastewater monitoring data retrievable through the EU4S-DEEP platform may be subject to registration and authentication when they are subject to specific rules of a Project Group or when different Exceptional Accessibility Regimes are imposed to different User Groups. The Authentication Service of the European Commission (EU Login) is used to control the Users' registration and authentication.	The data is provided aggregated by census section which on average consists of 150 individuals	We rely only on aggregated (anonymised) datasets, processing of individual datasets will be performed by the infrastructure of WP4.
	Please provide a detailed description of the security measures that will be implemented to prevent unauthorised access to personal data or the equipment used for processing (describe status of development or provide link to respective policy)	Open-Ended Response	(Security measures implemented for in-country analyses at each data hub to produce comparable aggregate and summary data that can be pooled and integrated by the orquestrator (IACS) into the federated infrastructure model are explain by each hub)	N/A	see https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9681795 The variables selected from the data sources are stored "in a protected and encrypted environment". Access rights to data in the data warehouse are granted to scientists involved in surveillance activities.	https://wwwfr.uni.lu/layout/set/print/universite/data_protection/data_protection_policy_and_faq
	Please provide a description of the anonymysation/pseudonymisation techniques that will be implemented (brief description or link to respective document/policy)	YES/NO reply	(Anonymysation/pseudonymisation techniques implemented are described by each data hub)	Surveillance of wastewater for signals of SARS-CoV-2 does not require the collection of individual names particularly when sampling the sewer network serving a large population as in the case of this project.	Collection of data on early markers of epidemic hotspots does not require the collection of individual names.	NO
	Is personal data transferred from the EU to a non-EU country or international organisation, confirmation that such transfers are in accordance with Chapter V of the General Data Protection Regulation 2016/679?	YES/NO reply	NO	NO	NO	NO
	If yes, note that confirmation must be kept on file. Please provide contact (name, email) and/or where to find information.	Open-Ended Response	N/A	N/A	N/A	NA
	Are personal data transferred from a non-EU country to the EU (or another third state)?	YES/NO reply	N/A	NO	NO	NO
9.4 - AI	Please provide a detailed description of all the AI systems, techniques, and activities involved in the project (link to document, policy or similar)		N/A	N/A	N/A	NA
	If any AI system, technique, or activity is involved in the project, provide a detailed explanation on how respect of fundamental human rights and freedoms will be ensured (eg. policy)		N/A	N/A	N/A	NA

Funded by the European Union

BY-COVID is funded by the European Union's Horizon Europe research and innovation programme under grant agreement number 101046203.