EOSC-Life

EOSC-Life: Building a digital space for the life sciences

D14.1 – Strategic plan for the development of a COVID-19 repository including specification of technical requirements, policies and procedures

WP14 – Design, development, implementation and use of a repository for individual participant data from COVID-19 trials
Lead Beneficiary: ECRIN and UiO
WP leader: Jacques Demotes & Gard Thomassen
Contributing partner(s): ECRIN, UiO

Authors of this deliverable: **S. Canham, C. Ohmann, G. Thomassen, M. Matei, J. Demotes, M. Panagiotopoulou**

Contractual delivery date: **30 September 2020**
Actual delivery date: **8 October 2020**
H2020-INFRAEOSC-2018-2

Grant agreement no. 824087
Horizon 2020
Type of action: RIA

Table of Contents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Executive Summary

The deliverable highlights the necessity of data sharing in the context of the COVID-19 pandemic and provides a “high level” concept document for the design, development, implementation and use of a repository for Individual Participant Data (IPD) from COVID-19 clinical trials. In the following pages, the most important aspects for setting-up and operating the repository such as legal challenges, functional specifications, quality assurance, implementation plan, sustainability and governance, evaluation of routine use, outreach, partnerships and scalability are discussed in detail.

Project Objectives

The overall aim of EOSC-Life WP14 is the design, development, implementation and use of a repository for Individual Participant Data from COVID-19 clinical trials that is compliant with European regulations and in particular with the GDPR.

With this deliverable, the project has reached the following objective:

a) The “high-level” concept of the repository is described in detail, including the legal context of data object sharing in COVID-19 research, functional specification, quality assurance, implementation plan, sustainability and governance, evaluation of routine use and impact, usability and user friendliness, outreach, partnerships and scalability.

Detailed Report on the Deliverable

1. Background and Scope

In recent years, many scientific organisations, funders and initiatives have expressed their commitment to more open scientific research. This cultural shift has been extended to also include clinical research and clinical trials in particular. Today, the results of clinical trials are increasingly considered as a public good, and access to the individual participant data (IPD) generated by those trials for further research is viewed by some as part of a fundamental right to health data, even if such a "right" does not exist from a legal perspective (Ohmann et al, 2018). Sharing data from clinical research can be justified on scientific, economic and ethical grounds. Data sharing generates and promotes more science and may result in better use of funding. Better use of data helps to better use healthcare resources, plan services more effectively, develop more evidence-based interventions and ultimately lead to better care for patients (O’Connell and Plewes, 2015). Sharing can therefore increase data validity, but it also squeezes more value from the original research investment, as well as helps to avoid unnecessary repetition of studies. The economic advantages of data reuse are one reason why governmental and intergovernmental agencies, as well as major research funders now support data sharing. Ethically, data

This project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 824087.

sharing provides a better way to honour the generosity of clinical trial participants, because it increases the utility of the data they provide and thus the value of their contribution (Ohmann et al, 2018).

To support the sharing of IPD in clinical trials, several organisations have developed generic principles, guidance and practical recommendations for implementation in recent years. Within the EU Horizon 2020 funded project CORBEL (Coordinated Research Infrastructures Building Enduring Life-science Services) and coordinated by the European Clinical Research Infrastructure Network (ECRIN), an interdisciplinary and international stakeholder taskforce reached a detailed consensus on principles and recommendations for data sharing of clinical trial data. One of the major principles formulated is that data and trial documents made available for sharing should be transferred to a suitable data repository to help ensure that the data objects are properly prepared, are available in the longer term, are stored securely and are subject to rigorous governance (Ohmann et al, 2017).

During a pandemic crisis, there is a need for timely and accurate collection, reporting and sharing of data within and between research communities, public health practitioners, clinicians and policymakers. Accurate and rapid availability of data will inform assessment of the severity, spread and impact of a pandemic to implement efficient and effective response strategies (RDA, 2020). The WHO statement on data sharing during public health emergencies clearly summarises the need for timely sharing of preliminary results and research data (WHO, 2015). There is also strong support from a series of funders and publishers for recognising open research data as a key component of pandemic preparedness and response. In the recommendations from RDA it is clearly stated that to facilitate data quality control, timely sharing and sustained access, data should be deposited in data repositories. Whenever possible, these should be trustworthy data repositories (TDRs) that have been certified, are subject to rigorous governance, and committed to longer-term preservation of their data holdings. By allocating persistent identifiers and requiring pre-specified formats and rich metadata, certified trustworthy repositories can guarantee a baseline FAIRness of data, as well as providing sustained access and a standardised citation (RDA, 2020).

There are different types of repositories available: general scientific repositories (e.g. DRYAD, Figshare), repositories dedicated specifically to clinical research (e.g. Vivli), repositories specialising in a specific disease area (e.g. BioLINCC) and institution-specific repositories (e.g. Edinburgh Datashare). Only very few, however, deal specifically with IPD sharing of COVID-19 trials. IDDO and ISARIC are partners in a global collaboration established to collect and share COVID-19 clinical data to inform clinical practice and public health response. The system will accept IPD datasets, collected as part of clinical care and follow-up, clinical trials or observational research. It is a data platform, where the data from multiple studies are included in a single database with a consistent and harmonised data structure (IDDO, 2020). The activity has no focus on the European area, however, and its standardised data structure means that it will not be easily applicable to many COVID-19 trials. Vivli, one of the major repositories to hold and share IPD data, is committed to advance the knowledge around the COVID-19 epidemic by waiving all fees for sharing and accessing clinical trials (Vivli, 2020). Unfortunately, Vivli is not fully compliant with the GDPR and thus applicability for European studies is limited.

What is still missing is a trusted repository dedicated specifically to all kinds of COVID-19 studies, focused on the European region (but not exclusively) and without restrictions on the standardisation of data, allowing secure and efficient data sharing and optimal use of clinical trial data of COVID-19 trials for re-analyses, for secondary analyses, and for patient-level data meta-analyses. The object of EOSC-LIFE WP14 is to provide such a repository, but not in isolation of the other services available - it will be of

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

major importance to explore harmonisation of procedures, services and tools between Vivli, the IDDO/ISARIC platform and the planned repository.

The work plan envisaged for this project is focused on delivering a pilot COVID-19 repository as early as possible. This is feasible due to the intensive preparatory work already done within H2020 CORBEL and other projects by ECRIN and by partners (ELIXIR, UiO). This work includes:

- The evaluation of existing repositories (Banzi et al. 2019)
- The definition of policies and processes for data sharing of IPD (Ohmann et al. 2017)
- The definition and assessment of quality criteria for trusted repositories (Tilki et al. 2020; Lin et al. 2020)
- The development of a business plan for a clinical study repository (CORBEL: D3.8[1])
- The exploration of technical solutions for repositories (Banzi et al. 2018)
- The development and launch of a metadata repository[2] for clinical trials and related data objects (2020)
- ECRIN's participation in the RDA COVID-19 Group recommendations for data sharing
- Production of a comprehensive overview of the ELSI ecosystems for sharing and reuse of clinical research and healthcare data in Europe (EOSC-Life WP4, co-led by ECRIN and BBMRI)
- A workshop on anonymization of health research and healthcare data in the context of the GDPR, (January 2020, Paris) within EOSC-Life WP4, and associated report
- International partnership with US data sharing repositories, in particular with Vivli.

Building on this experience, the objective of this WP is to first define the specifications, and then develop, implement and routinely operate a repository for individual participant data from COVID-19 trials, compliant with European regulations and in particular with the GDPR, allowing clinical trial data sharing after completion of the trial. Such a repository will be part of the European COVID-19 data hub, with a portal operated by ECRIN acting as the interface with the clinical research community, and with technical partners providing a secure environment and data sharing services. Development will include discussions with stakeholders (e.g. scientific organisations, funders, policy makers) to identify the most appropriate design and procedural options in the context of the GDPR and other relevant regulations, and the best business model for the long term.

2. Stakeholders to be involved

Through the planned stakeholder forum, the COVID-19 repository should address the interests of all stakeholders involved. These include:

a) Researchers: The researcher is a major stakeholder in the process of data sharing even if the decision to share data is usually made by the data controller, which can be a hospital or a sponsor (see b). For a researcher who may want to deposit data, it is important to have a trusted repository available, where he/she can deposit his/her data securely. The repository should have policies and procedures in place that guarantee legal and regulatory compliance, and guide researchers through the data preparation and uploading process. It should be implemented in a way that supports rewards for data sharing and allocate PIDs to the data objects, when necessary, for unique

[1] https://zenodo.org/record/3862715#.X3WgXZNKgbl
[2] https://ecrin.org/clinical-research-metadata-repository

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

identification. It should provide a flexible range of access methods. Researchers may also want to access IPD, and should therefore also be involved in the development of clear and transparent policies governing data access.

b) Trial sponsors/data controllers: Sponsors are usually the data controllers for the processing of personal data carried out in clinical trials, and must ensure that the legal obligations involved in data sharing resulting from the GDPR and CTR are fulfilled. Sometimes the trial sponsor and the investigating centres act as joint data controllers. The infrastructure, policies and procedures of the planned repository will need to demonstrate compliance with legal requirements to foster the necessary commitment from sponsors and, where necessary, from investigating centres.

c) Policy makers (e.g. government departments, especially public health agencies): Data holds the potential to drive rapid response and informed decision-making during public health emergencies, and to prevent wasteful initiatives based only on political calculations. There is a need for timely and accurate collection, reporting and sharing of data within and between research communities, public health practitioners, clinicians and policymakers, centred around a repository for IPD from COVID-19 trials.

d) Regulatory bodies: Proactive sharing of clinical trial data has long been a key strategic aim of the European Medicines Agency (EMA). The planned repository could contribute to an update of the EMA data sharing policy, currently discussing the possibility of sharing individual participant data from clinical trials. It is therefore essential to have a representative of EMA in the stakeholder forum.

e) Funders: Funders have an essential role in supporting timely data sharing in general and especially in health emergencies such as COVID-19. They have a key role in promoting the use of trustworthy data repositories for data sharing, built upon standards and assessed according to defined quality criteria (e.g. certification with CoreTrustSeal), and their involvement in the project is therefore essential. There will also be a need to develop, with input from funders, a sustainable business plan.

f) Publishers: Publishers require publishing of the data, software and code underlying a study, in an even more timely manner than usual in a pandemic situation. Transferring the data to trusted repositories, keeping them privately and securely, and allowing reviewers to inspect the data could be a way of better supporting the peer review process. The provision of such a service could be explored with publishers.

g) Patient organisations: Patient organisations should be actively involved in the planning and implementation of the COVID-19 repository to assure that their interests are considered. Equally, patient organisations could be major promoters of data sharing and could help to convince their members and the public to consent to data sharing if involved in clinical trials. They also have an important role to play in helping to balance the perceived scientific and social benefits of data sharing with the individual rights of the participants.

h) Scientific associations: ISARIC (International Severe Acute Respiratory and Emerging Infection Consortium) is a global federation of clinical research networks whose purpose is to prevent illness and deaths from infectious disease outbreaks. In response to the emergence of novel coronavirus it has developed a portfolio of resources to accelerate research. The Infectious Diseases Data Observatory (IDDO) brings together members of the global infectious disease community in the generation, analysis and application of research data. IDDO and ISARIC are partners in a global collaboration that collects and shares COVID-19 clinical data to maximise its utility. Both organisations should be represented in the stakeholder forum. In the context of the planned COVID-19 repository, possible cooperation with IDDO/ISARIC will be explored as part of the project.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

i) Standardisation bodies: It is a major target of the COVID-19 repository project to be compliant with common standards. CDISC has published the Interim User Guide for COVID-19 describing the most common biomedical concepts relevant to COVID-19, and the necessary structures to represent such data consistently, using CDASH and SDTM. ECRIN has itself formulated an extended metadata schema for clinical research data objects, which will be applied in the COVID-19 repository. This schema is based upon DataCite, a widely used metadata standard. For these reasons, representatives from CDISC and DataCite will be invited to join the stakeholder group.

j) COVID-19 platforms/initiatives: During the COVID-19 crisis several major initiatives have been launched and portals implemented. ECRIN, for example, has established a COVID-19 Taskforce with its national partners, and other initiatives include the COVID-19 Research Project tracker, the RDA COVID-19 working group, the COVID-19 Clinical Research Coalition, and the COVID network meta-analysis. The participation of a selection of these initiatives in the stakeholder forum is necessary to bring in specific COVID-19 related knowledge and should also help to disseminate information about the COVID-19 repository within the research community.

k) European Research Infrastructures: In the European Union, sustainable infrastructures have been successfully implemented to support biomedical and life sciences – with several directly involved in clinical research (e.g. ECRIN, BBMRI, EATRIS, Euro-Bioimaging). It is planned that the repository is operated by ECRIN together with its technical partner (Uio-ELIXIR), but the involvement of other infrastructures may provide valuable additional insights.

3. Legal context of data object sharing in COVID-19 research

The creation of a COVID-19 multinational data repository for research raises a number of legal challenges associated with sharing individual health data and the setting up of a dedicated data repository. While some of these challenges are common to any type of scientific research that uses and stores personal data, others are more closely associated with the data collection and sharing within clinical trials (Ohmann et al., 2017).

For instance, the setting up of a health data repository for research purposes raises specific concerns with GDPR compliance as health data is considered highly sensitive under the GDPR. Moving health data to a data repository raises concerns on security and privacy of information as the data providers no longer have complete control over the security of the information.

The data storage and the data sharing for further research purposes are falling into two distinct legal regimes. For instance, the question of choosing an appropriate legal basis that allows us to transfer such data into a repository is slightly different from the question of the legal basis for secondary use. Each data processing is subject to a separate legal regime and thus we cannot use, for instance the consent collected for secondary use as a legal basis for the transfer of data into a repository for further research purposes (CNIL, Warehouse or research project: a two-step reasoning and two distinct legal regimes, November 28, 2019).

Furthermore, national laws may restrict the transfer of health data to a data repository located in another country or could impose very strict legal requirements on national data providers which makes such a transfer of data almost impossible (Hallinan et al., 2020).

When it comes to clinical trial data sharing and reuse for further research, the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) must also be

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

considered. Even though both legislations support broad consent in the research context, the GDPR provides for alternative legal bases. What is then the most appropriate legal basis for further use? The EDPB as well as the European Commission have tried to provide clarification on the interplay between the two regulations (EDPB, 2019; European Commission, 2019).

Despite such efforts of clarification, it would be difficult to reach a harmonised position at EU level as the GDPR provisions concerning data processing for research purposes are subject to national adaptations (McCullagh, 2019; Jahns et al., 2019).

An additional issue is the lack of a clear and appropriate legal framework for data processing and research in epidemics/pandemics. Currently, neither the GDPR nor the Clinical Trial Directive/Regulation contain specific provisions for research and data processing for research purposes in a pandemic situation. The national authorities have taken emergency measures in order to cope with the coronavirus pandemic. Not all EU Member States have constitutional mechanisms in place allowing for such measures (EU Parliament, 2020).

Also, the solutions proposed vary from country to country and they are valid only for a limited duration (ECRIN, 2020). If the data to be stored in the COVID-19 Repository has been collected under such provisory emergency measures, the rules concerning the use and processing of such data might change in the future. For example, the French data supervisory authority – the CNIL – has decided to extend up to 6 months the time limit for the storage of pseudonymized data collected, during the emergency situation, for the purpose of epidemiological surveillance and research on COVID-19 (CNIL, 2020).

All these considerations should be taken into account when establishing a COVID-19 data repository.

The purpose of this section is to identify and address the specific legal challenges associated with the creation of a COVID-19 multinational repository for research, (practical and technical aspects are considered in following sections) and based on our findings, to provide stakeholders with an overview of the major challenges to be addressed to facilitate and encourage enhanced access and sharing of such data.

After looking into the current state of the EU legislation concerning IPD sharing in general and reviewing the existing recommendations/guidelines issued by the EU authorities during the COVID-19 pandemic (soft law), we will look into the challenges identified by different stakeholders (organizations that have already built up a COVID-19 data repository) and the solutions that have been proposed. Following consultation with various stakeholders, we will develop guidelines/recommendations concerning the COVID-19 repository platform.

3.1. Review of the current state of the EU legislation

In order to share individual-level data from clinical trials, we need to take into account several pieces of the EU legislation, including, but not limited to the GDPR (Council of the EU, 2016), the Clinical Trials Directive (Council of the EU, 2001) and, in the future, the Clinical Trial Regulation (Council of the EU, 2014). Unfortunately, the current EU legislation does not provide a straightforward answer to some of the legal issues that a research organization has to face when sharing individual participant data (IPD) across the EU.

- What is the most appropriate legal basis for collecting and sharing the data (informed consent, legitimate interest, public interest)?

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Who should be the data controller?
- How broad can the purpose of data processing be?
- Could anonymization be an alternative solution that would allow us to overcome the legal constraints?

In addition, the EU legislation left room for national adaptations. For instance, the GDPR was intended to provide much greater harmonization than at present in respect to data protection rules across the EU. When it comes to data processing activities for research purposes a lot of national variations remain.

For instance, concerning the question of the most appropriate legal basis for data sharing for research purposes, there is a consensus among several research organizations that “broad consent” would be more suitable in this context. Despite that, the national lawmakers have not always followed this position. In many EU countries, the “broad consent“ has not been recognized as a valid legal basis (Kaye et al., 2016).

Another legal issue is related to the interplay between the GDPR and the Clinical Trial Regulation. For instance, both legislations allow for “broad consent”.

The GDPR, for instance, in Article 25A, reads:

"*It is often not possible to fully identify the purpose of data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose*".

The CTR also contains specific provisions concerning the secondary use of clinical trials data and allows for broad consent. While the GDPR does not define the broad consent as blanket consent (the consent should cover “certain areas of research”), the CTR doesn’t say how “broad” this consent might be:

"It is appropriate that universities and other research institutions, under certain circumstances that are in accordance with the applicable law on data protection, be able to collect data from clinical trials to be used for future scientific research, for example for medical, natural or social sciences research purposes. In order to collect data for such purposes it is necessary that the subject gives consent to use his or her data outside the protocol of the clinical trial and has the right to withdraw that consent at any time". (CTR, Recital 29)

The two legislations are interconnected. In practice though, the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) raises several questions, including: “What is the meaning of the broad consent under the CTR?” and “What are the implications for the use of personal data outside the protocol of the clinical trial (secondary use) within the scope of the GDPR?” (European Commission, 2019) “Which legislation should prevail when sharing data collected in the course of a clinical trial?”.

Considering the lack of clarity of the current legal framework, guidelines and recommendations have been issued to provide some clarification on the application of the abovementioned legislations.

For instance, the European Data Protection Board (EDPB) has published an opinion on the interplay between the CTR and the GDPR. It considers that the GDPR should prevail.

This project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 824087.

With respect to the legal basis for secondary use, the EDPB doesn't seem to support the notion of "broad consent". The Committee refers instead to the presumption of compatibility, subject to the Article 89 of the GDPR. It states:

"For the time being, the presumption of compatibility, subject to the conditions set forth in Article 89,[the GDPR] should not be excluded, in all circumstances, for the secondary use of clinical trial data outside the clinical trial protocol for other scientific purposes. In any event, even when the presumption of compatibility will find to apply, the scientific research making use of the data outside the protocol of the clinical trial must be conducted in compliance with all other relevant applicable provisions of data protection as stated under Article 28(2) CTR. (EDPB 2019, paragraphs 31 and 32).

Despite the clarification provided by EDPB (their opinion is not legally binding), there are still some issues around the interplay between the two legislations:

- The CTR is not applicable yet.
- Should one refer to the GDPR or to the CTR in order to determine the "broadness" of the purpose of a broad consent in the context of a clinical trial? Which regulation should prevail as both regulations contain provisions concerning broad consent?
- How can subjects exercise their rights effectively in case of withdrawal of broad consent? What are the risks of complaint and litigation?
- On which legal basis should the data be further processed (secondary use) in case a broad consent was not used from the start?
- Is broad consent the primary option?

The position of the national supervisory authorities varies from country to country. Only a few of them have supported broad consent (Germany, in France only partially). An explicit consent is still required for the processing of data for the purpose of scientific research. Data that do not carry an explicit consent to data sharing (as from many past and current trials) could still be shared in circumstances where national or other regulations allow for exceptions to the normal restrictions on data sharing, for instance where obtaining consent is seen as too impractical for researchers or too burdensome for participants, and the risks are assessed as low. In such circumstances, it is anticipated that the proposed sharing request and data use may need the involvement of ethical committees or other review boards, dependent on national systems. In addition, the data may be required to undergo an increased level of de-identification, and the Data Use Agreement may impose greater restrictions on data access.

Effective anonymisation may also be an option, although there has to be a mechanism to agree that anonymisation has been truly achieved. It should be noted, however, that various jurisdictions define the threshold for anonymity differently (for example, the USA). Assessment of all the means reasonably likely to be used must consider not only the data on its own, but also the possibility of combination with other accessible data, including by third parties (RDA, 2020). The consequence of rendering data anonymous will often be that certain ethical and legal obligations which usually apply to identifiable data will no longer apply. In particular, anonymisation will usually render data protection law inapplicable; in the EU, for example, anonymous data falls outside the scope of the GDPR. With large datasets, and especially where datasets are cross-correlated, absolute anonymity will often be very hard to achieve. Researchers may need to take into account the possibility of future re-identification and manage this risk by means of a risk assessment. If that is the case the data protection regulations no longer apply. Anonymising data will itself usually be seen as data processing, and thus covered by data protection regulations. The anonymisation would, therefore, have to be done by someone who had been authorised to process the data.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

When developing a COVID-19 platform we will have to take into account all the above-mentioned issues concerning data storage and sharing in general.

In addition, we will have to take into account the current state of the EU legislation concerning data processing/data repository in research in a pandemic situation and the response of the EU authorities during the COVID-19 pandemic.

After looking into the current state of legislation, we have analysed the response provided by the authorities.

Often in cases of health emergencies such as the COVID-19 pandemic, fast track procedures are put in place, allowing the approval processes to be accelerated without diminishing the protection of the rights of persons (RDA, 2020). It has to be clarified whether due to the pandemic, specific rules have been issued in a country; giving easy access to such information and the conditions of their application are highly necessary. If this is the case, it has to be taken into consideration.

a. The case of research and data processing rules in pandemics: a certain legal unclarity at EU level

In order to effectively respond to the challenges raised by research conducted in a pandemic situation, we need to have an adapted and clear legal framework that, on one hand, protects the rights of research participants that are in a vulnerable situation (they are not necessarily in emergency situation) and on the other hand, allows the research organizations to rapidly set up multinational clinical studies and to share and/or transfer the data internationally.

If we look at the main EU legal instruments (hard law) that we apply to our health research activities, which are the GDPR and the Clinical Trial Directive/Regulation, neither of them contain any specific provisions for research in pandemics. The EU did adopt a regulation in July 2020 that addresses some of the concerns raised in the current pandemic. However, the scope of this regulation is limited to the conduct of clinical trials involving medicinal products for human use 'containing or consisting of genetically modified organisms intended to treat or prevent coronavirus disease' (Council of the EU, 2020).

In order to answer the challenges raised by research in the COVID-19 outbreak, simplification measures have been proposed at EU level through the soft law. These provisions, however, are valid only during the period of the COVID-19 outbreak in the EU/EEA (European Medicines Agency, 2020).

As for the GDPR, even though it allows data to be used in the event of epidemics/pandemics, this legislation includes no specific rules that aim at facilitating the sharing of data and consequently the setting up of multinational repositories in a pandemic situation. Furthermore, the GDPR fails to bring about sufficient harmonization of the national requirements across the EU.

Moreover, the interplay between the two legislations is already a major source of concern for the stakeholders that are sharing data internationally for research purposes.

b. The "Soft law" responses to COVID-19 emergencies: the EU position and reactions at national level (Library of Congress, 2020)

The EU has limited competency in public health (Brehon 2020) and its reaction to COVID-19 consists of a large number of "soft law" measures (decisions, recommendations, etc.). These measures, even though useful, have been adopted during and only for the period of the COVID-19 outbreak. Furthermore, these

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

measures respond only partially to the challenges that we are facing when setting up a multinational COVID-19 data repository. An analysis of each measure will be provided below.

European Data Protection Board, Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak.

The EDPB released guidelines on geolocation and other tracking mechanisms to combat COVID-19, relevant for studies investigating these kinds of interventions.[3]

European Data Protection Board, Guidelines on processing of health data in the context of COVID-19 (EDPB, 2020).

The EDPB addresses questions concerning international data transfers involving health data for research purposes related to the fight against COVID-19.

Decision No 1082/2013/EU of the European Parliament and of the Council of 22 October 2013 on serious cross-border threats to health and repealing Decision No 2119/98/EC Text with EEA relevance (Council of the EU, 2013).

At European regional level, EU Decision 1082/13 is the key legal instrument for cross-border threats to health. This explicitly recognises and endorses compliance with the (IHR) at Articles (6), (12) and (26) (see above). Both the IHR and Decision 1082/13 require signatory states to develop national plans for pandemic preparedness and response.

Joint European Roadmap towards lifting COVID-19 containment measures (EU exit roadmap, European Commission, 2020d)

To quote from Alemanno, 2020:

"The EU Exit Roadmap offers three main criteria to assess whether the time has come to begin to relax the confinement for each and every Member State

- An epidemiological criterion showing that the spread of the disease has significantly decreased for a sustained period of time;
- Sufficient health system capacity (i.e. the extent to which the different healthcare systems can cope with future increases in infection rates after lifting of the measures);
- Appropriate monitoring capacity, including large-scale testing capacity to detect and monitor the spread of the virus combined with contact tracing and quarantine capacity in case of the reappearance and further spread of infections.

This rather unusual guidance document strikes a fine balance between the need for EU-wide coordination and Member States' different country-specific needs and cost– benefit calculus. It essentially introduces a set of meta-criteria or benchmarks framing the exercise of Member States' public health prerogatives. In doing so, it also leaves each Member State the choice, depending on their size and organisation, regarding 'what level of compliance with the criteria above should be assessed' (e.g. the regional or macro-regional level rather than at the national level)".

COMMISSION RECOMMENDATION (EU) 2020/518 of 8 April 2020 on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data (European Commission, 2020a).

[3] https://iapp.org/media/pdf/resource_center/edpb_geolocation_covid19_guidance.pdf

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL AND THE EUROPEAN INVESTMENT BANK EU Strategy for COVID 19 vaccines (European Commission, 2020b)

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS Short-term EU health preparedness for COVID 19 outbreaks (European Commission, 2020c)

- At the national level, similar measures have been adopted by the different national authorities, who, for the most part, have transposed the measures adopted at the EU level.
- ECRIN has established a COVID-19 Taskforce with its national partners to look into the measures adopted by the national authorities during the COVID-19 pandemic. In this context, ECRIN has developed a database of the regulatory, ethical and data protection fast track approvals across all European countries (ECRIN 2020).

Other international instruments

In addition, there are several international instruments available, which tackle the management of emergency and health crises and which should be put into context with the EU laws and regulations.

WHO, 2005

WHO International Health Regulations (2005) (IHR) sets out key principles to guide national preparedness and response to pandemics.

(See WHO 2005, Speakman 2017)

Oviedo Convention, 1997

The Convention on Human Rights and Biomedicine (Oviedo Convention) is the only legally binding instrument at international level addressing human rights in the field of biomedicine. It provides a unique human rights framework, including in a context of emergency and health crisis management, to guide decisions and practices both in clinical and research fields.

Article 8 – Emergency situation. When because of an emergency situation the appropriate consent cannot be obtained, any medically necessary intervention may be carried out immediately for the benefit of the health of the individual concerned.

Article 16,17 – consent to research

(See Council of Europe 1997; Andorno, 2005)

Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research (Council of Europe, 2005)

Article 19 – Research on persons in emergency clinical situations.

This is of relevance to clinical research.

The Council of Europe statement on bioethics during COVID 19: (Council of Europe 2020)

In this Statement, the Council of Europe Committee on Bioethics (DH-BIO) wishes to highlight some of the human rights principles laid down in the Oviedo Convention which are particularly relevant and require particular vigilance in their application in the current pandemic.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Possibilities to make restrictions on the exercise of the rights and protective provisions contained in the Oviedo Convention are discussed with relevance to COVID 19. The document is relevant for clinical research.

3.2. Legal challenges associated with the COVID-19 repository

All the above legal instruments respond only partially to our concerns. In order to clarify the situation and come to a practical solution for the planned COVID-19 repository, we decided to include repositories, which are already dealing with data from COVID-19 trials in the stakeholder meetings. There are only a few so far but their experience is of utmost importance (e.g. IDDO/SARIC, Vivli). In addition, the stakeholder forum will cover more ethical and legal experts. The topics to be discussed with repositories, COVID-19 sponsors and coordinating investigators of clinical trials and the ethical/legal experts will cover, among others, the following topics:

- Legal basis for COVID-19 data storage in a data repository/ transfer/and further processing for research purposes
- Broad consent as legal basis for data sharing: national acceptability
- Transfer of personal health data to a third country (Norway): under which conditions and legal requirements?
- Health data repositories: specific legal requirements? specific legal regime?
- The processing of health data collected in the context of a research on COVID-19: the same rules during and after the emergency situation?
- The Agreements required for the transfer and the sharing of data (Consortium agreement, Data Transfer Agreement, Data Access Agreements, etc.)
- Data policy for the data repository to ensure compliance with the GDPR

3.3. Guidelines/recommendations for the COVID-19 platform

As a result of the analysis performed and the discussion at the stakeholder meetings, guidelines/recommendations concerning the COVID-19 platform will be formulated. These guidelines/recommendations will drive the procedural and technical specification of the COVID-19 repository.

Under the current situation the following pragmatic solutions concerning the legal basis for data sharing are under discussion:

- Limit the broadness of consent to “certain areas of scientific research" as provided under the GDPR
- Look for alternative legal basis such as public interest
- EDPB’s suggested solution: the presumption of compatibility provided under Article 5(1)(b) GDPR.
- Apply (effective) anonymization in all other cases if truly possible and legally compliant

4. Functional specification of the proposed repository

This section provides a high-level description of how we envisage the repository will function. The first part deals with the core task of managing data objects, while the second covers possible support

This project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 824087.

services. The focus is on the initial version of the repository, but a further section describes possible enhancements during later phases of development. Finally, there is a summary of the technical infrastructure to be used, at TSD in Oslo.

4.1. Core Functionality – Data Object Management

a) Core purpose and content: The repository will be designed to store data objects from completed clinical research projects related to COVID-19 (both clinical trials and non-interventional studies), for possible secondary use and analysis. Here 'data object' is used as the generic term referring to any object in electronic form, including documents as well as datasets. The belief is that it is impossible to fully understand or appraise data without the supporting documents (e.g. protocols, analysis plans, related journal papers) so that the data must form part of a complete 'package' of material from each study.

b) The data objects to be stored: We believe that in most cases the data to be stored will be pseudonymised and originate in the EU/EEC, and so fall under the GDPR, as well as, in many cases, the Clinical Trials Directive/Regulation. Derogations exist for member states in relation to some aspects of managing personal research data, and interpretations of GDPR and related legislation appear to be still evolving. Exactly what data can be stored will therefore depend upon the applicable national legal framework(s), but also on the data protection policy adopted by the data provider at institutional level. If data is accepted as being anonymised, the GDPR no longer applies and data management should be more straightforward. For data objects that are not personal data, e.g. documents, metadata files, transfer to a repository would be governed by the data provider's policies.

c) Division of legal responsibilities: Legally the repository will act as a data processor, with the data generators remaining as the data controllers. If and only if discussions indicate that stakeholders want the repository to become a data controller, at least in some cases, will this option be considered.

d) Linking or pseudonymising data: For the avoidance of doubt, the repository will not store any data that could link pseudonymised data back to the individual study participants or make any attempt to access such data. In other words, as far as the repository (and secondary users) are concerned, the data, once de-identified, is in practical terms anonymised (whatever its exact legal status). Maintaining the pseudonymising or linking data, e.g. to meet data retention requirements, will remain the responsibility of the data controllers, independent of any interaction with the data repository.

e) Decisions on transfer: The final decision about making data available for re-use and transferred to the repository will be taken by the data controller (e.g. sponsor). To protect study participants, however, as well as its own reputation, the repository will reserve the right to challenge a data controller's assertion that data is anonymised or fully de-identified, if that does not appear to be the case. Transfer will be dependent on prior Data Transfer Agreements (see f).

f) Data Transfer Agreements: Roles and responsibilities with respect to data processing activities will be defined in specific data transfer agreements (DTAs) between the data provider (data controller) and the data repository, or the organisations running the repository. (Whether or not the repository itself needs to become a legal entity needs further exploration). Such agreements should include the explicit allocation of controller/processor roles, the de-identification and pseudonymisation status of the material, the preferred access method, and the length of storage required, amongst other things. All material transferred to the repository, including anonymised

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

data and documents, will require such data transfer agreements. Each agreement will cover one or more 'object packages', each associated with a specific study.

g) Format of material: Material, whether data or documents, should be prepared in non-proprietary formats, e.g. CSV, TSV (or similar), or an XML schema (with schema definition) for data, plain text for scripts and code, and PDFs for almost anything else. Non-proprietary files can be uploaded (e.g. specific to a particular statistics program) but must be accompanied by a non-proprietary version. There is no size limit on file size.

h) Metadata required: Two forms of metadata are required.
   1. For data, descriptive metadata must be provided, usually as an additional file or set of files within the object package. The format can be CSV or XML, with CDISC's Define.xml being the preferred format. Data item information (code, name, type, categories used, definition if ambiguous) and data schedule details must be included in this metadata.
   2. For each object package, provenance and discoverability of data must be provided covering both the source study and the related data objects (of all kinds). This will make use of the ECRIN metadata schema designed specifically for this purpose. Tools will need to be developed to support the creation of this metadata.

i) De-identification required: Whether data is classed as pseudonymised or anonymised it must be de-identified, by the data controllers, so that recipients cannot in practice re-identify individuals, even with additional information taken from elsewhere. The normal expectation will be that:
   1. The direct identifiers listed on the HHS HIPAA site (HHS.Gov, 2020) are removed.
   2. Any dates are changed into integers (number of days post or pre a fixed event such as screening or randomisation) or are reset so that all participants appear to start at the same time.
   3. Large free text fields, used for comments, explanations or narratives, are removed.

   Further de-identification may be required. Data-identification should always be documented and that description should form part of the object package. In addition, the de-identified data should be re-analysed against primary and secondary outcome measures, and both similarities and differences in outcome should be documented. Note that protection of participant privacy stems from both de-identification and the use of Data Use Agreements (see below).

j) Application of Persistent Identifiers (PIDs): The intention is to ensure that all data objects in the repository are allocated DOIs (unless they already have one). These will be applied by the repository. DOI prefixes should be structured so that they indicate that objects have been generated by the same study. DOI suffixes should indicate different versions of the same material, when they exist. If allocation of DOIs proves to be uneconomic an alternative PID system will need to be developed.

k) Uploading data objects to the repository: This will be a multi-stage process:
   1. The data controllers agree a data transfer agreement with the repository, covering one or more object sets or packages, i.e. one or more studies.
   2. Data controllers are given permissions and instructions to upload their data objects (see technical specification for details). The data objects are stored in the repository but at this stage are not available to others or 'advertised' in the repository's catalogue.
   3. Repository managers check that descriptive metadata is present and sufficient, and that the data does appear to be de-identified.
   4. Data controllers complete provenance/discovery metadata using the ECRIN schema and repository provided tools.
   5. DOIs or other PIDs are allocated to all data objects that do not already possess them.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

6. The discovery metadata is incorporated into the repository's catalogue, and also integrated with the ECRIN Metadata repository.
7. If the study is new to the system a new 'landing page' will be constructed for that study, giving both the study details and those of the associated data objects.
8. The data objects are then available for download (e.g. in the case of protocols or other documents) or for application for access (in the case of pseudonymised datasets).

l) Updating material: If new versions of material are available data controllers should be able to upload that directly and a new DOI will be applied. Provenance metadata will need to be updated. All versions of documents and datasets will be retained but, by default, only the most recent version of a document/dataset will be available. Special requests would be necessary to access older versions. Completely new material, added to an existing study, will require associated provenance and discoverability metadata. In both cases a streamlined form of the processes described in k would be employed.

m) Length of storage: Data controllers will be able to specify the length of storage required. The default, however, will be up to 2050, which fits with the existing guarantee from TSD of keeping their material until 2050.

n) Types of access available: Data controllers will select the access they require for any data object, but the options available will depend on the object type.

1. Completely free download – objects are publicly available and can be downloaded directly from a link on the study landing page. Only available for documents and files without personal data (e.g. metadata files, aggregate result files). Data controllers may attach an embargo period on material, which will be shown on screen.
2. (If there is interest from stakeholders) 'Download with identification', when the user identifies themselves, so the controller knows who is downloading their material. Available as an option for documents and files without personal data, and for anonymised personal data sets. Identification might be from a registration mechanism (which would then require a log-in to access material protected this way) or from a token-based system that lasted for one browser session (after completing a form online) but which did not involve long term user management. Data controllers may also attach an embargo period, which will be shown on screen.
3. Download after review – available for all data object types but compulsory for pseudonymised datasets. The user is obliged to complete an online form explaining who they are, why they want the data, and agreeing that they will sign a Data Use Agreement - see below. Data controllers may optionally insist that they attach a protocol describing the work they wish to carry out and may further optionally insist that the protocol has ethics approval. This request will be reviewed by the repository staff to ensure it is complete, before being passed to the Data Controllers for their decision. If successful, users will be given an identity on the system and can login and retrieve the material.

In practice it is hoped that most documents and non-personal data will be made publicly available, albeit after a possible embargo period. Access to pseudonymised data sets will require review on a case by case basis.

o) Data Use Agreements: To access pseudonymised datasets, after successful review of the request, it will be necessary to complete a Data Use Agreement (DUA). This should demand, for example, that data recipients do not attempt to identify study participants, that they use the data only for the agreed purpose, that they do not pass it on to any other third party, that they store the data securely, and that they will acknowledge the original data generators in any published work. A

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

template (or templates) for data use agreements will be provided. The DUA is not foreseen as onerous and should not take too long to complete, but it is seen as a necessary mechanism for mitigating risk.

p) Data Access Committee: If there is interest amongst the stakeholder community, it may be possible to establish a Data Access Committee (DAC) of experienced trialists who could consider requests for access on behalf of data controllers, and make recommendations to them about the appropriateness of requests.

q) Monitoring and communication: Data deposition should be monitored by the storage of the relevant agreements (DTAs). Data requests will be monitored according to the mode of access - simple downloads can be counted, downloads after identification or review will include information about the destination of material and its intended use. This material will be passed to the relevant data controllers, but in summary form to all the main stakeholders, to EOSC and through periodic publication to the wider research community.

r) Long term arrangements: Arrangements will be sought to transfer all data securely to another repository, at least for static storage, if for any reason the repository facility has to close (e.g. through withdrawal of funding).

4.2. Supporting Functionality and Services

Preparing and managing data for secondary re-use is a relatively new activity for most trialists and it is envisaged that in many cases sponsors and investigators will need support. The following are possible services that could be developed and made available by the repository, at cost, if the stakeholder group thought that they would be a useful addition to the system.

a) Resource collections relating to legal issues and secondary use - hard and soft law, opinions pieces etc.

b) Consultancy/support for de-identification and its documentation.

c) Consultancy/support (including recreation of datasets) to generate descriptive metadata in standardised formats - in particular using Define.xml.

d) Consultancy/support (including recreation of datasets) using data standards to make data more inter-operable (e.g. using the CDISC COVID-19 data structures). If done retrospectively this would likely be relatively expensive, but such support could also be made available as part of study planning.

e) Support in applying provenance/discovery metadata to data objects. (N.B. Tools are planned to make this application much easier).

f) Monitoring of publications derived from secondary use of data objects, to ensure original data generators have been properly acknowledged.

In addition, the intention is to provide template Data Transfer Agreements and template Data Use Agreements, to speed the application of these agreements to data object packages, as well as comprehensive on-screen guidance to data generators and data requestors.

4.3. Later Enhancements

The functionality described above relates to the initial version of the repository. The list below summarises some possible enhancements in the latter stages of development.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

a) Offering on-screen access as an additional access type. The access described in the core functionality section all involves file download. If it seems an attractive option to stakeholders, providing on-screen access, with no download possible but with statistical programs available to carry out analysis and generate aggregate figures, can be investigated and, if possible, implemented.
b) Storing data and data objects for review purposes: For papers undergoing peer review, the repository could allow access to the data and other data objects to the reviewers, if this service was seen as potentially useful by publishers. It is not clear if this would fall into primary or secondary use, or whether the data would need de-identification.
c) Storing data and data objects related to pre-prints: For papers published publicly as pre-prints (and in such journals as F1000) the repository could allow access to the data and other data objects to those who wished to comment on the pre-print, on request, if this service was seen as potentially useful by pre-print publishers.
d) Storing data for ongoing studies: The TSD facilities in Oslo already provide a service for the storage of data from ongoing studies. Some COVID-19 researchers may want to use a similar service. This may be especially useful for platform trials or where data aggregation from different studies has been planned.
e) As an extension to d, the repository could also be used to provide a unified data platform for 'core' COVID data, assembling data from various studies into a single, searchable data platform. Such a service would need to be complementary, however, to the similar service planned by IDDO / ISARIC and not duplicate or compete with that service.
f) Storing data for non COVID-19 research. The most obvious extension of the repository would be to extend it to non COVID research, to create a repository for clinical research data in general that was particularly adapted to the European legal landscape. This would depend both on successful implementation of the repository as a COVID only system, and the development of a business model that would guarantee sustainability.

4.4. The Underlying Infrastructure

The underlying infrastructure is the TSD (Tjenester for Sensitive Data) system of the University of Oslo, here provided through Elixir-NO. The infrastructure is a multi-tenant remote access system with a strong set of built-in security measures: optional two factor login, controlled data access, strict access control and separation between tenants. It has its own IDP (Identity Provider) and IGA (Identity Governance and Administration system) to provide access to any user from anywhere. It supports a wide range of services as a 'Platform as a Service' or PaaS, including data collection, data storage and backend support for applications.

The system currently hosts more than 5000 users and approximately 1000 research projects. It includes over a 1000 windows and linux virtual machines, a 2400 CPU High Performance Computing system, and 5 PiBs (easily expanded) of IBM ESS (Elastic Storage Server) storage. It is well established, penetration tested and assessed and has been cleared for clinical use by the medical genetics departments at University Hospital Oslo and St. Olav University Hospital Trondheim.[4,5]

[4] https://www.uio.no/english/services/it/research/sensitive-data/about/index.html for background information
[5] https://www.uio.no/english/services/it/research/sensitive-data/about/description-of-the-system.html

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

5. Quality Assurance

5.1. Quality criteria and compliance

There are overarching general principles that address aspects of data management and data repositories at a very high level. The FAIR (Wilkinson et al., 2016) principles state that data should be Findable, Accessible, Interoperable and Reusable. The TRUST principles (Lin et al., 2020) represent guidance for repositories of research data, with a focus on Transparency, Responsibility, User focus, Sustainability and Technology. The COVID-19 repository to be developed should follow and implement these high-level principles.

Specifically, to meet each of the FAIR requirements for data:

Findable – We plan to make data objects findable by applying the ECRIN provenance / discoverability metadata, and then making that metadata searchable both within the repository and the MDR.

Accessible – We plan to make data objects accessible by explicitly giving clear instructions for accessing all data objects, having clarified the access arrangements required with data controllers, and by applying PIDs to all data objects. Download to be made straightforward for objects that will be publicly available.

Interoperable – We plan to improve interoperability by insisting on descriptive metadata and encouraging people to use a single format for that metadata (Define.xml) – though interoperability is ultimately the responsibility of the data providers.

Re-usable – We plan to make the data objects more re-usable – again by insisting on both descriptive and provenance metadata, and by stating explicit and clear access criteria.

Whilst to meet each of the TRUST requirements for a data repository:

Transparency – All data objects will be described in public catalogues, the procedures of the repository will be public, along with data relating to the level of activity and the results of individual requests for data.

Responsibility – The repository will check that the data objects in it are accurately described, maintained securely, are de-identified when necessary to protect research participants, and include the necessary metadata to be readily understood by others.

User focus – Through continuing engagement with stakeholders and users, and ongoing self and external assessment, we will aim to ensure that the expectations of all stakeholders are met.

Sustainability – Although this repository is built as part of EOSC-Life and that project is relatively short term, we will be seeking sustainability in the longer term once we have demonstrated the utility of the repository.

Technology – The infrastructure and technology of the University of Oslo's TSD service has already shown that it can support secure, persistent, and reliable services.

The services any repository provides should also conform to more specific quality standards, to give its users confidence that their data and documents will be stored securely and in accordance with the data transfer agreements they have agreed (Ohmann et al., 2017). Different approaches have been used to assess the quality of repositories dedicated to data sharing, each with an emphasis on different features. Examples include the Core Trustworthy Data Repositories Requirements (CoreTrustSeal, 2020) the

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

criteria listed by Burton et al (2015) for Data Safe Havens, and the recommendations listed by Hrynaszkiewicz et al., (2016).

In the H2020 CORBEL project, 8 quality criteria were defined that were seen as particularly important for repositories for clinical research data, using previous work assessing clinical data repositories (Banzi, 2019) as a guide. The implementation of a DSpace demonstrator[6] repository was then assessed with respect to these criteria (Tilki, 2020). These criteria will also be applied to the development of the COVID-19 repository, along with the more general criteria within the CoreTrustSeal system (although there is an overlap between them). We intend to explore the costs and benefits of seeking formal CoreTrustSeal accreditation for the repository. The aspects of repository functioning covered by these sets of criteria are listed below:

CoreTrustSeal criteria – Aspects covered:

Organizational Infrastructure:

1. Mission/Scope
2. Licenses
3. Continuity of access
4. Confidentiality/Ethics
5. Organizational infrastructure
6. Expert guidance

Digital Object Management

7. Data integrity and authenticity
8. Appraisal
9. Documented storage procedures
10. Preservation plan
11. Data quality
12. Workflows
13. Data discovery and identification
14. Data reuse

Technology

15. Technical infrastructure
16. Security

ECRIN Clinical Research Repository Criteria – Aspects covered:

- Guidelines for data upload and storage
- Support for data de-identification
- Data quality controls
- Contracts for upload and storage
- Exposure of metadata
- Application of identifiers

[6] Demonstrator is a prototype, a rough example or a first version of a conceivable repository that serves as proof of concept for showcasing the possible applications, feasibility, performance and usability.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Flexibility of access
- Plans for long-term preservation

It should be noted that the TSD infrastructure is currently seeking ISO 27001 certification, to help ensure that the high levels of security are maintained. ECRIN is currently seeking ISO 9001 certification, to recognise the extensive work that has been done in developing quality systems throughout the organisation.

Sustainability aspects are discussed in Chapter 7 and a detailed sustainability plan will be explored later on together with stakeholders and delivered in M37 as « D14.3 Report about use and user satisfaction of COVID-19 repository including a maintenance and sustainability plan ».

5.2. Ethical requirements and compliance

We are also conscious of the ethical dimension of developing and operating a data repository. Data sharing can raise new questions about familiar ethical concepts, such as privacy, confidentiality and informed consent, as well as generate novel ones (Kalkman et al., 2019). Amongst the ethical issues that have been identified (O'Connell and Plewes, 2015) are:

- the sharing of information that allows people to be identified
- the validity of consent when future uses of data are unclear
- a possible increase in social injustice (e.g. through stigma, discrimination)
- the impact on public trust in research if data are used inappropriately
- questionable or non-transparent decisions about who gets the data

A recent systematic review examined the ethical principles and norms formulated by international groups and organisations with respect to responsible data sharing in health research (Kalkman et al., 2019). Four main themes were identified:

- social benefits and value
- distribution of risks, benefits and burdens
- respect for individual and groups
- public trust and engagement

Some of the specific requirements that were listed in the same review, that have not already been explicitly listed in the functional specification, are listed below:

1. Ongoing discussion with major stakeholders, to be aware of and balance the needs of data generators, secondary users and the wider community.
2. Ensuring the repository operates within an explicit public ethics and governance framework.
3. Allowing requests to be submitted by qualified researchers but also by citizen scientists, journalists and others, who are able to justify their request and attest the use of rigorous scientific methods.
4. Ensuring equitable access to data by transparent rules, with conditions and procedures harmonised across all stakeholders.
5. Ensuring the purposes for which data is shared are consistent with the legal bases used for that sharing, and the spirit in which consent was given.
6. Maximising accessibility by avoiding fees (or applying low cost fees) for both data deposition and data access.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

7. Providing transparency about all procedures, workflows, and usage, including through comprehensive public documentation.
8. Cataloguing data objects in the repository in a consistent manner, using standardised metadata, ensuring that metadata is verifiable, accurate, and unbiased.
9. Ensuring that due credit and acknowledgement is given to all who contributed to the results, in particular the original data generators.
10. Including clear and easy-to-use processes to allow study participants to remove barriers for participants to withdraw their consent for data use, where the relevant legislative framework(s) allow this.
11. Ensuring an equitable policy with respect to duration of storage and disposal and destruction of data.
12. Ensuring appropriate data control, compliance with quality standards and feedback mechanisms at every stage of data processing.
13. Assuring long-term accessibility and sustainability of the repository.

To try and ensure that the repository meets these and other ethical requirements, the intention is to:

- Involve internationally acknowledged ethical experts with major experience in data sharing activities in the stakeholder group.
- Continue the participation of representatives from all involved stakeholder groups in all phases of the project
- Develop and publish an ethical framework for the repository and then carry out self-assessment against it.

6. Implementation plan for the repository

6.1. Physical Infrastructure and TSD services

The TSD physical infrastructure is a complex setup with main components consisting of approximately 1500 virtual computers, a 3000 cores HPC system, 2 factor remote login system based on VMware Horizon and Thinlinc. Additionally, TSD has a self services online enquiry formular system, and the possibility to act as a secure backend for smartphone and tablet apps. The whole system is secured by many layers from special routers, firewalls, separate physical and virtual infrastructures and networks. All access is by 2-factor and the rule is to never get more access than strictly needed. This implies also that the more than 1000 research projects hosted by TSD are separated by VLANs and NFS4 ACLs and storage is only accessed through Kerberos tickets in combination with VLAN separation. (System white paper can be handed out on request).

The overall goal is to launch a minimum viable product (MVP) by the end of January 2021. The progress plan for the COVID-19 clinical trial data repository is as follows:

1. Establish the ECRIN/CORBEL metadata schema system on the TSD online enquiry form. Already started as a proof of concept and found to be the way ahead. Along with fine tuning the schema TSD will also develop the lacking functionality that today makes a small mismatch between the ECRIN/CORBEL metadata schema and the TSD functionality. This development also needs to take

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

into consideration how to couple the metadata (trial project and data-objects) with the actual files. This will be finished by mid-November 2020.
2. TSD will establish a legal-data package for all data-uploaders in agreement with ECRIN-lawyers/legal experts. This package, once completed will enable the data-uploaders with the right to upload. This will be finished once the legal package is ready. Estimated work on the TSD side is approximately 1 week.
3. Data-deposit functionality in TSD will be reused, but must be extended to enable DOI/PID assignment. This will be enabled by December 2020.
4. TSD will establish a data-viewer functionality for quality assessment of data vs metadata. Parts of this quality-control will be automated. Finished by 15th of November 2020.
5. TSD will establish a landing page-system for all metadata on the project level, this landing page will enable users to drill down into metadata per data object per clinical trial. TSD will utilize the established UiO CMS (Vortex) and will have this finished by mid-December 2020. ECRIN will guide development to which metadata-fields that are relevant for data discovery.
6. TSD will make a "download-data legal package" in agreement with ECRIN lawyers and legal experts. By fulfilling the legal demands (that vary per dataset) TSD will enable datasets for further analysis inside TSD or for download to the data requestor. Finalized by mid-November 2020.
7. Testing, fine-tuning and further development will continue to a launch of an MVP in the end of January 2021.
8. Further development plans (de-intentifications etc.) will be decided upon within the consortium once the MVP is in progress.

6.2. Procedures and Data Management Systems

There are a series of organisational, procedural, and workflow issues that will need to be tackled within the implementation plan. These developments will involve all partners, and usually require both procedural and infrastructure components. In general, initial drafts of documents related to the issues below should be available by the first Stakeholders meeting, with working first versions available to support the pilot repository in the spring of 2021. Further development of these systems would then need to proceed throughout the length of the project, with final versions available at project end.

- Clarification of legal entities involved in the repository: A legal entity will need to be involved in any formal Data Transfer or Data Use Agreement. We need to clarify if that entity should be the University of Oslo, ECRIN, some combination of those organisations, the repository established as a separate legal entity in its own right, or some other legal structure. The various possibilities need to be explored as one of the preliminary steps in developing the repository.
- Clarification of insurance responsibilities and costs: The risks associated with the storage and re-use of data need to be identified, together with an exploration of how they can be mitigated (including disclaimers within data transfer and data use agreements). Any remaining risks should be identified so that the needs for liability and other insurance are known and any related costs can be estimated.
- Establishment of Quality Management System: As determined by the Steering group, the repository will need to establish a mechanism for developing, reviewing and approving quality systems and documents, including the SOPs and other documents listed below. Such a quality management system will be necessary to ensure consistent workflows, to provide data and quality assurance to stakeholders, and to demonstrate compliance with CoreTrustSeal and other external quality criteria (see section 5)

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Draft SOP on Data Ingestion: The stages involved in successful data ingestion need to be described and there should be clarification, at each stage, of:
- The responsibilities of each organisation/role within the repository,
- the information and objects required from data generators,
- the role of the Data Transfer Agreement
- the quality criteria that need to be checked and recorded
- the administrative information that needs to be recorded

More detailed technical work instructions are likely to be required for operationalising the ingestion process, as well as a variety of forms/record structures for collecting and storing the associated data.

- Development of Data Transfer Agreement templates: The purposes, sections and components of Data Transfer Agreements between data generators and the repository will need to be agreed and templates constructed that are legally robust. An important issue will be the degree of flexibility to be allowed within such agreements.
- Implementation of DOI minting: The plan is to provide DOIs (or a similar persistent identifier) for data objects that do not already have one. This process will need to be integrated with data ingestion quality management. It is likely to involve external organisations (e.g. DataCite) and distinct costs and may therefore need additional workflow support with suitable SOPs and administrative systems.
- Development of Guidance documents for data generators: A variety of guidance documents, available as web pages and/or downloadable PDFs, will be necessary to inform data generators about the policies and preferences of the repository with regard to data upload, how they can interact with the repository during ingestion, the purpose and nature of the Data Transfer Agreement, the quality assurance checks that will be required of the data 'package' and the operational details of data upload.
- Development of guidance for data requesters: Similarly, a variety of guidance documents will be necessary for potential data requesters, detailing the request process, indicating how study or data object specific details can be found, and indicating what is required from them to initiate the request.
- Draft SOP on Dealing with Data Requests: The stages involved in handling data requests need to be described, and with each:
  - the responsibilities of each organisation/role within the repository,
  - the information and objects required from data requesters,
  - the role of the Data Use Agreement
  - the quality criteria that need to be checked and recorded
  - the administrative information that needs to be recorded

More detailed technical work instructions are likely to be required for operationalising the request process, as well as a variety of forms/record structures for collecting and storing the associated data.

a) Development of Data Use Agreement templates: The purposes and contents of Data Use Agreements need to be agreed with templates constructed that are legally robust. Again, the degree of flexibility to be allowed within such agreements should be clarified beforehand.
b) Development of long-term monitoring systems and SOPs: The use and outcomes of secondary re-use need to be monitored, as well as levels of satisfaction with the repository and user feedback, from both data generators and requesters. This will require systems and workflows to be developed, supported by SOPs and other quality documents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

c) Development of management data flows and supporting systems: Workflow and infrastructure systems will be required to take the data created within the activities listed above, and summarise it for repository managers and stakeholders, with 'drill-down' into more detailed data when required. This will include the design and provision of a variety of reports.
d) Generation of cost estimates: Towards the end of the project, it should be possible to provide cost estimates for the different activities involved in managing the repository, to help assess both sustainability and scalability to a broader repository system.

The activities listed above are only the most obvious procedural systems required - others may emerge as the repository is developed. Additional services and enhancements (see sections 4.2 and 4.3) will require additional procedures and quality systems to be developed.

7. Sustainability and governance of the COVID-19 repository

Development phase

In the development phase, the repository's governance will be defined by the DOW of WP14 of EOSC-Life [7], an EC H2020-funded project. It is defined by the following structures:

a) Steering Group
b) Stakeholder Forum
c) Scientific Advisory Board

Steering Group

The Steering Group is formed by the WT-leaders of EOSC-LIFE WP14 and covers J. Demotes (ECRIN), G. O. Sundby Thomassen (Uio-TSD) and N. Blomberg (ELIXIR) and a legal expert, M. Matei (ECRIN). The Steering Group will start with the development of the concept and strategy for a COVID-19 repository as input to the other WTs (M19). This will include a plan for early development of a pilot repository, and later on development of optional services relevant for meta-analyses, including conversion into an interoperable (CDISC) data standard, as well as solutions for aggregation of data from multiple repositories (federated cloud, importing data transfer into the EU data repository). Thereafter, plans for governance, business plan, maintenance and sustainability will follow (M37) and will be updated till the end of the project (M48). All necessary decisions will be taken by the Steering Group.

Stakeholder Forum

A stakeholder forum will be formed to critically accompany the development and implementation of the repository. The stakeholder forum will cover representatives from research infrastructures/e-infrastructures, funders, publishers, regulatory bodies, patient organisations, standardisation bodies, scientific organisations related to COVID-19, trial registries/repositories and COVID-19 initiatives/platforms, as well as legal, ethical and technical experts. Two workshops with the stakeholder forum (currently planned as virtual workshops) will be held, one in November 2020 and another in October 2021. The results of the 1st workshop will be used as input for an early demonstrator of the repository (beta version: February 2021) and the 2nd workshop will assess the report about technical

[7] https://www.eosc-life.eu/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

implementation and validation of the COVID-19 portal (delivered August 2021). It is planned to keep the stakeholder forum (maybe at reduced size) after the development phase.

Scientific Advisory Board

In addition, a Scientific Advisory Board will be implemented, bringing experience from outside Europe into the project (e.g. Vivli, NIH). The Scientific Advisory Board will work via video conferences. It is planned to have at least 2 video conferences during the project.

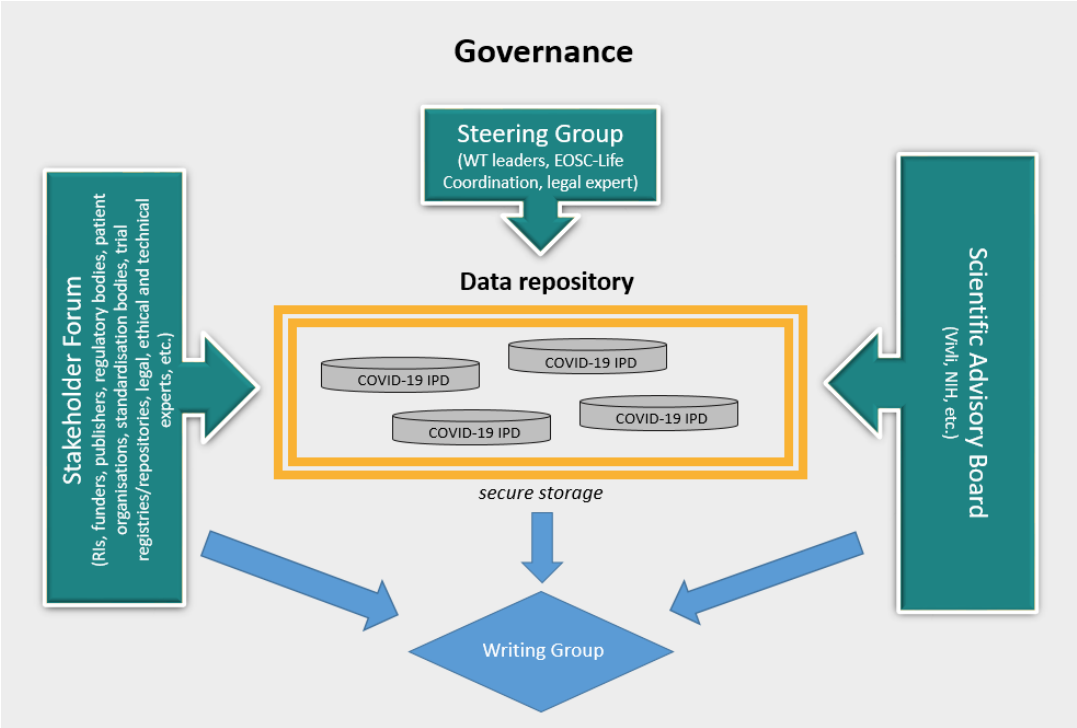

Figure 1: Governance structure and relationship between structures

Post development/routine use

The governance and sustainability plan for the routine phase will be developed during the H2020 project and will be provided as final deliverable (March 2022). At this stage, no definite structure and financing model can be presented. More input is needed from the stakeholders and external scientific advisors to come to feasible and sustainable solutions. By involving sustainable infrastructures (ECRIN, ELIXIR), there will be a good chance to implement and assure long term availability of the repository service. It is planned to integrate the repository service in the European COVID-19 data hub, with a portal operated by ECRIN acting as the interface with the clinical research community and with technical partners providing a secure environment and data sharing services (TSD, ELIXIR).

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

8. Evaluation of routine use and impact, usability and user friendliness

The intention is to evaluate the repository and its systems, both after the initial demonstrator has been constructed and, more comprehensively, once the repository has been established and is in use (this is likely to fall outside the planned project timetable and will therefore be reported independently). The results of the demonstrator evaluation will be fed back into the further development of the system.

Evaluation will encompass the main repository functions, as viewed by different user groups. Thus, features being examined will include, for:

a) Data providers:
- The clarity of guidance and instructions relating to data providers
- The suitability and flexibility of the access arrangements available
- The additional burden, if any, of the repository's demands for de-identification and descriptive metadata
- The ease of completing the Data Transfer Agreement
- Usability and user-friendliness of systems for upload
- Usability and user-friendliness of systems for applying provenance / discoverability metadata
- The systems available for ongoing feedback on requests
- The usefulness of the systems available for supporting the processing of data requests (e.g. Data Access Committees).

b) Repository managers (excluding purely technical management systems):
- The usability of systems for monitoring the status of individual packages of data objects as they go through the required quality checks.
- The ease of entering into communication with data providers / requesters.
- The accuracy and ease of use of systems providing overall monitoring data (of uploads, requests, downloads etc.).

c) Data requesters:
- The ease with which repository contents can be searched, and the usefulness of the descriptions provided
- The clarity of guidance and instructions relating to data requesters
- Usability and user-friendliness of systems for download
- The ease of completing the Data Use Agreement
- Usefulness of the provenance / discoverability metadata that has been applied
- Evaluation of impact of the repository on their work, and overall usefulness of the system

d) Steering Group, Stakeholder Forum and Scientific Advisory Group:
- Levels and volumes of activity within the registry
- Compliance of the registry with identified quality measures (see section 4.5)
- Compliance of the registry with ethical requirements.

Unfortunately, because of the limited time available for developing the initial demonstrator, only dummy or public data is likely to be available in the demonstrator repository, and the various 'users' in the initial evaluation may need to be volunteers (drawn from clinical research units) playing the role of different users, which may limit some aspects of this evaluation.

For both the initial and final evaluations the plan is to develop a detailed written protocol for each of the main evaluations listed (i.e. by user group) and then to publish those protocols beforehand.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

For consistency and reliability, we intend to use the EUCS (End User Computing Satisfaction) scale for usability testing. EUCS consists of 5 components: content, accuracy, format, ease of use and timeliness, and covers 12 items.

We will also collect details on the testers (e.g. role, experience), but exclude personal identifying data, to ensure that we cover a range of potential users. Evaluators will be provided with comprehensive guidance on using and testing the system, and – again for consistency – we may sometimes make use of artificial scenarios and ask them to work through those, recording their experience and impressions.

As well as the distinct evaluation exercises, it is the intention to provide cumulative statistics on the use of the repository, on a monthly basis. These will be similar to the metrics provided by the CSDR repository[8].

Soon after the launch of the completed repository we can also survey potential users, exploring their intentions about sharing IPD from COVID-19 studies and in particular in using this repository as the basis of that sharing. This will also be useful in 'advertising' the repository to those users. We can also attempt to monitor the reaction of social media to the launch of the repository (e.g. Facebook, Twitter.

In the longer term, we will attempt to assess the impact of the use of the repository by looking at the nature of the secondary use of data for research projects (e.g. re-analysis, meta-analysis of IPD, further analysis), the research outputs generated, and the impact of secondary use on research methodology (e.g. the development of new hypotheses and methods) and / or on medical decision making (e.g. a change in treatment policy due to secondary analysis). From the experience with other repositories, evaluation of impact makes sense only after several years of routine use. We will therefore perform the first evaluation of the wider impact of the repository at least 3 years after the official launch.

1. Outreach, partnerships and scalability

A successful adoption of the COVID-19 clinical trial data repository by the clinical research community will require a strong communication policy targeting the main stakeholders as listed in Chapter 2. This outreach activity will benefit from the EU Commission's communication on the COVID-19 data hub operated by the EMBL/EBI, but will also need the development of specific communication channels through websites, social media, webinars and videoconferences, and articles in scientific journals. Targets include in particular:

- COVID-19 investigators from European countries, and beyond
- clinical research data centres (including the network of ECRIN-certified data centres), and clinical trial units
- clinical study sponsors, either academic or industry sponsors, and their data controllers
- funding bodies, as open access to COVID research data is part of the requirements for funding
- COVID trial registries[9]
- the WHO ICTRP[10], as well as the WHO-affiliated trial registries (including the NIH[11])
- medical journals (through the ICMJE)

[8] https://www.clinicalstudydatarequest.com/Metrics.aspx
[9] https://ecrin.org/covid-19-trials-registries
[10] https://www.who.int/ictrp/en/
[11] https://clinicaltrials.gov/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- medicine agencies (the EMA, as well as national competent authorities), and ethics committees
- patient representatives

Adoption of the COVID-19 clinical trial data repository also requires a strong partnership with the COVID-19 scientific community, who will also contribute to designing and testing the instrument, and to make recommendations for improvement and optimal use. This is also true for the HTA bodies, Cochrane collaboration and other organizations specialising on systematic reviews and meta-analyses. Partnership with clinical trial data repositories outside the EU (Vivli and other initiatives), as well as with organisations providing additional services for secondary use of data such as IDDO transposing trial data into an interoperable data standard.

This also leads us to consider the scalability of the COVID-19 clinical trial data repository, through the development of additional functionalities or modules, or through its extension to other diseases conditions. This should be discussed in depth with investigators from other medical disciplines (ECRIN is involved in partnership projects with the paediatric community through PedCRIN, the rare diseases community through the EJP-RD, or the infectious diseases community through ECRAID). Based on a business plan taking into account the technical requirements and the needs of these communities, a proposal for an expansion towards a GDPR-compliant clinical trial data repository open to any medical specialty could then be discussed with potential funders.

References

Alemanno, 2020. The European Response to COVID-19: From Regulatory Emulation to Regulatory Coordination? Eur J Risk Regul. Apr 28: 1–10. doi: 10.1017/err.2020.44. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7218191/

Andorno, 2005. The Oviedo Convention: a European Legal Framework at the Intersection of Human Rights and Health Law. JIBL, Vol 2, 133-143. Available at https://www.academia.edu/457872/The_Oviedo_Convention_a_European_Legal_Framework_at_the_Intersection_of_Human_Rights_and_Health_Law

Banzi et al., 2018. Report about a workshop on sensitive data: Repositories for sharing individual participant data from clinical trials and existing tools/services for storing clinical trial data. doi: 10.5281/zenodo.1438261. Available at https://zenodo.org/record/1438261#.Xz6fvX-g-uc

Banzi et al., 2019. Evaluation of repositories for sharing individual-participant data from clinical studies. Trials 20, 169. doi:10.1186/s13063-019-3253-3, available at https://trialsjournal.biomedcentral.com/articles/10.1186/s13063-019-3253-3

Brehon, 2020. The European Union and the Coronavirus. Fondation Robert Schumann / European issues N°553 / 7th April 2020, available at https://www.robert-schuman.eu/en/european-issues/0553-the-european-union-and-the-coronavirus

Burton et al., 2015. Data Safe Havens in health research and healthcare. Bioinformatics, Volume 31, Issue 20, 15 October 2015, Pages 3241–3248, doi: 10.1093/bioinformatics/btv279. Available from https://academic.oup.com/bioinformatics/article/31/20/3241/195451

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

CNIL, Deliberation n° 2020-083 du 23 juillet 2020 , JORF n°0195 du 9 août 2020, Available at https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000042221437&categorieLien=id

Council of Europe, 1997. The Oviedo Convention. Convention for the protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine. Available at https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/164

Council of Europe, 2005. Additional Protocol to the Convention on Human Rights and Biomedicine, concerning Biomedical Research. Available at https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168008371a

Council of Europe, 2020. Committee on Bioethics: DH-BIO Statement on human rights considerations relevant to the COVID-19 pandemic. Available at https://rm.coe.int/inf-2020-2-statement-covid19-e/16809e2785

Council of the EU, 2001. Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1597941138059&uri=CELEX:32001L0020

Council of the EU, 2013. Decision No 1082/2013/EU of the European Parliament and of the Council, of 22 October 2013, on serious cross-border threats to health and repealing Decision No 2119/98/EC. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013D1082

Council of the EU, 2014. Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC Text with EEA relevance. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1597941138059&uri=CELEX:32014R0536

Council of the EU, 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1597941209291&uri=CELEX:32016R0679

Council of the EU, 2020. Regulation (EU) 2020/1043 of the European Parliament and of the Council of 15 July 2020 on the conduct of clinical trials with and supply of medicinal products for human use containing or consisting of genetically modified organisms intended to treat or prevent coronavirus disease (COVID-19). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020R1043&from=EN

CoreTrustSeal (2020) CoreTrustSeal Certification, available at https://www.coretrustseal.org/, accessed 13/08/2020

ECRIN, 2020. COVID-19 Taskforce. Available at https://ecrin.org/covid-19-taskforce

EDPB, 2019. Opinion 3/2019 concerning the Questions and Answers on the interplay between the Clinical Trials Regulation (CTR) and the General Data Protection regulation (GDPR) (art. 70.1.b)). Available at https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_opinionctrq_a_final_en.pdf

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

EDPB, 2020. Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak. Available at https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202003_healthdatascientificresearchcovid19_en.pdf

European Commission, 2019. Question and Answers on the interplay between the Clinical Trials Regulation and the General Data Protection Regulation. Available at https://ec.europa.eu/health/sites/health/files/files/documents/qa_clinicaltrials_gdpr_en.pdf

European Commission, 2020a. Commission recommendation (EU) 2020/518 of 8 April 2020 on a common Union toolbox for the use of technology and data to combat and exit from the COVID19 crisis, in particular concerning mobile applications and the use of anonymised mobility data. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020H0518

European Commission, 2020b. Communication from the Commission to the European Parliament, the European Council, the Council and the European Investment Bank: Strategy for COVID-19 vaccines. Available at https://ec.europa.eu/info/sites/info/files/communication-eu-strategy-vaccines-covid19_en.pdf

European Commission, 2020c. Communication from the Commission to the European Parliament, the European Council, the European Economic and Social Committee and the Committee of the Regions: Short-term EU health preparedness for COVID-19 outbreaks. Available at https://ec.europa.eu/info/sites/info/files/communication_-_short-term_eu_health_preparedness.pdf

European Commission, 2020d. Joint European Roadmap towards lifting COVID-19 containment measures. Available at https://ec.europa.eu/info/sites/info/files/communication_-_a_european_roadmap_to_lifting_coronavirus_containment_measures_0.pdf

European Medicines Agency, 2020. Guidance on the management of clinical trials during the COVID-19 (Coronavirus) pandemic. Version 3, 28/04/2020. Available at https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/guidanceclinicaltrials_covid19_en.pdf

European Parliament, 2020: States of emergency in response to the coronavirus crisis: Situation in certain member states (Briefing). Available at https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2020)649408

Hallinan et al., 2020. International transfers of health research data following Schrems II: A problem in need of a solution. Zenodo. http://doi.org/10.5281/zenodo.4017925

HHS.Gov (2020) Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, available at https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard

Hrynaszkiewicz, 2016. Publishing descriptions of non-public clinical datasets: proposed guidance for researchers, repositories, editors and funding organisations. Research Integrity and Peer Review, volume 1, Article number 6. Available at https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-016-0015-6

IDDO, 2020. COVID-19, Contributing data. Available at https://www.iddo.org/covid19/data-sharing/contributing-data

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Jahns et al., Broad donor consent for human biobanks in Germany and Europe: a strategy to facilitate cross-border sharing and exchange of human biological materials and related data, Available at : https://www.degruyter.com/view/journals/labm/43/6/article-p291.xml

Kalkman et al., 2019. Responsible data sharing in international health research: a systematic review of principles and norms, BMC Medical Ethics 20:2, doi: 10.1186/s12910-019-0359-9, available at https://bmcmedethics.biomedcentral.com/track/pdf/10.1186/s12910-019-0359-9

Kaye et al., 2016. Consent for Biobanking: The Legal Frameworks of Countries in the BioSHaRE-EU Project. Biopreserv Biobank. 14(3):195-200. doi:10.1089/bio.2015.0123

Library of Congress, 2020. European Union: Legal Responses to Health Emergencies, available at https://www.loc.gov/law/help/health-emergencies/europeanunion.php

Lin et al, 2020. The TRUST Principles for digital repositories. Sci Data 7, 144, doi:10.1038/s41597-020-0486-7, available at https://www.nature.com/articles/s41597-020-0486-7

McCullagh et al., 2019. (Eds.), National Adaptations of the GDPR, Collection Open Access Book, Blogdroiteuropeen, Luxembourg, 130 pages Available at: https://wp.me/p6OBGR-3dP

O'Connell and Plewes, 2015. Sharing Research Data to Improve Public Health in Africa, available at https://www.nia.nih.gov/sites/default/files/2017-06/sharing-research-data-to-improve-public-health-in-africa_0.pdf

Ohmann et al., 2017, Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ open, 7(12), e018647, doi: 10.1136/bmjopen-2017-018647, available at https://bmjopen.bmj.com/content/7/12/e018647

Ohmann et al., 2018. Classification of processes involved in sharing individual participant data from clinical trials. F1000Research 7:138, available at https://f1000research.com/articles/7-138

RDA, 2020. RDA COVID-19 Recommendations and Guidelines on Data Sharing, available at https://www.rd-alliance.org/system/files/RDA%20COVID-19%20Recommendations%20and%20Guidelines%2C%2030%20June%202020_Endorsed-Final_0.pdf

Speakman et al., 2017. Pandemic legislation in the European Union: Fit for purpose? The need for a systematic comparison of national laws. Health Policy. doi: 10.1016/j.healthpol.2017.08.009. Available at https://www.sciencedirect.com/science/article/pii/S0168851017302221

Tilki et al., 2020. Assessment of a demonstrator repository for individual clinical trial data built upon DSpace. F1000 Research, 9:311. doi: 0.12688/f1000research.23468.2, available at https://f1000research.com/articles/9-311/v2

Vivli, 2020. How to share your clinical research data on the Vivli COVID-19 portal, available at https://vivli.org/how-to-share-your-research-on-the-vivli-covid-19-portal/

WHO, 2005. International Health Regulations. Available at https://www.who.int/ihr/publications/9789241580496/en/

WHO, 2015. Developing Global Norms for Sharing Data and Results during Public Health Emergencies, available at https://www.who.int/medicines/ebola-treatment/data-sharing_phe/en/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Wilkinson et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3,160018; doi:10.1038/sdata.2016.18, available at https://www.nature.com/articles/sdata201618

Abbreviations

BBMRI	Biobanking and Biomolecular Resources Research Infrastructure
BioLINCC	Biological Specimen and Data Repository Information Coordination Center
CDASH	Clinical Data Acquisition Standards Harmonization
CDISC	Clinical Data Interchange Standards Consortium
CNIL	Commission nationale de l'informatique et des libertés
CORBEL	Coordinated Research Infrastructures Building Enduring Life-science Services
COVID-19	Coronavirus Disease 2019
CSV	Comma Separated Values
CTR	Clinical Trial Regulation
DAC	Data Access Committee
DOI	Digital Object Identifier
DOW	Description of Work
DTA	Data Transfer Agreement
DUA	Data Use Agreement
EATRIS	European Infrastructure for Translational Medicine
EBI	European Bioinformatics Institute
ECRAID	European Clinical Research Alliance on Infectious Diseases
ECRIN	European Clinical Infrastructure Network
EDPB	European Data Protection Board
EEA	European Economic Area
EEC	European Economic Community
EJP-RD	European Joint Programming - Rare Diseases
ELSI	Ethical, Legal, and Social Issues
EMA	European Medicines Agency
EMBL	European Molecular Biology Laboratory
EOSC	European Open Science Cloud
EU	European Union
EUCS	End User Computing Satisfaction
FAIR	Findability, Accessibility, Interoperability, and Reusability[12]
GDPR	General Data Protection Regulation

[12] https://en.wikipedia.org/wiki/Reusability

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

HTA	Health Technology Assessment
ICJME	International Committee of Medical Journal Editor
IDDO	Infectious Diseases Data Observatory
IGA	Identity Governance and Administration
IHR	International Health Regulations
IPD	Individual Participant Data
ISO	International Organization for Standardization
ISARIC	International Severe Acute Respiratory and Emerging Infection Consortium
MVP	Minimum Viable Product
NIH	National Institutes of Health
PaaS	Platform as a Service
PedCRIN	Paediatric Clinical Research Infrastructure Network
PID	Personal Identifier
RDA	Research Data Alliance
SDTM	Study Data Tabulation Model
SOP	Standard Operating Procedure
TDR	Trustworthy Data Repository
TSD	Tjenester for Sensitive Data
TSV	Tab-separated Values
UIO	University of Oslo
WHO	World Health Organisation
XML	Extensible Markup Language

Delivery and Schedule

The delivery is delayed:

No

Adjustments

Adjustments made:

None

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.