HEALTHYCLOUD
Health Research & Innovation Cloud

D2.4. Guideline on ELSI compliant Governance Models
Version 1.1

Document Information

Contract Number	965345
Project Website	http://www.healthycloud.eu/
Contractual Deadline	M27, May 2023
Dissemination Level	PU-Public
Nature	R-Report
Author(s)	Davit Chokoshvili (PNED G.I.E., WP2)
Contributor(s)	Michaela Th. Mayrhofer (BBMRI-ERIC, WP2) Regina Becker (PNED G.I.E., WP2) Francesco Vigna (PNED G.I.E., WP2)
Reviewer(s)	Carlos Luis Parra Calderón (SAS) Ramón Launa Garcés (IACS) Irene Kessissoglou (Sciensano)
Keywords	ELSI Guidance, GDPR, Governance, Compliance

Notice: *The HealthyCloud project has received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement Nº965345*

Change Log

Version	Author	Date	Description of Change
V 1.0	Davit Chokoshvili	2023/05/22	Complete Draft
V 1.1	Davit Chokoshvili Michaela Th. Mayrhofer Carlos Luis Parra Calderón	2023/05/30	Final document addressing reviewers comments
	Davit Chokoshvili	2023/06/07	Submission to EC

Table of contents

Executive Summary

This deliverable focuses on the ELSI-compliant governance of Health Research & Innovation Clouds (HRICs). The deliverable builds on the insights and learnings generated by multiple Work Packages over the course of the HealthyCloud project, synthesising these outputs into a comprehensive analysis of the ELSI compliance framework for the HRICs. In its conceptual analysis, this deliverable takes a broad view on data processing in the HRIC, examining the entire data reuse lifecycle taking place in the HRIC ecosystem. In doing so, the deliverable analyses compliance challenges encountered by relevant parties beyond the providers of HRIC Infrastructures. This bigger-picture approach focusing on the global compliance needs across the HRIC ecosystem helps identify potential gaps where the providers of the HRIC Infrastructure services could support relevant stakeholders achieve their legal, regulatory, and ethical governance obligations. The insights elicited through this analysis are subsequently transposed into actionable guidance primarily aimed at HRIC Infrastructures, as well as the ancillary support services being currently defined by the HealthyCloud consortium.

1. Introduction

The DoA of the HealthyCloud project envisages D2.4 as one of the summary deliverables, synthesising key insights, learnings, and recommendations into an overarching guidance focusing on Ethical, Legal and Societal Implications (ELSI). Thus, complementing the technical, operational, and organisational requirements of the planned Health Research & Innovation Clouds (HRICs), D2.4. seeks to contribute ELSI-relevant elements of the governance of HRICs.

The DoA primarily situates D2.4 within **T2.2: ELSI compliance of the governance of the future HRIC decentralized platform**:

"This task will focus on important aspects such as: roles of controller and processor; responsibilities over the data life cycle in an HRIC; requirements for GDPR; ethical compliance in access management and data use (including transfer outside the EEA); and consideration of country-specific regulations. The result will be a set of Guidelines for an ELSI compliant implementation of different HRIC governance approaches (D2.4) …"

Additionally, D2.4. is closely linked to, and complements, other tasks of WP2, while also drawing valuable insights from the output generated by other WPs. In this respect, of particular relevance are the following two lines of work completed by partners from other WPs:

- The comprehensive review of existing computational infrastructure models completed by WP5 (including ELSI and governance aspects), in conjunction with the functional requirements for HRICs identified by WP7; and

- Key scenarios involving data access and use through a HRIC, delineated by WP4 and subsequently transposed into the context of the HRIC FAIR Data Portal (WP6) as well as overall data flows (WP7)

Incorporating various insights from these work packages, and building upon the output generated internally in WP2, the present deliverable sets out to formulate generic, yet actionable ELSI guidance for the governance of HRICs, with a focus on the infrastructural components of the HRIC.

It is worth highlighting that the present guidance does not exhaustively cover all the elements of the HRIC being defined by the HealthyCloud Consortium. The draft strategic agenda (D8.1.) has proposed to conceptualise the HRIC "as an 'interface' of services, a specified list of resources and related functionalities designed to meet identified needs [of users and organisations within the research ecosystem]." However, the scope of this deliverable (D2.4), as per the DoA, is focused on the compliance challenges directly associated with the processing of patients' and research participants' data across the data lifecycle in the HRIC. Consequently, governance of HRIC services, resources and related functionalities not directly involved in the processing of patient/research participant data (and hence not facing the same compliance challenges) are largely beyond the scope of this deliverable.

This guidance views the ELSI concerning data lifecycle in the HRIC through the lens of the European General Data Protection Regulation (GDPR), particularly in terms of framing the roles and compliance obligations of various parties involved in the HRIC ecosystem. This approach reflects the pre-eminence of the GDPR in ethical and legal compliance of HRICs and helps link to the output of previous HealthyCloud project deliverables, which also primarily focused on GDPR-related aspects (e.g., D2.1 and D2.2). However, the ELSI challenges discussed in this document are not exclusively GDPR-derived issues and incorporate additional relevant areas such as legal frameworks other than the GDPR, as well as more general ethical governance considerations.

2. The General Data Protection Regulation

The EU General Data Protection Regulation (GDPR) came into force in May 2018 and regulates the processing of personal data. Personal data is defined broadly under the GDPR. Namely, any information that can be used to directly or indirectly (e.g., in combination with another piece of information) identify a natural person to whom the information concerns (i.e., the data subject), is considered personal data under the GDPR[1]. Owing largely to this broad definition of personal data, it is

[1] Article 4(1) and Recital 26 GDPR

assumed that most, if not all, individual-level (i.e., record-level) information processed in the HRIC should be treated as personal data whose processing is subject to the GDPR.[2]

In its conclusion section, D2.2 (Framework of modular contract clauses for HRICs), WP2 stated the following:

"While the GDPR has been in force for several years, given its relative novelty, it remains a disruptive force in terms of contractual obligations. There are still competing GDPR interpretations (e.g., assignment of controllership, legal basis, identifiability) that impact on data sharing agreements. The GDPR has also significantly increased the due diligence surrounding contract execution – before (due diligence, negotiation, signing) and after (monitoring of obligations)".

Given the scope of D2.2, the conclusion is formulated in relation to contractual arrangements under the GDPR. However, the broader point is that ensuring and demonstrating GDPR compliance is both conceptually challenging (owing to the legal uncertainties inherent in the Regulation) and highly resource-intensive, due to the broad scope of the GDPR. This is particularly problematic in the context of complex data processing environments such as the HRIC, which may be composed of a plethora of data-processing entities. These entities may have different architectural patterns, operating under a patchwork of local legal frameworks (including national implementations of the GDPR) that interpret and operationalise GDPR requirements in dissimilar, possibly inconsistent ways.[3]

As such, the present deliverable, D2.4., seeks to disambiguate some of the interpretive confusion around the GDPR by providing greater conceptual clarity for the context of the HRIC. To that end, D2.4 will examine the following phases in the data lifecycle.

1.Initial collection of personal health data and its primary use – while technically not part of data processing within the HRIC infrastructure, which focuses on the reuse of existing health-related data, this phase is crucial for the ELSI governance and legal compliance analysis. Initial data collection may take place in various healthcare and research contexts, including, for example, a hospital providing medical care to its patients, or a biomedical research institution collecting data to use it for its own research project.

2. Making data systematically available through the HRIC – this means that an existing dataset are made widely discoverable by external parties (through, for example, the HRIC FAIR Data Portal – WP6) who would be interested in accessing and using the data for own purposes. Of note, the party responsible for data

[2] See, for example, D2.1 First draft on legal framework for technical safeguards with a focus on cloud usage; Section 3.2 – "Anonymisation (anonymous use of data?)"

[3] Assessment of the EU Member States' rules on health data in the light of GDPR (DG Health and Food Safety) https://health.ec.europa.eu/system/files/2021-02/ms_rules_health-data_en_0.pdf

availability through the HRIC may not necessarily be the hospital or the research institution that originally collected the data in step 1.

3. Granting an external party permission to access and use the data (aka "data disclosure") for the external party's own purposes – following the discovery of a relevant data resource by the external party, the party requests data access permission. This may take place in one of the scenarios described in D6.3., depending on various contextual factors. For the purposes of this abstract conceptual framework, the specifics of how (and by whom) the prospective data user is granted the permission are not of decisive importance.

4. The external party (now the data user) accesses and uses the data for the approved purpose(s) – Depending on the architectural pattern of the infrastructures supporting the data user, this may or may not involve direct access to the data. For example, under certain scenarios, the data user may be allowed to access, or even download, entire underlying datasets, while in other cases, the data user would be permitted to query data remotely, without directly accessing the data. Once again, for the purpose of this high-level conceptual framework, the exact modality of access and/or use is not relevant; what matters is that processing of the data takes place for the purpose(s) pursued by the data user.

The subsequent section uses this general framework to elucidate GDPR roles of different parties involved in the processing chain.

2.1. GDPR Roles of parties in the HRIC

Under the GDPR, the most important categories of parties involved in the processing of persona data are the **controller** and the **processor**.

The controller is the party that, alone or jointly with others, determines the purposes and means of the processing of personal data, also commonly referred to as the "why and the how" of processing personal data. The processor, on the other hand, is a party that processes personal data on behalf of the controller. As such, the processor shall process personal data pertaining to data subjects in accordance with explicit instructions of a controller.[4]

[4] The GDPR affords the processors some discretion as to "how" the processing should take place. However, as per guidance by the EDPB (see the subsequent footnote), this is typically limited to specific implementation aspects, such the choice of software, which are referred to as "non-essential means". On the other hands, more substantive, or "essential means" of processing (for example, the choice of the categories of data or data subjects) are defined by the Controller.

It is worth mentioning, however, that controller and processor are functional roles that must be defined contextually, i.e., with respect to specific processing operations taking place across the data lifecycle.[5] In complex, multi-phased data processing environments such as data reuse envisaged under the HRIC, it might be insufficient, and potentially misleading to assign GDPR roles in a generic manner.

Of note, some of the earlier HealthyCloud deliverables did elect to assign the GDPR roles in a generic manner; however, those deliverables usually described a simplified scenario of data-sharing involving three parties: i) the data provider, ii) an intermediary such as a Health Data Hub or a Secure Processing Environment (SPE), and iii) the data user. Under these simplified assumptions, assigning GDPR roles in a more generic manner were not necessarily problematic. On the contrary, this approach offers significant advantages by simplifying the discussion. However, as the HealthyCloud consortium proceeded to map out the data lifecycle though the HRIC in a more comprehensive manner, it became clear that the upstream chain of processing operations taking place within the HRIC ecosystem could involve a longer list of entities, including: data producers, data providers, SPEs, Infrastructure Providers, and Health Data Hubs, among others. Although in certain scenarios some of these roles can be merged in a single entity, it is also possible to envisage considerably longer chains of data processing where these roles are handled by different parties[6]. This, in combination with the existence of additional entities beyond the upstream chain, (i.e., at a minimum, the HRIC FAIR data portal and the data user), requires a more granular approach to defining the GDPR roles in the HRIC.

Given the sheer number of possible ways to structure a data lifecycle in a HRIC, it is impractical to examine GDPR roles of the parties involved in data processing under each possible path. Rather, the analysis below will rely on the 4 phases of the generic data lifecycle model described earlier in order to correctly identify controllers.

2.2. Types of controllers within the HRIC data lifecycle

In complex processing environments involving multiple parties along a phased data lifecycle, the question “who is the data controller?”, absent any additional context, loses its relevance. Unlike, for example, the “data subject”, whose identity is inherently linked to the personal data relating to that data subject, the identity of the controller will change depending on the phase of the data lifecycle and the

[5] The European Data Protection Board (EDPB). Guidelines 07/2020 on the concepts of controller and processor in the GDPR: https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf

[6] In exceptional cases, it may be that all of these roles are assumed by a single entity. In practice, however, this is only foreseeable in the case of large, prospective population cohorts with own established data infrastructures, including an own purpose-built SPE.

specific processing operation. Therefore, the question should be rephrased in a more granular form: "who is a controller for a particular processing operation?"

In this regard, the four-step representation of the data lifecycle provides a helpful conceptual framework for correctly identifying controllers along the data reuse lifecycle in the HRIC ecosystem. As described in table 1 below, there can be up to four controllers along the data lifecycle, corresponding to the processing operations required to successfully complete each of the four phases[7]. The four potential groups of controllers are labelled as the **CTI** (Controller for the initial data collection), **CTA** (Controller for making data available via a HRIC), **CTD** (Controller for data disclosure to a specific user) and **CTU** (Controller for the (re)use of the data).

Table 1. Four possible categories of controllers across the data reuse lifecycle in a HRIC

Party	Controller for what?	Examples of processing operations
CTI	Initial collection of personal data and its primary use *(taking place outside the HRIC)*	- Data collection directly from the data subject (e.g., patient intake questionnaires) - Data generation by analysing biological samples provided by a patient/research participant - Using the data for the specific medical (e.g., diagnostic/treatment) or research purpose pursued by CTI (i.e., primary use)
CTA	Making the data widely available for a secondary use in the HRIC	- Generating aggregated metadata to aid discoverability (e.g., through the HRIC FAIR Data Portal) by CTUs - Data cleaning, structuring, curating to defined standards/formats, or otherwise making it ready for a reuse through the HRIC - Transfer of the data to an external Data Hub or another type of SPE; or, if retaining the data on premise: - Implementing a local instance of an SPE, allowing remote access and/or querying of the data by CTUs
CTD	Disclosing data to a particular data user (or otherwise granting permission to use the data)	All operations needed to enable data use by the CTU. Depending on the architectural pattern of the Data Hub, SPE or other data infrastructure service provider relied upon by the CTD, this may entail any of the following: - Data transfer to the CTU - Granting access to the relevant sub-set of the data in a dedicated digital workspace on an SPE - Generating a digital token that the CTU can use to instruct the SPE to perform a federated/distributed analysis

[7] For simplicity, this conceptual model does not capture the possibility of joint controllership.

		- Approving a bespoke data analysis algorithm brought to the SPE by the CTU.
CTU	Controller for using the data for an own (approved) purpose	All data processing operations needed to achieve the purpose(s) pursued by the CTU, specified by the CTU in the data access request, (and approved by the CTD)

The CTI, the controller that collected the data from the data subject, will in many cases be a Data Producer, as defined under the HealthyCloud Glossary (See Annex I). However, it's important to note that not all Data Producers will be CTIs. For example, it is possible for the CTI to instruct another entity, such as a diagnostic laboratory or a genome sequencing centre, to undertake initial data collection or generation. While such an entity would be a Data Producer, in terms of the GDPR, it would be acting as a processor on behalf of the CTI.

In practice, it is likely that the CTI and the CTA will be the same entity. Healthcare and biomedical research organisations that collect personal data for own purposes would often act as data custodians in relation to the resultant data collections, i.e., deciding on whether and how the data should be re-used. However, this need not necessarily be the case. For example, the CTI could transfer the data collection to another entity that intends to use the data for own purposes. This would make the recipient another independent controller in relation to the purposes pursued by this entity. Should the recipient controller subsequently decide to make the dataset available in the HRIC, (and provided that doing so is lawful under the GDPR), this would result in the CTI and the CTA being two different entities.

The CTD is the controller tasked with evaluating and deciding upon individual data access requests in relation to the dataset made discoverable through the HRIC. In some cases, the roles of the CTA and the CTD will be assumed by the same controller. This will always be the case where the CTA, after making data discoverable and requestable via the HRIC, also retains decision-making authority in relation to individual data access and use requests. However, there are numerous situations where one of the intermediaries in the data lifecycle could assume the role of the CTD. This situation may arise when a medical institution that holds the data engages an external repository for the purposes of sharing data with the downstream users and also delegates to the repository decision-making in relation to granting access to future users. In the context of the HRIC ecosystem, a review of the landscape of European Data Hubs revealed diverse governance patterns with respect to access decisions (D4.1.). While some data hubs acted as processors with respect to data access decisions (i.e., relied exclusively on the data provider's own DAC – Data Access Committee), others exhibited a more centralised data access governance, effectively making them the CTD. These insights obtained by WP4 highlights that the allocation of controllers' roles will vary across different scenarios under the HRIC ecosystem. Moreover, under the proposed European Health Data Space (EHDS) Regulation, the Health Data Access Bodies (HDABs) will likely act as CTDs, with data providers (i.e., the institutions whose data are being disclosed by HDABs to data users) becoming CTAs.

Finally, the CTU is a party interested in accessing and using an existing health dataset via a HRIC for an own purpose, such as a specific research project pursued by the CTU. As the name suggests, this party acts as the controller in relation to data processing operations pursued by the CTU.

Differentiating among these four sub-categories of controllers is important in designing a GDPR-compliant data governance framework for the HRIC.

3. ELSI-compliant processing of personal data in the HRIC: General Considerations

Due to the heterogeneous composition of the European HRIC ecosystem, it is to be expected that the different actors participating in the HRIC will face multitude of dissimilar ELSI compliance needs and challenges. Although each organisation participating in the HRIC is ultimately responsible for its own compliance, there are opportunities for the HRIC to support the participating organisations' compliance efforts by providing interpretive guidance at the central level.

To this end, WP2 is currently defining the HRIC Legal/Regulatory Pillar[8], one of the potential HRIC support services to be proposed as a concrete output of the HealthyCloud project. This HRIC Legal/Regulatory Pillar is envisaged to provide a compliance support service as one of its main components. An important role of the compliance support service would be to establish and maintain an up-to-date overview of the data flows within the HRIC ecosystem, including a complete list of all relevant actors involved in data processing. Subsequently, the personnel operating the HRIC compliance support service can map out the key substantive and procedural compliance obligations applicable to each party, ensuring that the service is well-positioned to provide ELSI guidance to various actors in the HRIC ecosystem, as needed.

The groundwork for establishing data flows and mapping out the relevant actors has been performed by other WPs, culminating in the generalised HealthyCloud data flows prepared by WP7. However, the data flows overview required for ELSI compliance purposes will be more extensive, additionally capturing data processing activities that are, from a technical and operation point of view, not part of the data processing in the HRIC infrastructures. This includes, for example, initial data collection by the CTI, as well as downstream uses of the data by a CTU beyond the scope of the purpose for which the CTU has been granted permission to process the data (if applicable).

[8] Preliminary description of the proposed service can be accessed here

Following the allocation of the GDPR roles (i.e., controllers and processors) to various parties involved in the HRIC, it is essential to delineate, and document envisaged data processing in a Record of Processing Activities (ROPA), in accordance with Art. 30 of the GDPR. While each controller is required to maintain a ROPA for the processing operations under the controller's responsibility (Art. 30(1) GDPR), in a closely integrated data reuse lifecycle facilitated by the HRIC, controllers' responsibilities will be inextricably linked. Therefore, also in this respect, a central HRIC Legal/Regulatory Pillar will be of substantial value, maintaining a global view on the data flows and, if/as needed, aiding the different parties in completing their local ROPAs.

In mapping out the roles of the parties involved in the HRIC-facilitated data reuse lifecycle, the Legal/Regulatory pillar should rely on the framework described in Table 1 to identify different categories of controllers. Other entities involved in the data reuse lifecycle will be processors acting on behalf of the controllers.

The following sub-sections further elaborate on the key GDPR (and other ELSI) compliance obligations applicable to controllers and processors participating in the HRIC ecosystem. With respect to controllers, particular emphasis is placed upon elucidating which compliance obligations are of the greatest relevance to each category of controllers involved in the HRIC-facilitated data reuse lifecycle.

4. Compliance obligations in the HRIC: Controllers

Under the GDPR, controllers must ensure that the processing of personal data takes place in compliance with the principles of the GDPR, including the Accountability Principle, which requires controllers to demonstrate this compliance. The principles of the GDPR are listed below:

Box 1. Principles relating to processing of personal data (Article 5 GDPR)

1. Personal data shall be:

a. processed lawfully, fairly and in a transparent manner in relation to the data subject ('**lawfulness, fairness and transparency**');
b. collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('**purpose limitation**');
c. adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('**data minimisation**');
d. accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('**accuracy**');
e. kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the

> appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('**storage limitation**');
>
> f. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('**integrity and confidentiality**').
>
> *2. The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1* ('**accountability**').

The Art. 5 GDPR principles are further implemented by subsequent articles of the Regulation, which describe substantive and procedural obligations adherence to which will allow controllers to achieve and demonstrate compliance with the principles, and hence the spirit, of the GDPR. Controllers must additionally take special care to comply with the Art. 25 GDPR framework of Data Protection by Design and by Default (DPbDD). This requires a deliberative approach to the design of processing, a careful assessment of the implications of the envisaged processing for the data subjects' rights and freedoms, and, subsequently, selection of the appropriate technical and organisational measures (TOMs).

The aforementioned general obligations incumbent upon controllers give rise to a wide range of concrete substantive and procedural requirements. Some of these may be clearly specified and highly prescriptive by nature, such as the requirement to maintain a ROPA for the relevant processing activities (Art. 30(1) GDPR), or to enter in a data processing agreement (DPA) when engaging a processor, with the mandatory content elements of the DPA laid down in Art. 28(3) GDPR. Most compliance aspects incumbent upon controllers, however, are amenable to greater interpretation, which necessitates significant judgement on the part of the controller. For example, performing a Data Protection Impact Assessment (DPIA), an important instrument for demonstrating the controller's compliance, is required where certain criteria are met, such as: "*systematic* and *extensive* evaluation of *personal aspects* … based on automated processing", or "processing *on a large scale* of special categories of data" (Art. 35(3) GDPR; emphases added). These are, however, largely qualitative criteria, which need to be contextually assessed and interpreted by the controller. The GDPR is also not prescriptive as to the structure and format of the DPIA, leaving considerable room for the controller to decide how to perform the assessment. Finally, where after completing the DPIA the controller concludes that the intended processing results in a high risk to data subjects, the controller is required to consult the relevant supervisory authority (Art 36(1) GDPR). However, ascertaining whether the threshold of "high risk" is reached is also subject to the controller's interpretation.

Arguably, the task of achieving and demonstrating compliance with the GDPR by controllers becomes even more opaque in relation to the principles of the Regulation (Box 1). In particular, translating the controllers' principles-based general obligations into concrete TOMs, workflows, and suitable compliance documentation policies poses a significant strategic challenge. This makes controllers' compliance with the GDPR principles extremely broad and resource-

intensive.[9] In the context of the HRIC, challenges to achieving and demonstrating GDPR compliance by controllers are further exacerbated by the complexity intrinsic to the HRIC ecosystem. This complexity engenders interdependencies among controllers in terms of GDPR compliance obligations. In other words, each controller's ability to achieve and demonstrate its GDPR compliance, particularly with respect to the principles of the Regulation is, in large part, contingent upon the extent to which other controllers throughout the data lifecycle are GDPR-compliant. To partially mitigate this complexity, HealthyCloud WP2 has developed a comprehensive framework and modular templates for contractual agreements, intended to be used by parties in the HRIC ecosystem (see D2.2). However, while the contractual clarity enabled by such agreements is a prerequisite for GDPR-compliant HRICs, it is not a sufficient condition, owing to the intrinsic functional interdependencies among the controllers in relation to their GDPR compliance obligations. As the examples below illustrate, the longer and the more complex the chain of data processing in the HRIC ecosystem, the more difficult achieving and demonstrating GDPR compliance for each controller.

4.1. Compliance interdependencies among the controllers

To illustrate complex compliance-related interdependencies among controllers across the HRIC-facilitated data reuse lifecycle, consider the principle of lawfulness, fairness and transparency (Art 5(1)(a) GDPR). The first component of the principle, "lawfulness"[10] is a broad concept, but in the context of GDPR compliance discussions it is commonly used synonymously with the "legal basis" for the processing of personal data under the GDPR. The GDPR legal basis, in turn, refers to the six options under Art. 6(1), in conjunction with the ten additional options under Art 9(2) GDPR, when the personal data undergoing processing constitutes a "special category of personal data" (including health data). Thus, in order to lawfully process special categories of data, the controller must have a valid GDPR legal basis, meaning the controller must choose a suitable Art. 6(1) condition in conjunction with a corresponding Art. 9(2) option.[11]

The fundamental issue with the GDPR legal bases, in the sense of both Art. 6(1) and Art. 9(2) GDPR, is that their availability to controllers varies across the Member States as well as types of organisations. In other words, a particular GDPR legal basis may be available to a governmental or publicly funded entity, but not a private entity for the same data processing activity (or vice versa, depending on the specific

[9] See, for example, https://arxiv.org/pdf/1808.07338.pdf

[10] It is also common to refer to "Lawfulness", "Fairness" and "Transparency" as separate principles, even though, strictly speaking, they are components of the same principle.

[11] Becker, Chokoshvili, Dove. Forthcoming (2023). See also DG Health and Food Safety cited in [3]

GDPR legal basis and the context of data processing).[12] In principle, each controller in the HRIC ecosystem should be able to assess and demonstrate that it has a valid GDPR legal basis to undertake the intended processing of personal data, as organisations are expected to understand the applicable laws and regulations under which they operate. However, in the context of a complex environment such as the HRIC-facilitated data reuse lifecycle, controllers cannot be expected to systematically ascertain that other controllers also have a valid legal basis in relation to the processing operations under the other controllers' responsibility. Doing so would, in many cases, require an in-depth knowledge of the legal framework applicable to the other controllers. To partially mitigate this issue, the relevant controllers can enter in a contractual agreement (using modular clauses described in D2.2) whereby they declare that each controller has a valid legal basis to process personal data. However, this contractual clarity, while necessary, may not be sufficient due to functional interdependencies among the controllers in relation to their GDPR compliance.

For example, assume the simplest possible scenario in the HRIC, where there are only two entities acting as controllers: the sole "upstream controller" taking on the roles of CTI, CTA and CTD; and the downstream controller, i.e., the CTU.[13] The upstream and the downstream controllers can enter in a contractual agreement whereby both controllers confirm that they each have a valid GDPR legal basis to process personal data for their respective purposes. However, without asking the downstream controller to specify the GDPR legal basis, as well as to demonstrate its validity, the upstream controller runs a risk of non-compliance with respect to its other obligations, including the remaining components of the same Art. 5(1)(a) principle, "Fairness" and "Transparency". Fairness and transparency of processing are further implemented in Chapter III of the GDPR (data subjects' rights). Under Chapter III provisions, the controller that initially collected the data (i.e., the CTI, or the "upstream controller" in this simplified case) has certain obligations vis-à-vis the data subjects. First, unless explicitly exempted from doing so under a relevant Member State or European law, the CTI must notify the data subject about further processing of personal data and inform them of any substantial changes to the nature of processing.[14] This includes, for example, changes in the legal basis,

[12] Ibid.

[13] NB: the number of controllers should not be equated with the number of parties involved in the HRIC-facilitated data lifecycle. For example, in this scenario with only two controllers, there will likely by other parties acting as processors, such as the Data Hub provider, the HRIC federated Computational Infrastructure service, and the HRIC FAIR Data Portal, among others. Otherwise, in the absence of processors, this would be a simple bilateral data-sharing between two parties taking place outside the HRIC ecosystem.

[14] Reuse of previously collected personal data collection in a new research project will typically constitute "further processing of personal data" under the GDPR. For a detailed legal analysis of this issue, see Becker et al. 2022: https://brill.com/view/journals/ejhl/30/2/article-p129_1.xml

alongside explaining the implications of this change (e.g., whether and to what extent the data subject can exercise its GDPR rights).[15] Owing to the uneven availability of GDPR legal bases across member states and types of institutions, it is expected that different controllers involved in data processing within the HRIC will rely on different GPDR legal bases. Moreover, as the CTI remains the principal point of contact for the Data Subject, the CTI may, at any time, receive requests from its data subjects to exercise their GDPR rights. As data subjects are progressively aware of their GDPR rights, such requests are becoming increasingly common, including at medical research institutions.[16] The CTI is required, under Art. (19) GDPR, to not only respond to such requests, but also to act as an intermediary between the data subject and the downstream recipients of the data. Should, at this point, the CTI discover that the downstream controller relies on a legal basis under which a particular right does not apply – and hence the data subject cannot exercise the right – the CTI would be in a clear breach of its transparency and fairness obligations vis-à-vis the data subject, having failed to notify the data subject in advance. Worse still, following this inquiry by a data subject, the CTI could discover that the downstream controller does not have a valid GDPR legal basis, despite having claimed to the contrary in the contractual agreement signed between the parties. In all cases, the CTI would be liable to the data subjects' complaints to the relevant supervisory authority and, potentially, subsequent monetary fines.

Importantly, the example above is simplified in that it assumes there is only one "upstream controller", taking over the roles of CTI, CTA and CTD. It is likely that in practice, upstream controllership within the HRIC ecosystem, encompassing three roles, will often be fragmented between two or more entities, for reasons elaborated in the next section.

The emphasis on the GDPR legal basis is warranted because absent or insufficient legal basis, within the meaning of Articles 6(1) and 9(2) GDPR, has been the single most cited violation in monetary fines imposed on controllers by supervisory authorities.[17] This highlights the significance of challenges associated with ensuring that each controller participating in the HRIC has a valid GDPR legal basis in relation to data processing operations for which a particular controller is responsible. However, the complex interdependencies among controllers in terms of

[15] The rights afforded to data subjects under the GDPR are: "Right of information" (Arts. 12-14 GDPR), "Right of access" (Art. 15), "Right to rectification" (Art. 16), "Right to erasure" (Art. 17), "Right to restriction of processing" (Art. 18), "Right to data portability" (Art. 20), and "Right to object" (Art. 21). Crucially, whether and to what extent these rights apply depends, among other factors, on the GDPR Legal basis used by the controller and the applicable national laws in the controller's country.

[16] See, for example, Narayanasamy et al(2020): https://www.frontiersin.org/articles/10.3389/fgene.2020.00303/full and Mladinić et al (2021): https://hrcak.srce.hr/clanak/383527

[17] Saemann et al. 2022: https://petsymposium.org/popets/2022/popets-2022-0111.pdf

demonstrating their GDPR compliance can be even better highlighted using other Art. 5(1) GDPR principles.

Consider, for example, the closely related principles of **purpose limitation** (Art. 5(1)(b)), **data minimisation** (Art. 5(1)(c)), and **storage limitation** (art. 5.(1)(e)). Collectively, these principles require the controller to demonstrate that it has a valid purpose for processing personal data, and that the nature of the processing is proportionate to the purpose being pursued. In other words, controllers must be able to demonstrate that processing is limited to what is strictly necessary for achieving the purpose. In the context of scientific research, a "valid purpose", within the meaning of the GDPR, is to be interpreted as a specific research project pursued by a researcher.[18] In the case of the HRIC-facilitated data reuse lifecycle, this seemingly places the onus of compliance exclusively on the CTU, the party that must clearly define its research question(s), specify the subsets of datasets required for this project, and describe a detailed data analysis plan.

However, a more holistic examination of the role of controllers in the HRIC ecosystem makes it clear that the upstream controllers' ability to meet their own GDPR compliance obligations under the same principles is inextricably linked to the CTU's compliance. For example, consider the CTD's perspective: after the CTU has defined its purpose(s) of processing personal data in a sufficiently detailed manner that complies with the GDPR requirements for purpose limitation, data minimisation, and storage limitation, it is the responsibility of the CTD to ensure that the permission to access and/or analyse personal data is granted in a manner that follows CTU's request specifications. This may include granting permission to only process relevant sub-sets of a dataset, and for a defined duration only. In the absence of these specifications by the CTU, and their adherence by the CTD, the CTD won't be able to achieve and demonstrate its own compliance with the relevant GDPR principles. For example, the CTD won't have the means to ascertain that data disclosure[19] to the CTU was done in a manner that ensures compliance with the GDPR principles of purpose limitation, data minimisation, and storage limitation. However, the compliance interdependencies are not limited to the CTD – CTU pair of controllers. The ability of the CTD to demonstrate that data disclosure to CTUs is done in a compliant manner directly influences whether it is permissible for the CTA to make the data available (i.e., discoverable and requestable by prospective CTUs)

[18] Becker et al. 2023 (forthcoming)

[19] "Data disclosure" is to be understood broadly. Depending on the architectural pattern of the technical infrastructure used by the CTD, this can be data transfer ("data release") to the CTU, granting access within an SPE ("data visiting"), or permitting the CTU to initiate a remote analysis without actually accessing the data ("model-to-data" federated/distributed approaches). (Terms in quotation marks correspond to SPE types in HealthyCloud D5.1)

via the CTD.[20] Importantly, the requirements the CTD must follow when evaluating data access requests from CTUs for their eligibility won't be solely dictated by the GDPR. Some of them will be additional requirements communicated by the CTA itself, based, for example, on the known preferences and objections of data subjects (i.e., patients and/or medical research participants) as expressed in informed consent forms.[21] Increasingly, when asked to provide informed consent for future research uses of their biological samples and associated data, prospective research participants are given options whereby they can opt in or out of certain categories of research, such as research activities pursued by commercial entities, or uses falling under purposes that conflicts with participants' personal values and preferences. Such expressed preferences subsequently become part of data access conditions and restrictions (also known as "ELSI Metadata"), which must be respected throughout the data lifecycle.[22] There could be additional sources of data use conditions and restrictions applicable to the dataset contributed by the CTA to the HRIC. It is crucial that the CTD has in place a well-defined data access governance framework which includes processes, workflows, and safeguards ensuring that compliance obligations are met. On the one hand, the CTD must implement and operationalise data use conditions and restrictions prescribed by the CTA. On the other hand, the CTD must ensure that requests to access and/or use the data by the CTU are formulated such that they respect GDPR compliance obligations incumbent upon the CTU, including under the principles of purpose limitation, data minimisation, and storage limitation.

To further emphasise the compliance interdependencies among the controllers, it is worth highlighting the role each of the four controllers plays with respect to ensuring that data access and use conditions or restrictions are respected throughout the data lifecycle in the HRIC:

– The CTI is the party that defines most of the additional (i.e., beyond legally required) use conditions/restrictions, such as those based on the research

[20] The substance of this statement also applies in situations where the same entity seeks to assume the roles of both the CTA and the CTD. The entity can only make its data available (i.e., discoverable and requestable) in the HRIC if the entity has the means to ensure and demonstrate that the subsequent data disclosures to CTUs will be done in a GDPR-compliant manner. In other words, the entity must be able to demonstrate its compliance with the GDPR under both capacities.

[21] "Informed consent", a standard ethical (and in some cases legal) requirement for biomedical research, should be distinguished from consent in the sense of the GDPR, i.e., consent as the legal basis for processing personal data (Art. 6(1)(a) and Art. 9(2)(a) GDPR). Consent under the GDPR concerns personal data only, must be obtained for a specific purpose, and entail an affirmative action (that is, only an opt-in consent is valid under the GDPR). By contrast, an informed consent for medical research concerns all aspects of the research (data and sample use, as well as medical interventions), is usually broader in scope and, depending on the jurisdiction, can be either opt-in or opt-out."

[22] E.g., Cabili et al. 2021: https://www.sciencedirect.com/science/article/pii/S2666979X21000380; Dyke et al. 2022: https://link.springer.com/article/10.1007/s12021-022-09577-4

participants' expressed values and preferences. Digital consent enables this.[23]

- The CTA translates these conditions/restrictions (including conditions or restrictions reflecting the research participants' choices captured in the digital consent, as applicable) into dataset description following a structured ontology (i.e., ELSI Metadata) that will be associated with the data throughout its lifecycle in the HRIC including the consent itself in these conditions that are part of this ELSI metadata associated with the dataset. Doing so enables an accurate and structured representation of the data use conditions/restrictions (including, via the HRIC FAIR Data Portal defined in D6.2), further enhancing meaningful discoverability of the dataset by prospective CTUs.[24]
- The CTD is responsible for ensuring that each subsequent data disclosure to a CTU is done in a manner that respects the data use conditions/restrictions communicated by the CTA. This responsibility of the CTD applies irrespective of whether the CTD itself operates the supporting technical infrastructure or engages a processor (such as the provider of the federated HRIC Computational Infrastructure outlined in D5.3) to that end.
- The CTU must formulate its data access request such that it meets the additional use conditions/restrictions associated with the data. For example, the CTU may need to attach a research ethics approval, and/or a completed DPIA when applying for the use a particular dataset, or – in some cases – modify its research project so that it meets the additional data use conditions/restrictions.

4.2. Fragmentation of upstream controllership in the HRIC-facilitated data reuse lifecycle

The analysis above highlights the complex compliance-related interdependencies among controllers involved in the HRIC ecosystem. This creates major challenges in terms of tracking and demonstrating controllers' compliance as the chain of processing gets longer. Therefore, from the point of view of the overall HRIC compliance, it is preferable to minimise the number of controllers involved in the HRIC data lifecycle. Ideally, the three sub-categories of controllers responsible for the upstream processing of personal data throughout the HRIC, (the CTI, CTA, and CTD) would be assumed by the same legal entity. While the totality of substantive GDPR and other compliance obligations incumbent upon controllers would remain essentially the same, this would reduce the number of necessary contractual

[23] E .g., Parra-Calderón et al. 2018: https://link.springer.com/article/10.1007/s12687-017-0355-z

[24] In other words, enabling a prospective CTU to not only discover what data collection exists, but also to assess whether the collection can be lawfully used in the specific research project pursued by the CTU.

agreements and make tracking compliance, at the global HRIC level, considerably easier.

However, from the narrow perspective of the sole upstream controller assuming the three roles, this would be less desirable, as the entity would now be required to demonstrate its compliance with three different sets of legal obligations. As the CTI, the entity would be responsible for, among other things, directly interfacing with the data subjects, informing them about further processing of the data by each CTU (as required under Art. 13 GDPR), and supporting the data subjects in exercising their rights, if/as needed. In its capacity as the CTA, the entity would also need to undertake additional data processing (e.g., data cleaning, transformation, making the data research-ready for CTUs), as well as defining a comprehensive set of data access and use conditions (the ELSI Metadata) to be associated with the dataset. Finally, as the CTD, the same entity would be responsible for verifying and demonstrating that: i) all applicable use conditions are respected by each subsequent CTU; and ii) the CTU has met the legal obligations incumbent upon the CTU (e.g., correctly specifying purposes of processing in order to comply with purpose limitation, data minimisation and storage limitation principles, among others; as well as demonstrating that the CTU has a valid legal basis under the GDPR). Owing to these increased compliance burdens, there is a strong incentive for the institution acting as the controller for the initial data collection to limit its controllership to, at the most, a dual role of the CTI and CTA. As also highlighted in the review of the governance models of European Data Hubs by WP4 (reported in D4.1.), in practice, the role of the CTD is often delegated to the Data Hub and/or the Data Hub's associated central Data Access Committee (DAC), with the medical institution providing the data to the Data Hub acting as the CTA (or CTI and CTA). However, in order to demonstrate that its participation in the HRIC is legally compliant, the institution would require contractual assurances and demonstrable compliance from the subsequent controller (e.g., the CTD), resulting in the fragmentation of the upstream controllership.

Apart from this inherent incentive among medical institutions to reduce their own controllership-associated GDPR compliance obligations, there are other contextual as well as external reasons leading to the fragmentation of upstream controllership in the HRIC, as briefly discussed below.

If there is only one entity acting as the "Upstream Controller" in the HRIC data lifecycle, that entity must have a valid GDPR legal basis for the processing carried out under all three capacities: for data collection, data availability, and data disclosure. Owing to the uneven availability of the GDPR legal bases across the Member States and types of legal entities, currently, few medical institutions would satisfy this condition. It's more likely that the controller that initially collected the data for own purposes (thereby acting as the CTI) does not have a valid legal basis to either assume the role of the CTA, - making the data available (i.e., discoverable and requestable) via the HRIC – and/or act as the CTD, granting downstream controllers (CTUs) permission to reuse the data for purposes pursued by CTUs. Under these circumstances, "upstream controllership" will be necessarily fragmented between at least two entities, one acting as the CTI/CTA, and the other

as the CTD. Inclusion of the data in the HRIC is only possible if the CTI (or CTI/CTA, as applicable) finds a way to lawfully provide the data to another entity that can then lawfully act as the controller for the next phase of processing (i.e., the CTA or CTD, as applicable) in the HRIC-facilitated data reuse lifecycle.

Additionally, fragmentation of upstream controllership may have occurred before data inclusion in the HRIC is considered. For example, it may be that the entity intending to make the data available in the HRIC (i.e., become the CTA) was not responsible for data collection from the data subjects. Instead, it received the data from another entity, in the context of a research project pursued by the CTA. These situations commonly arise as part of research projects aimed at the reuse of existing datasets, where the modality of data-sharing between parties follows the traditional bilateral data transfers. In this case, the fragmentation of upstream controllership has already taken place between the CTI and CTA, and – provided that the dataset currently held by the CTA is to be processed in the HRIC – cannot be reversed. Crucially, even though the CTI does not directly participate in the operational data flows in the HRIC, the CTI remains involved in the processing chain for GDPR compliance reasons. Namely, the CTI remains the primary contact point vis-a-vis data subjects and retains its responsibilities associated with supporting data subjects in exercising their rights under the GDPR (unless explicitly exempted from these obligations under a relevant Member State or Union law). Another common form of fragmentation of upstream controllership occurs between the CTA and the CTD: as also highlighted in D4.1., it is not uncommon for medical institutions to transfer their data to dedicated Data Hubs, from which point the Data Hub, through its central DAC, becomes the controller for subsequent data disclosures to CTUs.

Where, owing to the reasons described above, the fragmentation of upstream controllership does occur, mapping out and demonstrating compliance with each controllers' compliance obligations throughout the HRIC-facilitated data reuse lifecycle becomes a significant challenge. As the controllers operating under different legal frameworks may be subject to diverging GDPR requirements,[25] ensuring the controllers' compliance at the global, HRIC level, - unless coordinated centrally -, is extremely difficult to understand, track, and enforce. Having in place a well-resourced, central compliance support service, as part of the HRIC Legal/Regulatory Pillar being defined by WP2, will be a crucial instrument in this respect.

[25] For example, some controllers may be exempt from their Chapter III GDPR obligations vis-à-vis data subjects, particularly in the context of scientific research, where their GDPR legal basis is performance of a task in the public interest (i.e., Art. 6(1)(e) in conjunction with Art. 9(2)(j) GDPR). Other controllers, however, could be required to comply with additional obligations laid down in their national laws implementing the GDPR, as permitted under Art. 9(4) GDPR].

5. Compliance obligations in the HRIC: Processors

Under the GDPR, a processor is the party that processes personal data on behalf of the controller (Art. 4(8) GDPR). The compliance obligations applicable to processors under the GDPR differ from those of controllers. Most notably, processors, unlike controllers, are not required to demonstrate their compliance with the Art. 5 principles, or the Art. 25 DPbDD framework GDPR. Instead, the onus of demonstrating GDPR compliance, in the sense of the Accountability Principle, remains with the controller that engages the processor. As a consequence, it is the controller's responsibility to "use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject" (Art. 28(1) GDPR). Although the processor's obligations, particularly vis-à-vis the controller are non-trivial, considering the interpretive uncertainty of the controller's obligations under the GDPR, this allocation of responsibilities benefits processors by providing them with legal clarity. These benefits are particularly tangible in complex, multi-phased, and cross-border data processing chains envisaged under the HRIC, where the legal uncertainties inherent to the GDPR give rise to controllers' non-compliance risks which are difficult to fully understand, let alone eliminate.

As the relationship between the controller and the processor is governed by a contractual agreement in accordance with Art. 28(3) GDPR, the processor should receive its formal instructions from the controller regarding the processing of personal data. By demonstrating that it has acted based on the controller's instructions, the processor will typically be able to fulfil its GDPR obligations in relation to the data processing activities within the scope of the agreement.[26]

In order to leverage these advantages and to maximise the legal certainty for the parties operating infrastructural components of the HRIC, it is preferable that such parties generally act as processors engaged by relevant controllers (CTI, CTA, CTD, and/or CTU, as applicable). Encouragingly, this is currently the case under most scenarios envisaged for the two key infrastructural components of the HRIC proposed by the HealthyCloud consortium: the HRIC (federated) Computational Infrastructure (HRIC-CI) component defined by WP5, and the HRIC FAIR Data Portal described by WP6. Of note, it is foreseeable that, under certain circumstances, the providers of these HRIC components could act as controllers with respect to the special categories of personal data undergoing processing in the HRIC. [27] However,

[26] Assuming that: i) the agreement is valid in the sense that the GDPR roles (controller/processor) were correctly defined; and ii) entering into the agreement did not constitute a breach of an obligation incumbent upon the processor.

[27] For example, the HRIC-CI service provider may be required to perform a particular processing operation for own purposes using the personal data of patients/research participants. This may include data preservation and/or creation of audit logs in order to comply with a legal obligation

such cases would be rare, while the providers' role as a/the controller would be limited to specific processing operations with a narrowly defined purpose. Hence, the GDPR obligations associated with controllership will be correspondingly limited for these providers. On the other hand, with respect to most processing operations, the providers of the HRIC-CI and the FAIR Data Portal services will be processors. These considerations make it relatively straightforward to ensure the compliance of the providers operating these infrastructural components of the HRIC.

The most important GDPR compliance obligations for processors are summarised below. The section thereafter further transposes these requirements into recommendations aimed at the specific infrastructural building blocks of the HRIC.

5.1. General Obligations

Establishing and maintaining an internal policy framework for data protection. Such a framework, at a minimum, should include: i) organisational controls aimed at ensuring logical access control to personal data of data subjects; ii) procedures concerning handling and reporting of data breaches; iii) a consistent methodology for identifying, evaluating, and addressing privacy risks. The latter element could also be incorporated into the organisation's overall risk management strategy.

Training personnel on data protection and policy framework. Considering the organisation's role as a processor, particularly in relation to special categories of personal data, it is important to ensure that the relevant personnel employed by the organisation are intimately familiar with the GDPR as well as other applicable privacy and data protection laws. At a minimum, such a training should be aimed at the organisation's internal Data Protection Officer (if applicable), other supporting data privacy officers, data stewards, as well as the personnel routinely coming in contact with the personal data, even if in a robustly anonymised form. This training could be provided through the HRIC Legal/Regulatory Pillar.

Establishing suitable safeguards and other technical measures aimed at reducing the risks of data breaches. Such safeguards have been mapped out and described in various HealthyCloud deliverables, including D2.1 and, more recently, D5.4. The processor deciding on implementing a particular safeguard or technical measure should pay particular attention to documenting the decision-making process, explaining why the selected measures were deemed adequate. This will be highly relevant for the controllers engaging the processor towards ensuring and demonstrating their compliance with the GDPR.

incumbent upon the HRIC-CI provider. In this case, the HRIC-CI service provider would become the controller for the relevant specific processing operation(s). Likewise, there is a scenario where the provider responsible for the HRIC Fair Data Portal service could become a (joint) controller with respect to facilitating data access/disclosure between the CTD and the CTU (See Scenario S6.b in D6.3 for more details).

Maintaining a Record of processing activities (ROPA) under the processor's responsibility. Importantly, the ROPA maintained by the processor (Art. 30(2) GDPR) need not be as detailed as the controller's ROPA (Art. 30(1) GDPR). Nevertheless, in order to simplify the controller's GDPR compliance, the processor should generally aim for a granular description. The processor should be particularly attentive to address compliance documentation concerning data transfers to third countries or international organisations (if applicable), and incorporate the relevant technical and organisational measures employed by the processors (Art. 30(2)(c-d) GDPR). For instance, as noted in D2.1 Section 5, when processing personal data on a cloud infrastructure operated by US providers, regardless of the location of the servers, data could still possibly end up in the hands of American authorities because of extraterritorial application of US laws. Capturing any residual privacy risks stemming from the nature of the applicable legal framework is important for ensuring the controllers have a complete view on the risks associated with data processing.

Entering in a Data Processing Agreement with the controller, in accordance with Art 28(3) GDPR. The agreement should include several mandatory content elements, listed under Art. 28(3)(a-h). In order to enable the processor to effectively abide by contractual obligations listed under this article, the processor requires certain technical capabilities as well as expertise. For example, the processor must, under certain circumstances, support the controller in the pursuit of enabling data subjects to exercise their rights under Chapter III GDPR. Similarly, upon request, the processor should support the controller in achieving and demonstrating its other GDPR obligations, including, for example, carrying out a DPIA (Art. 28(3)(f) GDPR). The list of the processor's contractual obligations under Art. 28(3) should be used by the processor to perform a gap analysis, identifying potential missing technical capabilities and expertise. Addressing the identified gaps by subsequently building these capabilities, particularly in terms of legal expertise (e.g., expertise required for assisting the controller in performing a DPIA), can be supported by the proposed HRIC Legal/Regulatory Pillar.

5.2. ELSI-Compliant Governance of the HRIC: Recommendations for the providers of key Infrastructural Components

This section covers the recommendations concerning the key output of the HealthyCloud consortium, including the main infrastructural components of the HRIC: the HRIC (federated) Computational Infrastructure (HRIC-CI) service, and the HRIC FAIR Data Portal.

The federated HRIC-CI service, being defined by WP5, is the central component of the technical infrastructure supporting data flows within the HRIC. The HRIC-CI will be designed to connect upstream controller(s), which could be Data Hubs, Health Data Collections, Data Producers, Data Providers, and/or Infrastructure

Providers[28] (as applicable and required), with the data (re-)users, i.e., CTUs. As highlighted in D5.3., "Computational Infrastructures [should] be built in such a way that researchers are enabled to comply with existing regulatory and ethical requirements more easily". This consideration is indeed crucial to the design and implementation of a HRIC-CI that provides meaningful benefits to the broader HRIC community.

Collectively, the relevant previous HealthyCloud deliverables (D2.1, D2.2, D5.3) have already successfully mapped out most of the key governance requirements for the HRIC-CI service. These efforts have significantly benefited from the study of the existing computational infrastructures, including SPEs (reported in D 5.1), which has been further supplemented by preliminary insights and observations eventually reported in D5.4. As such, the high-level guidelines and recommendations formulated in D5.3. are substantially complete. They can be used, in their current form, to inform the design of the HRIC-CI backbone. This section of the present deliverable D2.4. essentially complements D5.3., offering additional insights and points to consider when translating the D5.3. guidance into concrete design choices as part of implementing the HRIC-CI service.

Although generally not subject to the GDPR obligations associated with controllership, the provider(s) of the HRIC-CI services should nevertheless approach the design of their technical and organisation measures (TOMs), including safeguards, from the point of view of the controller's requirements. Doing so will allow the providers to not only support controllers towards demonstrating their GDPR compliance, but also increase the controllers' confidence in the HRIC-CI services, thereby ensuring that HRIC-CI is widely utilised by the target community. In this regard, it is recommended to consider research and development for applying homomorphic encryption techniques as it facilitates situations where calculations are performed by parties who are denied plain text access to sensitive data.

At the same time, however, it is important to recognise that there are inherent limitations to the extent to which the HRIC-CI service providers can support controllers in demonstrating their compliance. For example, in the context of controllers' compliance with Art. 5(1)(a) GDPR, providers of HRIC-CI service have no capacity to ascertain and help demonstrate that the CTU has a valid GDPR legal basis to process the data being disclosed by the CTD, as doing so would require an in-depth understanding of both the CTD's and the CTU's national legal frameworks. Similarly, unless explicitly instructed by a relevant controller to perform a particular

[28] For the definitions of these terms, see Annex I.

action, there are few steps HRIC-CI service providers can proactively undertake to demonstrate that the processing of personal data in the HRIC-CI meets the general transparency and fairness obligations. It is incumbent upon the relevant controller to correctly interpret its compliance obligations under, among others, Art. 5(1)(a) GDPR and, to the extent applicable, instruct processors such as HRIC-CI service providers to undertake certain actions. Supporting controllers along the lines of broader GDPR compliance would fall under a separate legal compliance support service, which is part of the proposed HRIC Legal/Regulatory Pillar defined by WP2.

On the other hand, there are various other areas of controllers' GDPR compliance where the providers of the HRIC-CI services could undertake rational design choices to meaningfully support controllers' compliance efforts. These aspects have already been outlined in the aforementioned HealthyCloud project deliverables; they are further elaborated in a more nuanced manner below, within the context of controllers' GDPR compliance.

The architectural pattern of the HRIC-CI backbone. Early on in the HealthyCloud project, WP5 explored various architectural patterns, including topological configurations, of infrastructures supporting data sharing and/or analysis. Those architectural patterns varied from the simple "data release" approaches, whereby the data-providing controller (i.e., the CTD) essentially transfers the entire dataset to the recipient controller (i.e., the CTU), to much more sophisticated, layered, and multifaceted design blueprints (see D5.1). Following this comprehensive landscape assessment, WP5 arrived at the preliminary conclusion that federated architectures offer greatest benefits in terms of privacy and security (D5.3).

The conclusions of WP5 are encouraging, as federated models of data analysis provide significant benefits also in the context of broader GDPR compliance, from the standpoint of both upstream and downstream controllers. The overall compliance with the GDPR would be the easiest to demonstrate by the controllers where the architecture follows the model described in the sub-section 4.5.4.3 of Deliverable 5.1: "Distributed Compute SPEs: Compute-to-data". Under this scenario, the CTU is not directly granted access to the record-level data held by the CTDs, but instead is allowed by the CTD(s) to bring its data analysis algorithm to the HRIC-CI, subjecting the data to the analysis in this manner. Importantly, reliance on the "compute-to-data" approach of federated analysis, while preferable, does not automatically guarantee GDPR compliance. As also correctly noted in D5.1, "a clear assessment of the risks of using this approach in the context of the specific project must still be conducted". For example, in the broader context of GDPR compliance, it could be that the prospective CTU did not describe its research project in a manner that meets the GDPR principles of purpose limitation and data minimisation. Such a request, if approved by the CTD without significant changes,

would still result in the processing of personal data in a manner that fails to comply with the GDPR. As such, this approach does not eliminate the responsibility of CTDs to review each federated analysis request submitted by a prospective CTD. "Compute-to-data" architectural patterns, however, do make it easier to demonstrate that after a permission to perform analysis has been granted in a GDPR-compliant manner, the analysis carried out in the HRIC-CI was indeed carried out as agreed upon between the CTD and the CTU.

This is, however, not to say that "compute-to-data" is the only modality of the HRIC-CI that would enable controllers to demonstrate their GDPR compliance. In principle, also design choices that allow the CTU to view or even download record-level personal data could be implemented in a manner that make it possible for controllers to demonstrate their GDPR compliance. However, under these circumstances, the providers of the HRIC-CI services would need to adopt additional safeguards to ensure that the relevant principles of the GDPR are met, most pertinently the principles of Purpose Limitation (Art. 5(1)(b)) Data Minimisation (Art. 5(1)(c)) and, especially if data download is allowed, Storage Limitation (Art. 5(1)(e)). This can be accomplished by rational deployment of suitable technical safeguards, supplemented with appropriate contractual tools. For example, the HRIC-CI service provider could implement a data partitioning/segregation capability, whereby data users (CTUs) are granted access to the relevant sub-set of the data, as opposed to an entire dataset. In a similar vein, data access approaches enabling CTUs to freely explore the available data at the record level, in the absence of a well-defined research question (and pre-approved by the CTD), should be avoided. Moreover, in the context of controlled-access modalities outlined in D5.1, the providers of the HRIC-CI services should adopt suitable user monitoring technologies that record user behaviour within the secure processing environment. Additionally, appropriate contractual agreements, such as a data use agreement binding on the CTU, should be signed, whereby the CTU commits to restricting data access to its personnel on a need-to-know basis, while mandating concrete actions whereby the CTU can demonstrate its compliance with, for example, the purpose limitation and the storage limitation principles of the GDPR.

With respect to the GDPR principle of Integrity and Confidentiality (Art. 5(1)(f) GDPR), alongside the relevant requirements incumbent upon the controllers under the DPbDD framework (Art. 25 GDPR), WP5 has identified various solutions that would be helpful in enabling controllers' compliance. For example, the user Authentication and authorisation infrastructure (AAI) solutions described in D5.3 are robust and fit for purpose. Another important advantage of these AAI solutions is that they also minimise the extent to which the HRIC-CI processes the personal

data of the researchers themselves.[29] There are numerous additional steps incorporated into WP5 guidance, including commitment to implementation the "5 Safes" framework, that should further strengthen the security of processing special categories of personal data via the HRIC-CI.

There is, additionally, a clear link between the data storage/retention processes and capabilities recommended by WP5 on the one hand, and the relevant controllers' ability to ensure and demonstrate their compliance with the storage limitation principle. For example, D5.3. states that a "User data [must] be removed from a compute node and operational storage after a project ends to avoid data leakage or misuse. Data [should] be archived in appropriate archives for FAIR usage". More broadly, it is likely that the mixture of "persistent" and "ephemeral" data storage capabilities described under section 2.2 of D5.3 will be adequate for HRIC-CI providers to support controllers in demonstrating their compliance with the GDPR's storage limitation principle under various scenarios.

One more important set of technical capabilities described by WP5 concerns the potential adoption of a structured ELSI Metadata model that would enable at least partially automatable data processing workflows whereby the processing of personal data respects the data access and use conditions defined by the relevant upstream controller (e.g., the CTA). Not only would this be invaluable in the context of demonstrating (upstream) controllers' compliance with the GDPR's purpose limitation principle, but it would also ensure that additional restrictions and/or conditions – including, those imposed by data subjects themselves – are respected. The specific ELSI Metadata model explored by WP5 makes use of the structure proposed by the GA4GH in their Data Use Ontology (DUO) (See section 1.3 in D5.3). Recommending the adoption of a particular ELSI Metadata structure is beyond the scope of the present deliverable D2.4. It is nevertheless important to emphasise the crucial value of implementing a suitable ELSI metadata structure, capturing the relevant data access and use conditions. Beyond its role in demonstrating (upstream) controllers' compliance, - GDPR-related or otherwise -, adoption of a comprehensive ELSI Metadata model would also be invaluable for enabling meaningful discoverability of the HRIC data collections by prospective downstream controllers (CTUs). This can be done by ensuring that the data use conditions and restrictions are clearly reflected in the description of the datasets as displayed via the HRIC FAIR Data Portal (Deliverables 6.2).

[29] NB: With respect to processing the personal data relating to the users of the HRIC-CI, the provider(s) of the HRIC-CI services will be controllers. However, this is relatively unproblematic, as the providers will have a valid GDPR legal basis to do so (presumably, Art. 6(1)(b) GDPR), while the scope of processing will be limited to the specific purpose of providing HRIC-CI services. Moreover, it is unlikely that special categories of personal data (i.e., "sensitive" personal data) will be processed by the HRIC-CI in this context.

In an ideal scenario, assuming full technical interoperability between the HRIC-CI and the HRIC Fair Data Portal, it would be feasible for prospective data users (i.e., prospective CTUs) to perform all of the following steps:

i) Discovering, through the HRIC FAIR Data Portal, what HRIC data collections exist, and, among these, which datasets are available and can be lawfully used in the specific research project pursued by the CTU;
ii) Requesting access to the relevant datasets via the HRIC FAIR Data Portal. These requests would be reviewed and, provided they meet the eligibility criteria, approved by the relevant CTD(s) (i.e., the DAC(s) of the relevant CTD(s) such as Data Providers, Data Hubs, Infrastructure Providers, as applicable)
iii) Following an approval, initiating data analysis (federated or otherwise, as applicable) through the HRIC-CI. Assuming technical interoperability between the HRIC FAIR Data Portal and the HRIC-CI, the user journey between the two components would be smooth, without requiring the user to create different accounts for the two infrastructural components of the HRIC.

However, the author and the contributors of this deliverable are well-aware of the technical challenges associated with the adoption and, especially, operationalisation of ELSI Metadata models within automatic data processing workflows. As such, the ideal scenario described above should be seen as the aspirational goal for the infrastructural components supporting data flows in the HRIC ecosystem. Nevertheless, the concrete steps taken by the HealthyCloud consortium towards achieving this goal, including proposing the Metadata Standards and Data Interoperability Support Service as one of the HRIC services, highlights that the consortium members are well-placed to address ELSI Metadata-related challenges, albeit this would require continued collaboration beyond the lifetime of the HealthyCloud project. In order to achieve full technical interoperability in terms of ELSI Metadata, a close partnership will be necessary among the providers of: i) the HRIC-CI backbone; ii) the HRIC-CI FAIR Data Portal service; and iii) the proposed Metadata Standards and Data Interoperability Support Service.

One of the few technical aspects with a potential relevance to controllers' GDPR compliance not yet fully explored by the HealthyCloud consortium concerns infrastructural capabilities required for exercising data subjects' rights, in the sense of Chapter III GDPR. However, this should not be seen as a significant gap in the proposed design of the relevant HRIC infrastructural components under the envisaged federated data analysis architecture. Since in federated data analysis architectures there is limited or no long-term ("persistent") centralised data storage by the HRIC-CI, data subjects' rights would only apply in a limited way. For example, there could be rare cases where data subjects' rights need to be exercised during the short-term ("ephemeral") storage of the data on the HRIC-CI, and/or, during the

active data analysis phase, where the data are being processed by the user (CTU). It is recommended that WP2 and WP5 jointly explore the relevance and the applicability of GDPR Data Subject rights to the scenarios currently envisaged by WP5.

Finally, following the establishment of the HRIC infrastructural components, (including, but not limited to the HRIC-CI and the FAIR Data Portal services), the providers of these components should demonstrate clear links between their TOMs and the corresponding GDPR compliance obligations incumbent upon the relevant controller. This can be accomplished by performing a compliance Gap Analysis from the controller's point of view, which could potentially serve as a source of input for the controller's DPIA, if required.[30] However, owing to the broad scope of such a Gap Analysis, it is recommended that the providers operating HRIC Infrastructural components engage a competent external party to help with the provider's efforts. The compliance support services of the proposed HRIC Legal/Regulatory Pillar, if implemented, would be of significant assistance in this respect.

6. Conclusion

This deliverable has summarised ELSI governance and compliance challenges associated with the processing of personal data in the HRIC ecosystem, followed by recommendations for the design and implementation of ELSI-compliant HRIC governance, with a focus on the HRIC Infrastructural Components.

Overall, the providers of the key infrastructural components of the HRIC defined by the HealthyCloud consortium are well-positioned to meet their compliance obligations as processors in the sense of the GDPR. Several potential areas of improvement, such as implementing an ELSI Metadata standard, and considering capabilities for supporting data subjects in exercising their GDPR rights, have been proposed. These aspects will be further explored by WP2, in coordination with other relevant WPs, over the remaining months of the HealthyCloud project.

However, the principal compliance challenges to the implementation and, especially, widespread adoption of the HRIC lie beyond the specifications of the underlying technical infrastructures. Namely, these compliance challenges stem from the complex, fragmentary, and uncertain legal framework under which controllers operate. The challenges are further exacerbated by the inherent compliance interdependencies among the controllers involved at different steps

[30] An indicative example of such a Gap Analysis for cloud-based health organisations can be found in Georgiou and Lambrinoudakis (2020): https://www.mdpi.com/1999-5903/13/3/66

throughout the data lifecycle. These compliance interdependencies among controllers, as discussed in section 4, are of functional nature and cannot be fully resolved through contractual instruments and accountability tools. Identifying and adequately mitigating controllers' non-compliance risks in the context of the HRIC is resource-intensive and requires significant coordination at the central level. In this respect, the HRIC Legal/Regulatory Pillar proposed by the HealthyCloud consortium, if implemented, could be of significant added value.

Importantly, these conclusions have been formulated under the relevant European legal framework applicable as of May 2023. However, legislative discussions around privacy and data protection laws in Europe are evolving rapidly, meaning that the analysis and the conclusions presented in this deliverable may soon be outmoded. In particular, there are legislative changes on the horizon that could help bring the European HRIC closer to reality. Most pertinently, the forthcoming EHDS Regulation could facilitate streamlined access and reuse (i.e., secondary use) of personal health data, including for scientific research purposes. Not only does the EHDS Regulation seek to create valid GDPR legal bases for each controller involved in the data reuse lifecycle, but it also aims to stimulate the establishment of infrastructures, data-sharing intermediaries, as well as common data and metadata standards required for secondary use of data in a scalable manner. Whether the ambitious vision of the EHDS Regulation can be realised and in which timeframe remains to be seen. Additionally, the GDPR itself will undergo a comprehensive evaluation in 2024, in accordance with the procedure laid down in Art. 97 GDPR. Depending on the findings of this upcoming evaluation, the legislator could decide to undertake a revision of the Regulation.

It is the hope of HealthyCloud Consortium members that the European legislators will leverage these unique opportunities to create a more research-friendly legal framework in Europe. In doing so, they would facilitate the implementation and, perhaps more importantly, widespread utilisation of the HRIC and other data-sharing initiatives in Europe.

Annex I – Relevant Definitions (from the Glossary)

Data controller (or simply **Controller**) the party that, alone or jointly with others, determines the purposes and means of the processing of personal data. The actual processing may be delegated to another party, called the data processor. The controller is responsible for the lawfulness of the processing, for the protection of the data, and respecting the rights of the data subject. The controller is also the entity that receives requests from data subjects to exercise their rights.

Data processor (or simply **Processor):** a processor shall mean "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." The essential element is therefore that the processor only acts "on behalf of the controller" and thus only subject to his instructions.

In some cases, the processor may choose not to process the data himself, but may have recourse to a subcontractor who processes the data on his behalf. In practice, this will depend upon the processor agreement entered into with the controller.

Data producer: person or entity that generates health-related data either as product of their activity or as mandated by another individual or organization.

Data Provider A person or organisation (other than the data subject) who has the right to grant access to or to share certain data.

Data user: A natural or legal person/organisation who has lawful access to certain personal or non-personal data and is authorised to use that data for commercial or noncommercial purposes.

Health Data Hub:

Minimal inclusion criteria:

1. A digital technical infrastructure with the core mission of enabling health data sharing
2. It provides health data from different sources
3. It allows discovery of health datasets
4. It has a metadata discovery service
5. It has a data accessibility mechanism in accordance with existing regulation
6. It has an authorization functionality, provided by the same Data Hub or by an external institution.

Health Data Collection:

A technical infrastructure that holds datasets, makes datasets available for use, and organises data in a logical manner. The datasets may come from different sources, hospitals and/or research institutes from the same country (national data repositories) or different countries (international data repositories). Data collections may also cover appropriate, subject-specific locations where researchers can submit their data. Data collections may have specific requirements concerning subject or research domain; data reuse and access; file format and data structure; and the types of metadata that can be used.

Minimal inclusion criteria:

1. A digital platform that receives and stores data
2. It receives data from a single source and/or multiple sources

3. Allows discovery of the stored health data
4. It must have control over the data stored

Other possible characteristics of a data collection:

1. It could have a specific thematic, data type that it collects (e.g. a particular disease, a particular data type: genomic data, clinical data, EHRs…)
2. It could be part of one or more overarching data hubs
3. It could generate data

Infrastructure provider is the responsible organization to support the physical management of health-related data following existing regulations.

Secure processing environment (SPE): The physical or virtual environment and organisational means to provide the opportunity to re-use data in a manner that allows for the operator of the secure processing environment to determine and supervise all data processing actions, including to display, storage, download, export of the data and calculation of derivative data through computational algorithms.

Acronyms and Abbreviations

AAI - Authentication and Authorisation Infrastructure
CTA - Controller for making data available via the HRIC
CTD - Controller for data disclosure to a specific user
CTI - Controller for the initial data collection
CTU - Controller for the (re)use of the data
D – Deliverable
DAC- Data Access Committee
DoA - Description of Action (Annex 1 of the Grant Agreement)
DPA - Data Processing Agreement
DPbDD - Data Protection by Design and Default
DPIA - Data Protection Impact Assessment
DUO - Data Use Ontology (Developed by the Global Alliance for Genomics & Health; GA4GH)
EDPB - The European Data Protection Board
EHDS - European Health Data Space
ELSI - Ethical, Legal, and Societal Implications/Issues/Impacts
FAIR (principles) - Findable, Accessible, Interoperable, Reusable data
GDPR - The General Data Protection Regulation
HDAB - Health Data Access Body
HRIC - Health Research Information Cloud
HRIC-CI - HRIC Computational Infrastructure
ROPA - Record of Processing Activities
SPE - Secure Processing Environment
TOMs - Technical and Organisational Measures
WP - Work Package


HEALTHYCLOUD
Health Research & Innovation Cloud