Project Title	AI-based long-term health risk evaluation for driving behaviour change strategies in children and youth
Project Acronym	SmartCHANGE
Grant Agreement No.	101080965
Project Start Date:	1 May 2023
Project Duration:	48 months
Project Website:	https://www.smart-change.eu/

D2.3 – SELP Compliance Framework

Work Package	**2**
Lead Partner	VUB
Contributing Author(s)(Partner)	Renato Sabbadini, Paul Quinn, Dario Fenoglio (USI), Martin Gjoreski (USI, Marc Langheinrich (USI)
Due Date	2024.01.31
Date	2024.02.16
Version	V1.4

Dissemination Level

X	PU – Public, fully open
	SEN – Sensitive, limited under the conditions of the Grant Agreement
	Classified R-UE/EU-R – EU RESTRICTED under the Commission Decision No2015/444
	Classified C-UE/EU-C – EU CONFIDENTIAL under the Commission Decision No2015/444

Funded by the European Union

	Classified S-UE/EU-S – EU SECRET under the Commission Decision No2015/444

Abstract:	This document builds on deliverable D2.1 and summarizes findings of T2.3, dealing with specific regulatory issues emerged in relation to this project and their contextual application vis-à-vis the other work packages.
Keyword List:	Legal and ethical issues, SELP, Data protection, Privacy, GDPR, Artificial Intelligence Act, AI Act, Medical Device Regulation, EHDS, European Health Data Space, artificial intelligence
Licensing information:	This work is licensed under Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) http://creativecommons.org/licenses/by-sa/3.0/
Disclaimer:	This project (GA No. 101080965) has received funding from the Horizon Europe R&I programme. The information provided in this document reflects solely the author's views. The European Community, Agency, and Commission are not liable for any use that may be made of the information contained herein. The content is provided without any guarantee or warranty of fitness for a particular purpose. Users utilise the information at their own risk and liability. In the case of proprietary information of the SmartCHANGE Consortium, it shall not be used, duplicated, or communicated to third parties without prior consent.

Funded by the European Union

Versioning history

Version	*Date*	*Author(s)*	*Notes &/or Reason*
1.0	05/01/2024	Renato Sabbadini (VUB)	TOC and chapters 2
1.1	31/01/2024	Renato Sabbadini (VUB), Paul Quinn (VUB), Dario Fenoglio (USI), Martin Gjoreski (USI, Marc Langheinrich (USI)	Chapters 2, 3, 4, 5
1.2	12/02/2024	Renato Sabbadini (VUB), Paul Quinn (VUB), Dario Fenoglio (USI), Martin Gjoreski (USI, Marc Langheinrich (USI)	Ex. Sum, Chapters 1, 2, 3, 4, 5, 6, References, Appendixes
1.4	16/02/2024	Renato Sabbadini (VUB), Paul Quinn (VUB)	Integrations after review

Quality Control (peer & quality reviewing)

Version	*Date*	*Name (Organisation)*	*Role & Scope*
1.2	07/02/2024	Mitja Luštrek (JSI)	General review
1.2	12/02/2024	Fawad Taj (VUmc)	General review

Funded by
the European Union

Table of contents

Funded by the European Union

List of figures

List of tables

Funded by
the European Union

Executive summary

Following the completion of deliverable D2.1, i.e. the Benchmark of regulatory and ethical frameworks, this deliverable is meant to provide additional elements for the contextualised applicability of the frameworks described in D2.1 to the task of the Work Packages dealing with the creation of the risk-prediction model and the SmartCHANGE tool and with the proof-of-concept study. While societal issues will be left for a later stage, i.e. once sufficient information has been gathered from the participatory design sessions and the proof-of-concept study, and further ethical and legal requirements will be finetuned upon collection of the responses to the questionnaire at the basis of the SELP Impact assessment (D2.4). This deliverable deals with three main issues that have become more relevant and in need of attention at this stage of the life-cycle of the project: i) data protection, ii) the AI Act and iii) compliance with Article 5(5) and Annex I of the Medical Device Regulation.

As far as data protection is concerned, this document provides an answer as to how to deal with anonymous data in light of different legal interpretations of the notion. This document also explores the issue of risks to data protection in the context of data processing for machine learning purposes. Moreover, in light of the presence of two non-EU countries among the partners, one 'founding' partner of the consortium, i.e., USI from Switzerland, and a future partner, i.e., TMU from Taiwan, information on the data protection laws of these two countries, is provided.

In relation to the AI Act, an update on its legislative process is provided, highlighting the most pragmatic course of action for the SmartCHANGE consortium vis-à-vis this new piece of legislation.

As seen both in the original proposal of the project and the discussion of section 1.3 of D2.1, the SmartCHANGE tool does not aim yet at a full certification as a medical device at this stage, but will nevertheless comply with Article 5(5) and Annex I of the Medical Device Regulation to 'prepare the ground' in case partners decide towards the end of the project that the tool can be exploited after the completion of the project as a proper medical device.

List of abbreviations

Abbreviation	*Definition*
AI	Artificial Intelligence
AIA	AI Act
CJEU	Court of Justice of the European Union
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
EDPS	European Data Protection Supervisor
EU	European Union
GDPR	General Data Protection Regulation
LLM	Large Language Model
MDR	Medical Devices Regulation
MDCG	Medical Device Coordination Group
SELP	Societal, ethical, legal and privacy
SRB	Single Resolution Board
WP	Work package

Funded by
the European Union

1 Introduction

With the completion of deliverable D2.1 (Sabbadini et al. 2023), the main pieces of legislation affecting or likely to affect the SmartCHANGE project were introduced to the partners based on the three main areas at the centre of the project in terms of legal and ethical considerations, i.e. data protection, artificial intelligence and the 'aspiration' for the SmartCHANGE tool to be ready to undergo a medical device certification process if partners so decide at the end of the proof-of-concept study.

1.1 Purpose and scope

The original purpose of this deliverable has been adjusted as a result of the continuous dialogue among project partners in the months since the delivery of D2.1. Originally, this document should have been a framework for the contextual application of the norms laid out in the frameworks described in D2.1. Various considerations have led us to reframe this document. First, at least in relationship to the GDPR and the MDR, deliverable D2.1 already discussed in certain sections concrete elements for the specific application of the relevant norms to the project, as in the case, for instance, of the issue of whether or not partners intend to pursue a full MDR compliant certification of the SmartCHANGE tool or limit themselves to take advantage of the possibilities offered by Article 5(5) of the MDR (more on this in section 3 below). Second, ethical concerns and related norms and guidelines were addressed to the extent possible at this stage of the project in deliverables D2.1 and D2.2. Third, and related to the previous point, finer granularity in terms of contextual application of ethical and legal frameworks will be more realistically achievable once the contours of the SmartCHANGE tool and the proof-of-concept study will acquire greater definition over the course of the following months, including by means of the interactive questionnaire at the basis of task T2.4, i.e. the SELP impact assessment. Fourth, while the acronym SELP also refers to societal issues, we have matured the conviction that a proper analysis of this aspect will be possible only once a clearer understanding of the opinions and attitudes towards the SmartCHANGE tool among the various stakeholders will have emerged by means of the participatory-design related tasks of WP3 (User engagement) and those of WP7 (Proof-of-concept study) , these analyses of societal aspects will therefore feature in D2.5a (m24) in relation to what will emerged in WP3 and inD2.5b (m48) in relation to what will emerge in WP7.

While leaving further contextualised legal considerations for D2.4, i.e. once the questionnaires described in T2.4 will have been completed by the partners and once the

contours of the SmartCHANGE will have acquired more shape, the present deliverable D2.3 will therefore focus on three main topics, described in the section 1.3 below, together with an explanation as to their choice.

1.2 Contribution to other deliverables

The contribution to other deliverables can be derived from the table of correspondence between WP tasks and relevant legislation in section 4. While the table establishes a link between the various tasks and the main, relevant pieces of EU legislation, the interactive questionnaire of task T2.4 should expand the understanding of the role of tertiary legislation or 'soft law', including industry standards and guidelines, in defining the scope of actions leading to the creation and deployment of the SmartCHANGE tool.

1.3 Structure of the document

This document starts by analysing key areas of concern in relation to data protection, as this is the area that drew so far the most attention from the partners, particularly those involved in the processing of data for the creation of a risk-prediction model and those involved in the collection of data during the proof-of-concept study. The discussion deals first with the *vexata quaestio* of 'anonymous' data, as opposing legal interpretations differ on whether or not this kind of data is still personal or not. Then the attention goes to the risk of re-identification of data subjects when processing data through machine learning. Finally, a section outlines the principal aspects of data protection legislation in two pivotal non-EU countries relevant to this project:, i.e. Switzerland and Taiwan, for different reasons: in the former case, one of the key partner in terms of data processing necessary for the creation of the risk-prediction model is based in Switzerland (USI), in the latter case, a new partner from the island of Taiwan has been invited to join the project consortium.

The document will then provide a small update on the legislative process of the AI Act and its consequences for the SmartCHANGE project, followed by a section on the Medical Device Regulation providing elements for the consideration of the partners in addition to those already provided in D2.1.

After the table of correspondence described in 1.2 above and the conclusions, the document ends with four appendixes: two Data Transfer Agreements templates prepared for the partners who will receive data from third parties, one template for a Data Protection Impact Assessment and a letter from the new partner of Taiwan mentioned in section 1.6.

2 Data Protection

As the SmartCHANGE project partners working with personal data learned more about data protection, by way of a dedicated training held towards the end of month five and of section 1.2 of deliverable D2.1 (Sabbadini et al. 2023), and as the legal partner learned more about the use of personal data in the context of the project, the discussion within the consortium focussed on three areas, namely: i) anonymous data, ii) the risk to data protection when the data is used to train an AI model, and iii) the legislation of two non-EU countries involved in the project, i.e. Switzerland and Taiwan, the former being the country of USI, a partner since the beginning of the project, and the latter being the country of a new partner, Taiwan Medical University (TMU), joining the project in the near future.

2.1 Anonymous data

The data for the research work of the SmartCHANGE project falls under 3 categories:

1. Anonymous data acquired because of public availability (e.g. the nutritional dataset of the Centre for Disease Control, U.S.A.) or with permission (with or without payment) from third-parties, i.e. non-partners of the SmartCHANGE project;

2. Pseudonymous data acquired from project partners, e.g. ABCD (from VUMC), Young Finns (JAMK), etc.;

3. Data collected during the proof-of-concept study of SmartCHANGE.

As discussed also in section 1.2.2.1 of D2.1 (Sabbadini et al. 2023), the data in point 2. and 3. above are clearly personal – and sensitive – data, processed on the basis of Article 9(2)(j) (i.e. scientific research) in the first case, and collected and processed on the basis of Article 9(2)(a) (i.e. consent) in the second case. The data under 1. above are excluded from the scope of the General Data Protection Regulation (GDPR, Recital 26). At present there is much academic discussion and some important legal cases (ongoing concerning the limits of the concept of anonymity as outlined in the GDPR. The original (and previously considered authoritative) Article 29 Working Party Opinion on Anonymisation Techniques (WP29 2014) is very strict on what it considers ‘anonymous data’, stating:

> “[…] it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), **the resulting**

> **dataset is still personal data**." (Article 29 Working Party, Opinion 5/2014 on Anonymisation Techniques, p. 9, emphasis added)

Various scholars have struggled with the implications of Opinion 5/2014 (Dautlich et al. 2021, Finck/Pallas 2020, Groos/van Veen 2020 and Cruyt 2023). This is because its maximalist reading opens the door for an interpretation of the notion of 'personal data' that leaves very little room for the notion of 'anonymous data', and at the minimum leaves any controller or processor in a state of perennial legal uncertainty, because the status of the anonymous data should be regularly checked against new techniques of re-identification that might be developed (WP29 2014: 9, 24). Groos/van Veen (2020) rightfully question the soundness of the approach at the basis of Opinion 5/2014, as this would imply, for instance, that open data published by statical agencies would continue to remain personal data, as the microdata at its basis cannot and should not be deleted, also for reasons of reproducibility. In their words: "Nobody in their right mind would consider the highly aggregated data which we read in the news or scientific papers personal data because the data on which they are based have not been deleted at event level." (Groos/van Veen 2020: 3)

In other words, following the Article 29 Working Party opinion down this road would lead us to a situation where there would be increasingly few instances in the real world where one could speak of anonymous data. This clearly cannot have been the intention of the legislator when drafting the GDPR, i.e. a piece of legislation that follows the publication of the 2014 Opinion on anonymisation techniques by the Article 29 Working Party and that clearly refers to the existence of anonymous data in recital 26.

Moreover, in a very practical situation such as the one faced by SmartCHANGE partners receiving data declared anonymous by third-parties, it is not clear how the former would or should have the power, the authority and the means to ascertain whether or not the original, non-anonymised, dataset owned by the latter is still in existence. All the SmartCHANGE can do is to rely on the nature of the data as declared by the provider in the Data Transfer Agreement and to verify that the data does not contain elements the allow the re-identification of the data subjects behind the dataset.

Groos/van Veen 2020 rightly reminds us that in the end it is "a Court [that] decides about the interpretation of an Act and not the regulator, and in a more substantive sense, [...] the law should not have a scope of application which is infinite and can be arbitrarily executed" (Groos/van Veen 2020: 2).

From this perspective, then, the anonymous nature of a dataset should not be seen as an absolute property but rather as a relative (i.e. contextual) one, assessed by means of the double test created by the CJEU in the case of *Breyer v Bundesrepublik Deutschland* (CJEU, C-582/14), briefly presented in deliverable D2.1 (Sabbadini et al. 2023) in section 1.2.2.1. In the Breyer case the CJEU provides a two-step test to ascertain whether or not data can be re-identified by a controller, i.e.:

1. "For a controller who is not prohibited by law to identify, the data would be anonymous if the identification requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant.
2. For a controller who does not meet the first test, so the risk of identification is in reality not insignificant, the data would still be anonymous if identification either by that controller or with the help of a known third party was prohibited by law." (Groos/van Veen 2020: 5)

In the case of anonymous datasets acquired by SmartCHANGE partners, both legs of the test are fulfilled in the sense that, on the one hand de-anonymisation of the dataset would indeed require a disproportionate effort in terms of time, cost and human resources, while on the other hand, attempts to re-identify data subjects would go either against the law of the country providing the data, as stated by the entities providing the data themselves, or because they would violate a specific contractual obligation in which the data recipients have stated clearly that they would not attempt to re-identify the data subjects at the origin of the anonymous dataset (see Annex I and II).

2.1.1 Anonymous or pseudonymous?

In month 3 of the SmartCHANGE project VUB sent an initial Data Protection compliance questionnaire to the partners working with personal data. From the responses it emerged that several partners seemed to be using the terms 'anonymous' and 'pseudonymous' interchangeably, to the point of producing new verbs such as "pseudo-anonymise". In legal terms, of course, the distinction between 'anonymous' and 'pseudonymous' data is crucial vis-à-vis the GDPR, because the former is not considered personal data and falls therefore outside the scope of the regulation, while the latter is considered personal data and falls therefore within the scope of the regulation. In other words, as far as the GDPR is concerned, anonymous and pseudonymous are not synonymous. It would be unfair, however, to attribute the failure to recognise the difference between the two terms to lack of knowledge or understanding. The reality is that legal scholars and courts over the years have been trying

to find a way to translate in practical and actionable terms the difference between the two concepts, hence the attempt of Opinion 5/2014 to get rid of – so to speak – of the concept of anonymous data altogether, and the use by Groos/van Veen (2020) of the Breyer case to rebut Opinion 5/2014, although the kind of data at the centre of the case should be considered more pseudonymous than anonymous, as a third-party (the internet service provider) was in possession of the additional information necessary to re-identify the data subjects.

More recently, the General Court of the CJEU ruled on a case (T-557/20) that dealt with an "alleged breach of information obligations to data subjects when transferring their data to a third party" (Cruyt 2023) and saw the Single Resolution Board, i.e. the central resolution authority within the Banking Union, against the European Data Protection Supervisor, i.e. the independent supervisory authority that monitors and ensures that European Union institutions and bodies respect the right to privacy and data protection when they process personal data. The SRB performed a valuation of the Banco Popular Español, which included a procedure for affected parties – including employees of the bank – to make comments and sent the valuation after to the Deloitte company for an assessment. The personal data of the employees who made comments had been pseudonymised by SRB through an alphanumeric code and Deloitte did not possess the additional information to re-identify the data subjects.

Five employees of the bank made a complaint to the EDPS about this data transfer to a third party, i.e. Deloitte, and the EDPS eventually reprimanded the SRB, that objected however the reprimand, arguing that "although a key indeed existed, Deloitte had no way of accessing this key without breaking the law, nor was there any other practically feasible way to re-identify the data subjects, meaning that this transfer did not constitute a transfer of personal data." (Cruyt 2023). The General Court subsequently annulled the decision by EDPS and blamed the EDPS for failing to assess the possibility of Deloitte re-identifying the data subjects.

"With this decision, the Court confirms that these data could be anonymous data from the perspective of Deloitte, even though it was pseudonymized data from the perspective of (some employees of the) SRB. *The analysis whether information is pseudonymous or anonymous must depart from the perspective of the data holder.*" (Cruyt 2023, emphasis

added). The judgement by the General Court has been appealed[1] on 5th July 2023 by the EDPS on the ground that the Court did not consider the notion of 'pseudonymisation'.

At the centre of the problem of distinguishing between anonymous and pseudonymous data there is the issue of whether one adopts an absolutist, abstract, 'objective' approach or a relativist, concrete one. According to the former, no conceivable third party should be in possession of the additional information necessary to re-identify the data subjects for the data to be truly anonymous, but that, of course, would indeed require that the original dataset to be destroyed or anonymised also in the eyes of its originator, as the original dataset would represent not 'additional information' needed for re-identification, but *the* information itself. Moreover it has been remarked that in the era of Big Data, no one can ever be sure that a given dataset is effectively rendered anonymous, as a cross-checking with other datasets with new techniques might still enable the re-identification of the data subjects (Spiecker et al. 2023: 147). In the SRB v EDPS judgement, the General Court seems to have opted for a more relativist approach, where the capacity of the dataset holder to re-identify the data subjects must be the guiding principle.

The picture is further complicated by the fact that pseudonymisation is normally understood as involving procedures that simply remove direct identifiers, e.g. names and dates of birth, whereas anonymisation implies a greater intervention on the dataset, including noise introduction, elimination of outliers, and much more. But even so, the distinction between personal identified data, pseudonymised and anonymised data should be seen more in the context of adjacent categories on a continuum, as shown in figure 1 below (Dautlich et al. 2021: 20):

1 https://curia.europa.eu/juris/document/document.jsf?text=&docid=276483&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=323544

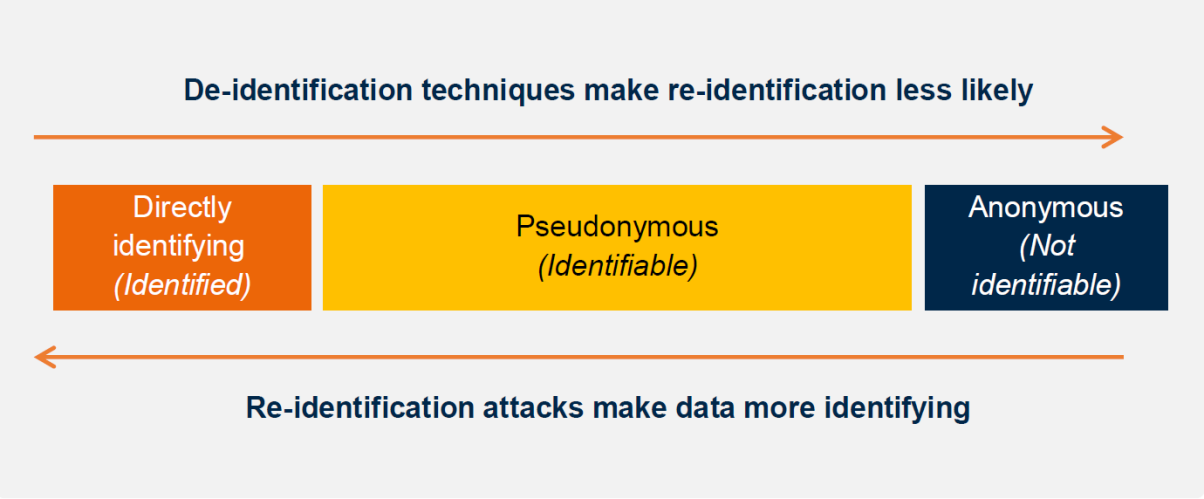


Figure 1: The identifiability spectrum (Dautlich et al. 2020: 20)

The difficulties highlighted so far are the consequence, no doubt, of the fact that technologies and techniques evolve at a faster pace for legislators and regulators to be able to adapt accordingly. We suspect this might be one of the reasons for the delay of the European Data Protection Board in issuing a new set of guidelines vis-à-vis anonymisation.

The question is: what is the most legally viable and pragmatically sensible approach for SmartCHANGE partners when processing data received from third parties? The answer to this question, we believe, relies on the consideration that behind any dataset received from a third party there is a Data Transfer Agreement, with the set of obligations both parties have committed themselves to. From this point of view, SmartCHANGE partners cannot but take at face value the anonymous or pseudonymous nature of a dataset as declared by the providing party, while exercising duty of care by asking the providing party to describe the process of pseudonymisation or anonymisation used and by verifying upon reception of the dataset that there are indeed no direct identifiers and by adopting technical and organisational measures in all cases, i.e. both for anonymous datasets and pseudonymous ones, to prevent access to the data by unwanted and potentially malicious third parties. This approach also reduces the risk that any personal data is inadvertently processed as non-personal data (i.e. without the protections that the data subject would require). Acting in this way will reduce the risks of harm to any individuals that may have had a link to the original data.

Funded by the European Union

Recommendation to partners

1. Have the provider of data specify whether the data is anonymous or not in the DTA, providing a description of the anonymisation technique used

2. Avoid any attempt of re-identification of data subjects of an anonymous or pseudonymous dataset and state this commitment in the DTA

3. Check the dataset upon reception for any element that might lead to re-identification and re-send to provider if any such element is found.

4. Technical and organisational measures as per GDPR Article 89(1) should be maintained for all datasets, regardless of their status vis-à-vis anonymity/pseudonymity.

Implementation:

The Data Transfer Agreements (see Appendixes 1 and 2) include clauses related to 1. and 2.

The entity storing anonymous or pseudonymous datasets received from third parties, situated in a partner's organisation, on behalf of all the other SmartCHANGE partners, should consult its own DPO as to the need of a Data Protection Impact Assessment (a model is provided in Annex III in case of a positive response from the DPO).

2.2 Machine learning

With the arrival of ChatGPT in November 2022, the public has gradually become more aware of the possibilities of AI in general and some of the risks connected to it vis-à-vis data protection, particularly in relation to the data used to train the model. While the study of risks connected to the protection of data used in machine learning is an evolving field,[2] articles about attacks (see Nasr et al. 2023) based on different kind of prompts, able to elicit a large

[2] We are grateful to Mykola Pechenizkiy (TUE), Martin Gjoreski (USI), Mitja Lustrek (JSI) and Andres Algaba (VUB) for the exchanges behind the content of this section. Any mistakes left are the sole responsibilities of the authors of this section.

Funded by the European Union

language model like ChatGPT to regurgitate personal data used to train it, have raised concerns as to how safe it is to use personal data when training an algorithm. Even before the arrival of ChatGPT scholars worried whether models trained on personal data should be considered personal data as well (Veale et al. 2018).

While this document cannot possibly delve into the intricacies of machine learning and what exactly happens to data used to train a model, it is important to note that data used in machine learning is not 'copied and pasted' as such into the model, but rather ran through it many times and used to compute the loss of ability of the model to predict target values.[3] The computation of the loss makes it possible to adjust the weights of the model. Nevertheless, if a malicious third party can access the weights of the model, these can be used to reconstruct in part the original training data, though never in a manner that provides more information than the training data. In other words, a model trained on a genuinely anonymous dataset cannot undergo a so-called reconstruction attack and thus provide information about the data subjects that was not present in the anonymous dataset.

It is also important to note that SmartCHANGE is not aiming at creating a large language model[4] (LLM), i.e. such as ChatGPT, but rather a risk-prediction model to be used in very specific contexts and trained, by comparison with an LLM, on significantly smaller set of data. The risk of Membership Inference Attacks, i.e. the kind of attack where a malicious third-party is attempting to assess whether or not a given sample of data was part of the dataset used to train an algorithm, is reduced considerably in part by eliminating outliers, e.g. a person whose height is 2,05 metres in a set of people with "normal" height, in part by isolating the model in a manner such that queries from the outside, i.e. by third parties, are not possible, and in part by ensuring that participants in the proof-of-concept study do not share publicly data that might be used to perform said attacks in the future.

Privacy and security evaluation of the AI processing pipelines[5]

The development of the AI-based processing pipelines in SmartCHANGE can be divided into three phases, as shown in Figure 2: Phase 1 – Data Exploration and Model Development; Phase 2 – SmartCHANGE Interinstitutional Collaboration. Phase 3 – Implementation & Federated Personalization (Pilot studies). Each phase has its respective threats and

[3] https://www.dp-institute.eu/wp-content/uploads/2024/02/privacycafe-Andres.pdf
[4] For more information related to privacy and LLMs see Brown et al. 2022.
[5] This section was written by: Dario Fenoglio, Dr. Martin Gjoreski, Prof. Marc Langheinrich, Università della Svizzera italiana, Lugano, Switzerland

Funded by the European Union

countermeasures, which are essential to ensure the project's integrity. In this document, we focus on the currently active project phases, i.e., **Subsection 1** explains Phase 1, and **Subsection 2** explains Phase 2. Phase 3 is scheduled for Month 21 in the project, and we will provide a more thorough overview of it in a follow-up document. In that document, our primary focus will be on privacy risks in federated learning scenarios, particularly those anticipated for SmartCHANGE. Nevertheless, as a precautionary step, in **Subsection 3,** we provide an explanation for why the final SmartCHANGE system offers greater security compared to a public AI-based system, such as ChatGPT.

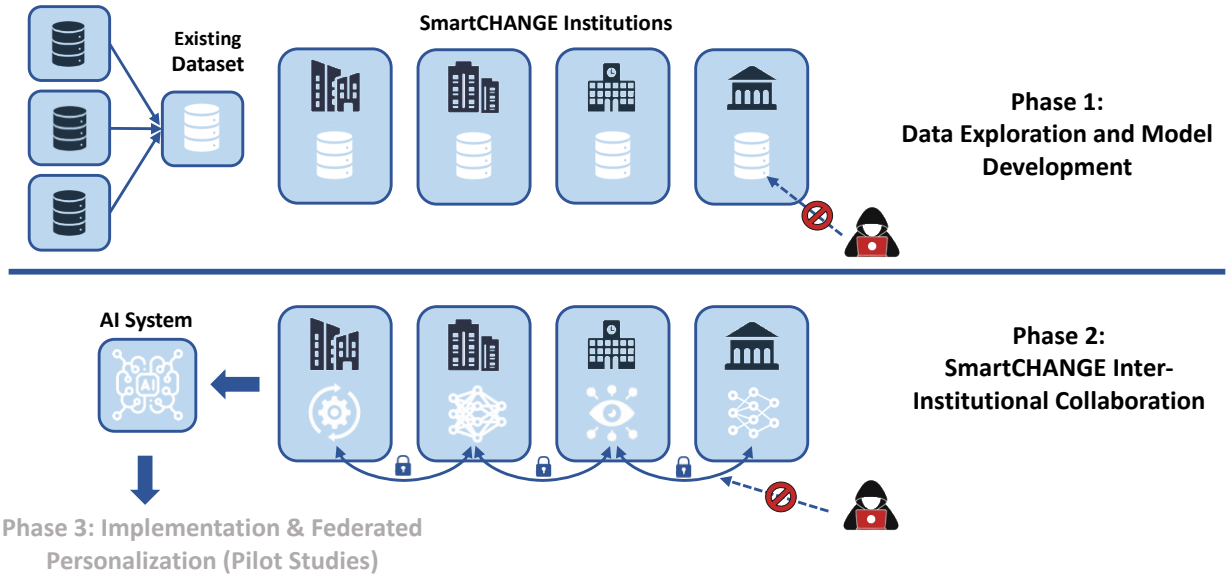


Figure 2: SmartCHANGE development phases. Phase 1: Data Exploration and Model Development. Phase 2: SmartCHANGE Interinstitutional Collaboration. Phase 3: Implementation & Federated Personalization (Pilot Studies).

Phase 1: Data Exploration and Model Development

The initial phase involves the collection and secure distribution of the existing dataset to each participating institution. These institutions are entrusted with the responsibility of maintaining the confidentiality of the data, ensuring no exposure to privacy threats by working on it exclusively within their local environments. This phase is critical for developing various components of the SmartCHANGE AI system. Institutions will focus on creating data preprocessing tools, devising visualization techniques, generating counterfactuals, and constructing robust prediction models. The prime objective during this phase is to leverage the data locally to its full potential.

Funded by the European Union

Phase 2: SmartCHANGE Inter-Institutional Collaboration

As the project progresses into the second phase, collaboration becomes essential. SmartCHANGE collaborators involved in the development of the AI-based pipelines need to share processed data, code, or model architectures to integrate the various components developed in the initial phase. This is a normal part of any interinstitutional project (e.g., EU projects). Within SmartCHANGE, we will collaborate effectively by establishing secure data-sharing protocols, implementing strict access controls, and anonymizing sensitive information during data preprocessing. Additionally, we will share code and AI models through trusted platforms with version control to ensure security. It should be noted here that all the communication and data transfer will be done among the SmartCHANGE collaborators who have signed the Data Sharing Agreements. There will not be any access points exposed to the internet (e.g., there will not be any APIs for clients to communicate automatically). Thus, the risk of privacy attacks at this stage is close to zero.

Client Verification in SmartCHANGE vs. public AI-based services

A crucial distinction between the SmartCHANGE system and public AI-based services (e.g., ChatGPT) is the verification process required for clients to submit queries. The SmartCHANGE system is designed to accept queries exclusively from verified SmartCHANGE clients, as depicted in Figure 2. These clients will be used by SmartCHANGE consortium members, medical professionals, or users participating in the SmartCHANGE pilot studies. If a client cannot provide a SmartCHANGE verification, the access (query) to the system is blocked, and no response is given. This verification step prevents unauthorized access to the SmartCHANGE system (including the AI models).

In contrast, public AI-based services allow any user to register and submit queries. This level of openness presents an opportunity for attackers to construct and leverage multiple targeted queries alongside their responses to deduce sensitive information about the training dataset. However, this is not possible for the SmartCHANGE system. By restricting query access to verified clients, the SmartCHANGE system enhances security, offering a more protected environment compared to public AI systems. This is essential because the majority of security attacks on AI systems are possible only in scenarios where the attackers can continuously query the AI model.

Funded by
the European Union

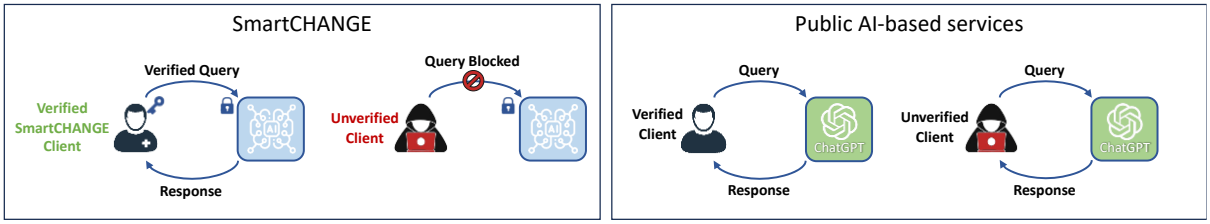


Figure 3: SmartCHANGE system in comparison with public AI-based systems (e.g., ChatGPT). The SmartCHANGE system can be only queried by verified clients, making it more secure compared to public AI- based systems.

Recommendations

Check with DPOs of the institutions processing personal data (i.e. pseudonymised) for machine learning purposes if said processing requires a Data Protection Impact Assessment based on national legislation, past history, and practice (see Annex III).

2.3 National laws

2.3.1 Switzerland

As seen in D2.1 in section 1.2.4 (Sabbadini et el. 2023: 45), data transfer outside of the EU is permitted, among other possibilities, when the country of destination is deemed to offer a level of protection comparable to the one provided in the EU by the GDPR. In the case of Switzerland there is an adequacy decision by the European Commission dating 26 July 2000[6] that was reviewed und upheld very recently, i.e. on 15 January 2024.[7]

Switzerland also adopted in 2020 a new law on data protection that entered into force as of 1st September 2023, named the Federal Act on Data Protection (FADP), which reflects most

[6] Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32000D0518
[7] For the report see: https://commission.europa.eu/document/f62d70a4-39e3-4372-9d49-e59dc0fda3df_en

of the content of the GDPR.[8] In particular, the FADP recognises a special category of data equivalent to those described in GDPR Article 9 (i.e. sensitive data) among its definitions as laid out in Article 5 FADP:

"c. sensitive personal data means:

1. data relating to religious, philosophical, political or trade union-related views or activities,

2. data relating to health, the private sphere or affiliation to a race or ethnicity,

3. genetic data,

4. biometric data that uniquely identifies a natural person,

5. data relating to administrative and criminal proceedings or sanctions,

6. data relating to social assistance measures;"

The processing of this kind of data is considered to be lawful if, among other possibilities:

"e. The controller processes the personal data for purposes not related to specific persons, in particular for **research**, planning or statistics, provided the following requirements are satisfied:

1. The controller anonymises the data as soon as the purpose of processing permits; if anonymity is impossible or if it requires disproportionate effort, the controller shall take appropriate measures to prevent the identification of the data subject.

2. **If the matter involves sensitive personal data, the controller shall disclose such data to third parties in such a manner that the data subject is not identifiable; if this is not possible, it must be guaranteed that the third parties only process the data for purposes unrelated to the data subject's person**.

3. The results are published in such a manner that data subjects are not identifiable." (FADP, Article 31(2)(e), emphasis added)

It is to be noted that the Swiss partner of the SmartCHANGE project, i.e. Università della Svizzera Italiana (USI) will only process data provided by other partners and third-parties, as Switzerland will not feature among the four sites of the proof-of-concept study.

[8] For a complete comparison between the two pieces of legislation (In French) see: https://swissprivacy.law/55/

2.3.2 Taiwan

As SmartCHANGE invited one entity of Taiwan, i.e. the Taipei Medical University (TMU), to become a partner in the project, and as TMU will provide an anonymous dataset derived from medical health records created by several hospitals on the island within the framework of the National Health Insurance agency, this section reports on current data protection legislation of Taiwan and its extra-territoriality or lack thereof, and also deals with ethical review bodies and procedures as currently regulated on the island. It is important to bear in mind that at the moment, as far as personal data transfers are concerned, there is no adequacy decision by the European Commission vis-à-vis Taiwan and that in the context of the SmartCHANGE project the flow of data is meant to go only from Taiwan to the EU, never in the opposite direction. In other words, TMU will not have access to the other datasets processed or collected and processed by the other partners of the project.

The Personal Data Protection Act 2015 (PDPA 2015) reflects in part the content of the GDPR, most notably in relation to some data subjects' rights (PDPA 2015, Articles 8 – 13), some controller's and processor's obligations (PDPA 2015, Article 8) and data breach notifications (PDPA 2015, Article 12). There are, however, many differences between the EU legislation and the Taiwanese one. Of relevance for this section are the following:

1. The PDPA does not refer to or define the notion of 'sensitive data'.
2. The PDPA does not refer to or define the notions of 'pseudonymous' and 'anonymous' data.

In lieu of a definition of 'sensitive data' the PDPA singles out, however, "Data pertaining to a natural person's medical records, healthcare, genetics, sex life, physical examination and criminal records" (PDPA, Article 6), that:

> "shall not be collected, processed or used unless on any of the following bases: […]
> 4. where it is necessary for statistics gathering or **academic research** by a government agency or an **academic institution** for the purpose of healthcare, public health, or crime prevention, **provided that such data**, as processed by the data provider or as disclosed by the data collector, **may not lead to the identification of a specific data subject**; […]" (PDPA, Article 6(4), emphasis added)

At least in relation to health data there is a very close parallel between Article 6 PDPA and Article 9 GDPR, and to be more specific, between the above-mentioned Article 6(4) of the PDPA and Article 9(2)(j) of the GDPR, i.e. the scientific research exception for the processing

of sensitive data provided that the fundamental rights of the data subjects are protected by means of de-identification of the data itself.

Taiwan case law, in case dealing with Article 6(4) PDPA above, interprets the Article 6(4) as giving the right to academic medical institutions to pursue research even without consent of the data subjects, provided that the data is processed after the adoption of the necessary measures for data not to "lead to the identification of a specific data subject" (TCC 2022: (54-46)). It should be noted that while the Court refers to said necessary measures as "pseudonymisation measures" (TCC 2022: (54-56)), neither the Court nor the legislation introduces or defines the notions of 'anonymous' and 'pseudonymous' in the same way Recital 26 GDPR does. In other words: the collection of the personal data and its processing in de-identified form for research purposes in the way it was done in Taiwan would be considered legal also in the EU, also taking into account that the measures for the keeping of medical health records, in terms of safety, are strictly regulated by the Regulations on the Production and Management of Electronic Medical Records (see Annex IV), by Article 79 and 80 of the National Health Insurance Act[9] and Article 70 of the Medical Care Act[10].

Moreover, the description of the techniques used by the hospitals operating within the framework of National Health Insurance Act[11], such as TMU, to render anonymous the datasets before they are sent to third-parties, such as other hospitals on the island or foreign parties such as the SmartCHANGE partners, however, as listed in a letter to the consortium (annex IV), include: removal of direct identifiers, de-identification of indirect identifiers, data perturbation, data aggregation, followed by risk assessment and testing. The SmartCHANGE partners are satisfied that these procedures have rendered the dataset anonymous for the purposes of the SmartCHANGE project.

As far as extra-territoriality is concerned, although the PDPA itself is not clear in relation to this matter, according to Ken-Ying-Tseng of the Taiwanese law firm "Lee and LI", "The current position of the National Development Council (NDC) is that the PDPA does not have extra-territorial application."[12]

The SmartCHANGE partners, however, before proceeding with the use of the dataset provided by TMU, will have the ethics committee of either VUB, VUMC or one of the other

[9] https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=L0060001
[10] https://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=L0020021
[11] https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=L0060001
[12] https://www.mondaq.com/privacy/1005652/data-privacy-comparative-guide

EU-based project partners review the relevant Taiwanese ethics committee's decision[13] vis-à-vis the creation of said dataset, and include an opinion by SmartCHANGE's very own Ethics Board on the processed carried out in Taiwan at the time of creation of the dataset. In particular, the ethics committees based in the EU will ascertain how compatible the relevant ethics principles followed by the Taiwanese committees are with those of the Horizon Europe principles as laid out in Articles 18 and 19 of Regulation 2021/695 (Horizon Europe Regulation) and the European Code of Conduct for Research Integrity (ALLEA). Report on these reviews will feature in the ethics review deliverable due on month 12 or at latest in deliverable D2.4, the SELP Impact Assessment (month 14).

Proof-of-concept study

TMU, if granted permission to join the SmartCHANGE consortium, will be also the 5th site for the proof-of-concept study. Research on human subjects in Taiwan is regulated both by Articles 79 and 80 of the Medical Care Act[14] and by the Human Subjects Research Act[15]. Before the proof-of-concept study, TMU will submit a request to its own Institutional Review Board, i.e. the ethics committee of the institution as regulated by the Regulations for Organization and Operation of Human Research Ethics Board[16], to be reviewed also by one of the Ethics Committees of the other four sites for the proof-of-concept study based in the EU.

Recommendations

1. Data received from TMU can legitimately be considered anonymous.

2. No EU data should be transferred to Taiwan.

3. Relevant Taiwanese Ethics Committee's decision at the time of the creation of the dataset to be reviewed by EU-based Ethics Committee of one of the SmartCHANGE partners, either VUB, VUMC or another partner's and also by the SmartCHANGE Ethics Board. No data from Taiwan is to be processed before this review.

[13] The establishment and function of ethics committees in Taiwan is regulated by the following act (in English): https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=L0020179
[14] https://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=L0020021
[15] https://law.moj.gov.tw/Eng/LawClass/LawAll.aspx?PCode=L0020176
[16] See note 13

Funded by the European Union

4. Proof-of-concept study: the Ethics Committee of one of the four partners engaged in the proof-of-concept study in the EU will review the assessment by the Institutional Review Board of TMU of the proof-of-concept study to be carried out in Taiwan.

Funded by
the European Union

3 AI Act: Update

At the time of writing this document, the final version of the AI Act has not yet been approved by the European Parliament, although a final version (AIA Compromise) of the compromise between the European Commission, the European Council and the European Parliament is available. While the final version of the Act relies on an agreement between the parties as to the use of facial/biometric recognition in real time by law enforcement agencies in certain circumstances, most of the norms of this piece of legislation in relation to non-high-risk AI systems have not changed. The vote of the Parliamentary Committees on the final vote on the Act is expected on 13 February 2024, while the vote of the whole Parliament is foreseen on April 10-11, 2024.[17]

In addition to what has been written in deliverable D2.1 (Sabbadini et al. 2023), under section 3.1, the following remarks need to be added and taken into consideration by the SmartCHANGE partners.

First, while as such the SmartCHANGE AI tool does not fall under the category of high-risk AI systems, things might change if and when the tool were to be certified as a medical device in full alignment with the MDR, because in Article 6(1) of the AIA (Compromise), high-risk AI systems are those where:

"[...] (a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex II;

(b) the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II. [...]" (AIA Compromise, Article 6(1))

Although the MDR features in annex II of the AIA, it is not clear at this stage if all medical devices that use an AI component will fall under the regulation: this is the kind of question that the European Commission will presumably answer in the guidelines on the regulation it will release, otherwise a question on this topic will have to be posed to the future national AI

[17] The EU AI Act Newsletter, 25/01-05/02 2024, available at: https://artificialintelligenceact.substack.com/p/the-eu-ai-act-newsletter-45-ai-office?r=2c6nee

authorities/boards. For the sake of clarity it is important to note that the definition of AI system in the regulation is as follows (AIA compromise, Article 3(1)):

"'AI system' is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"

Second, regardless of the future status as medical device of the SmartCHANGE tool, the obligation of transparency under AIA Article 52(1) still stands, i.e. the obligation of ensuring that humans interacting with the AI tool understand that they are interacting with an AI system and not a human, as is mandatory for systems:

"intended to **interact with natural persons**", where the obligation for providers is to inform natural persons "that they are interacting with an AI system, unless this is obvious from the point of view of a natural person who is reasonably well-informed, observant and circumspect, taking into account the circumstances and the context of use" (AIA, Compromise Article 52(1), emphasis added)

In terms of concrete impact on the SmartCHANGE project, however, what is also relevant are the dates of application of the regulation and the drafting and issuing of guidelines by the European Commission and/or delegated bodies vis-à-vis the practical application of the norms. The AI Act will enter into force 20 days after its publication on the Official Journal of the European Union, i.e. at least 20 days after the final vote in the Parliament. There are different dates of application for the regulation: 6 months (after the entry into force) for Title I and II (prohibitions), 12 months for Title III Chapter 4, Title VI, VIIIa and X, and 36 months for Article 6(1), i.e. the very article that defines high-risk systems. In other words, the SmartCHANGE tool will be ready before the application of the AI Act, even if it were to follow a certification process in full compliance with the MDR.

There are, moreover, further factors to be considered before the full application of the regulation, as EU and national supervisory bodies will need to be established, guidelines and codes of conduct will need to be issued, regulatory sandboxes for the testing of high-risk systems will need to be instituted at the national level. Furthermore, EU member states can

Funded by the European Union

still shape further the implementation of the regulation by means of circa 20 secondary legislative acts.[18]

In sum, even if choosing to embark on an MDR-related certification process, adhering to the obligations for high-risk systems described in Chapters 2, 3 and 4 of Title III of the AI Act in the absence of the bodies and secondary/tertiary norms described above would prove to be very difficult, not to mention administratively and financially cumbersome, in relation to the scope of the actual deployment of the SmartCHANGE tool. Nevertheless, we will follow the developments of these issues and report back to the SmartCHANGE partners if anything of consequence vis-à-vis the project will turn up.

Recommendations

Proceed with creation of risk-prediction model and SmartCHANGE tool as a non-high-risk AI system, while following developments in the emergence of tertiary law at EU and national level.

[18] *Ibid.*

Funded by the European Union

4 SmartCHANGE as a Medical Device

As stated at the time of the submission of the proposal of the SmartCHANGE project:

> "The SmartChange [sic] project is a proof-of-concept project. It does not intend to directly bring a medical device to market. Given this it is not intended to achieve certification within the project's lifetime. It will however follow the relevant aspects of the Medical Device Framework in order to ensure safety of all participants within the clinical study envisaged within the project." (Grant Agreement, p. 158)

On this basis, and following the discussion in section 1.3.9 of deliverable D2.1 (Sabbadini et al. 2023), it makes sense to treat SmartCHANGE as a research project for "devices, manufactured and used only within health institutions established in the Union" (MDR Article 5(5)), thus falling under the scope of Article 5(5) of the MDR and with the obligation to fulfil requirements listed in Article 5(5) MDR itself and Annex I MDR. While we refer to section 1.3.9.1, 1.3.9.2 and 1.3.9.3 of deliverable D2.1 for an analysis of the above-mentioned requirements, the partners of the four sites of the proof-of-concept study need to ensure that all four sites fall under the definition of 'health institution' as stated in the MDR Article 2(36), i.e. "an organisation the primary purpose of which is the care or treatment of patients or the promotion of public health". *Prima facie* all four sites of the proof-of-concept study seem to fit with the description of Article 2(36) MDR, though entities that are not hospitals should check with their respective administrations if their status as health institution can be corroborated by documentation or other means attesting said status, e.g. proof that the institution was the site of clinical trials in the past or a history of promotion of public health, etc.

A dedicated workshop for the computer science, medical and technical partners of the project will be held in the course of the consortium meeting of May 2024, i.e. M13 of the project, for a collective analysis of annex I of the MDR and a careful repartition of responsibilities among partners vis-à-vis the requirements listed in Annex I. In the same workshop partners will have to agree on how and where to keep the record describing the fulfilment of each requirement listed in Annex I and Article 5(5) MDR.

It is important to note that apart from requirements listed in the MDR there are various guidance documents issued by the Medical Device Coordination Group, i.e. an EU-level body of experts established on the basis of Article 103 MDR and whose tasks are described in Article 105 MDR and include, among others, the following:

Funded by the European Union

"[…]

(c) to contribute to the development of guidance aimed at ensuring effective and harmonised implementation of this Regulation, in particular regarding the designation and monitoring of notified bodies, application of the general safety and performance requirements and conduct of clinical evaluations and investigations by manufacturers, assessment by notified bodies and vigilance activities;

(d) to contribute to the continuous monitoring of technical progress and assessment of whether the general safety and performance requirements laid down in this Regulation and Regulation (EU) 2017/746 are adequate to ensure safety and performance of devices, and thereby contribute to identifying whether there is a need to amend Annex I to this Regulation;

(e) to contribute to the development of device standards, of CS and of scientific guidelines, including product specific guidelines, on clinical investigation of certain devices in particular implantable devices and class III devices;

[…]" MDR, Article 105

The documents of the MDCG, however, represent an example of 'soft law', i.e. a tertiary form of legal rule that is "not directly enforceable through criminal or civil proceedings, but which may, nevertheless, produce indirect legal effect." (Cane/Conaghan 2008:721). In other words, and as stated by the MDCG on the first page of their documents, these documents are "not a European Commission document and [it] cannot be regarded as reflecting the official position of the European Commission. Any views expressed in this [these] document[s] are **not** legally binding and **only the Court of Justice of the European Union can give binding interpretations of Union law**." (MDCG 2023-4: 1, emphasis added)

In light of the fact that the SmartCHANGE tool will consist of an *ad hoc* created software to be used in combination with an already existing wearable device, the main three MDCG documents to be considered in the case of SmartCHANGE are:

1. *Guidance on Cybersecurity for medical devices* (MDCG 2019-16)
2. *Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software* (MDCG 2020-1)
3. *Medical Device Software (MDSW) – Hardware combinations. Guidance on MDSW intended to work in combination with hardware or hardware components* (2023-4)

Funded by the European Union

The first document, also referred to in 1.3.9.1 of deliverable D2.1 (Sabbadini et al. 2023), deals with cybersecurity requirements in general and those related to Annex I MDR, while also introducing basic concepts of cybersecurity, instructions for secure design and manufacture, for documentation and instructions for use and reference to other relevant legislation and international standards.

The second document deals with the performance evaluation of medical device software, laying out basic principles and describing the evaluation procedure, while providing at the end, an example of the evaluation of different kinds of devices, including one for the analysis of sleep quality data, which would be of interest for the SmartCHANGE tool as sleep is one of the parameters the tool is intended to measure.

The third document deals with the integration of hardware and software components to create a medical device, exploring different scenarios, including the one the SmartCHANGE tool would fall into, i.e. where the manufacturer of the hardware (in this case the wearable device) and the manufacturer of the software app are two different entities.

While it is safe to assume that technical and computer science partners have already familiarity with these guidance documents, having encountered them in the past in relation to other projects, the process leading to deliverable 2.4, i.e. 'SELP Impact Assessment' will help ascertaining whether this is the case or not.

Recommendations

1. Adhere to requirements of Article 5(5) and Annex I MDR

2. Take into consideration guidance documents provided by the MDCG

3. Technical, Computer Science and Medical Partners to coordinate at ad hoc workshop session in May 2024 (3rd Consortium in-person meeting) repartition of above-mentioned requirements and coordination vis-à-vis record-keeping of fulfilment of requirements.

5 Table of correspondence

Below is a table that sums up succinctly the main relevant legislation for each WP, taking into account only those WPs that deal with the creation of the risk-prediction model, the creation of the SmartCHANGE tool and the proof-of-concept study. Additional sectoral guidelines and standards will be identified by querying the partners through the questionnaire leading to the SELP Impact Assessment (T2.4), which will be presented in deliverable D2.4.

Table 1. Correspondence between task and main relevant legislation

WP	Task	Description	Regulation	Remarks
4	4.1	**Dataset collection and predictor selection**	GDPR	Pseudonymous and personal data
	4.2	**Data pre-processing and harmonisation**	GDPR	Pseudonymous and personal data
	4.3	**A federated-learning approach to predicting NCDs**	GDPR	Pseudonymous and personal data
	4.4	**Baseline predictive models for NCDs**		
	4.5	**Unified predictive model on heterogeneous data for NCDs**		
5	5.1	**Self-audit, tuning models' performance, robustness and generalisation**	AIA and related	For future exploitation (if so decided)
	5.2	**Explainable AI tools**	AIA and related	For future exploitation (if so decided)
	5.3	**Visual analytics for model and prediction understanding**		
6	6.1	**Technical specification**	GDPR, MDR: 5(5) and Annex I and MDCG related	

Funded by the European Union

	6.2	**Development of the mobile application for families**	GDPR, MDR: 5(5) and Annex I and MDCG related	
	6.3	**Development of the system for healthcare professionals**	GDPR, MDR: 5(5) and Annex I and MDCG related	And in future AIA related if exploitation decided
	6.4	**System integration and configuration**	GDPR, MDR: 5(5) and Annex I and MDCG related	And in future AIA related if exploitation decided
7	7.1	**Study design and recruitment**	Ethical requirements as per D2.1	Other regulations depending on final version of study protocol
	7.2	**SmartCHANGE solution provisioning and deployment**		Other regulations depending on final version of study protocol
	7.3	**Study implementation**		Other regulations depending on final version of study protocol
	7.4	**Evaluation of feasibility and usability**		Other regulations depending on final version of study protocol
	7.5	**Evaluation of potential efficacy in cardiovascular and metabolic risk lowering**		

Funded by the European Union

6 Conclusions and next steps

Partners involved in the tasks of WP 3, 4, 5, 6 and 7 should read this deliverable and re-read deliverable D2.1 if necessary and verify their understanding and applicability of both vis-à-vis the tasks they are leaders of and come back with questions to the Authors. An interactive questionnaire will be sent by month 13 whose replies will serve as the basis for D2.4, i.e. the SELP impact assessment.

Funded by
the European Union

7 References

7.1 Legislation, Treaties, Case Law and Opinions

AIA: Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

AIA COMPROMISE, 26 January 2024. Available at: https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf

ALLEA: *The European Code of Conduct for Research Integrity*, available at: https://allea.org/code-of-conduct/

CJEU, C-582/14, *Breyer v. Bundesrepublik Deutschland*, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62014CJ0582

CJEU, T-557/20, SRB v EDPS, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62020TA0557

FADP 2020: Federal Act on Data Protection. Official translation in English, available at: https://www.fedlex.admin.ch/eli/cc/2022/491/en

GDPR: 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, and repealing Directive 95/46/EC (general Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679

HORIZON EUROPE REGULATION: Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013, available at: https://eur-lex.europa.eu/eli/reg/2021/695/oj

MDCG 2019-16. *Guidance on Cybersecurity for medical devices*. Document by the Medical Device Coordination Group, available at: https://ec.europa.eu/docsroom/documents/41863

MDCG 2020-1. *Guidance on Clinical Evaluation (MDR) / Performance Evaluation (IVDR) of Medical Device Software*. Document by the Medical Device Coordination Group, available at: https://ec.europa.eu/docsroom/documents/40323

MDCG 2023-4. *Medical Device Software (MDSW) – Hardware combinations. Guidance on MDSW intended to work in combination with hardware or hardware components.* Document by the Medical Device Coordination Group, available at: https://health.ec.europa.eu/system/files/2023-10/md_mdcg_2023-4_software_en.pdf

MDR: Regulation on medical devices 2017/745 of the European Parliament and of the Council of April 2017, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R0745&qid=1697462381286

PDPA 2015: Personal Data Protection Act of 2015. Official translation into English of the Taiwan law available at: https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=I0050021

TCC 2022: Taiwan Constitutional Court judgement 111-Hsien-Pan-13 (Case on the National health Insurance Research Database), available at: https://cons.judicial.gov.tw/en/docdata.aspx?fid=2170&id=347736

WP29 2014: Article 29 Data Protection Working Party. *Opinion 05/2014 on anonymisation techniques*. Available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf accessed on October 14 2023

7.2 Bibliography

BROWN, HANNAH, LEE, KATHERINE, MIRESHGHALLAH, FATEMEHSADAT, SHOKRI, REZA, TRAMER, FLORIAN, 2022. "What Does it Mean for a Language Model to Preserve Privacy?", in: In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3531146.3534642

CANE, PETER & CONAGHAN, JOANNE (eds.), 2008. *The New Oxford Companion to Law*. Oxford, Oxford University Press.

CRUYT, KAREN, 2023. "Anonymity is in the eye of the beholder… or is it?", VUB HALL Blog entry available at: https://hall.research.vub.be/anonymity-is-in-the-eye-of-the-beholder-or-is-it

DAUTLICH, MARC, COHEN, GUY, GRAZETTE, MARCUS, 2021. *Introduction to Anonymisation.* Own publication by Bristows and Privitar, available at: https://go.privitar.com/rs/588-MYA-374/images/2021-07-Privitar-Bristows-Intro_to_Anonymisation.pdf

FINCK, MICHÈLE, PALLAS, FRANK, 2020. "They Who Must Not Be Identified: Distinguishing Personal from Non-Personal Data under the GDPR" in SSRN Electronic Journal 10(1):11-36, DOI 10.1093/idpl/ipz026

FRA/COE: European Union Agency for Fundamental Rights and Council of Europe, 2018. *Handbook on European data protection law*. Luxembourg, Publications Office of the European Union.

GADOTTI, ANDREA, HOUSSIAOU, FLORIMOND, ANNAMALAI, MEENATCHI, DE MONTJOYE, YVES-ALEKANDRE, 2023. "Pool Inference Attacks on Local Differential Privacy: Quantifying the Privacy Guarantees of Apple's Count Mean Sketch in Practice" Preprint available at: https://www.researchgate.net/publication/370058817_Pool_Inference_Attacks_on_Local_Differential_Privacy_Quantifying_the_Privacy_Guarantees_of_Apple's_Count_Mean_Sketch_in_Practice accessed 16 October 2023

GROOS, DANIEL, VAN VEEN, EVERT-BEN, 2020. "Anonymised Data and the Rule of Law " in EDPL (4)2020: 498-508. Available at: https://edpl.lexxion.eu/data/article/16563/pdf/edpl_2020_04-007.pdf

KISELEVA, A., 2019. "Decisions made by AI versus transparency: Who wins in Healthcare?" In T. C. Bächle & A. Wernick (Eds.), *The futures of eHealth. Social, Ethical and legal challenges*. Berlin, Germany: Humboldt Institute for Internet and Society.

LI, TIANCHENG, LI NINGHUI, 2009. "On the Tradeoff between Privacy and Utility in Data Publishing". Conference paper available at: https://www.cs.purdue.edu/homes/ninghui/papers/privacy_utility_kdd09.pdf accessed 15 October 2023

NASR, MILAD *et alii*, 2023. "Extracting Training Data from ChatGPT". Available at: https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html

SABBADINI, RENATO, QUINN, PAUL, ALTENBURG, TEATSKE, FAWAD, TAJ, 2023. SmartCHANGE (EU funded project) D2.1 *Benchmark of regulatory and ethical frameworks*.

SPIECKER GEN. DÖHMANN, INDRA, PAPAKONSTANTINOU VAGELIS, HORNUNG, GERRIT, DE HERT, PAUL (eds.), 2023. *General Data Protection Regulation: Article-by-Article Commentary.* London, Beck/Nomos/Hart Bloomsbury Publishing.

VEALE, MICHAEL, BINNS, REUBEN, EDWARDS, LILIAN, 2018. "Algorithms that remember: model inversion attacks and data protection law", in *Philosophical Transactions of the Royal Society A* 376:20180083, available at: https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0083

Funded by the European Union

Appendix I. EU Data Transfer Agreement

Data Transfer Agreement

between

…… [Full name and address of legal entity], henceforth named the "**Provider**",

and

1. **INSTITUT JOZEF STEFAN (JSI)**, established in Jamova 39, LJUBLJANA 1000, Slovenia,
2. **UNIVERSITY OF PIRAEUS RESEARCH CENTER (UPRC)**, established in Al. Papanastasiou 91, Piraeus 185 33, Greece,
3. **TECHNISCHE UNIVERSITEIT EINDHOVEN (TUE)**, established in Groene Loper 3, 5612 AE Eindhoven, The Netherlands,
4. **UNIVERSIDADE DO PORTO (UPORTO)**, established in Praca Gomes Teixeira, Porto 4099-002, Portugal,
5. **UNIVERSITA DELLA SVIZZERA ITALIANA (USI)**, established in Via G. Buffi 13, 6900 Lugano, Switzerland,

henceforth named the "**Recipients**". The Provider and the Recipients are collectively named the "**Parties**". This Agreement governs the transfer of the following dataset:

[name and description of the dataset]

from the Provider to JSI, on behalf of all Recipients,

in

[please, choose one]

□ anonymised format

□ pseudonymised format

obtained through the following procedure:

[short description of the procedure used to anonymise or pseudonymise the dataset]

for the following research purposes within the framework of the EU-funded project "AI-based long-term health risk evaluation for driving behaviour change strategies in children and youth" (a.k.a. "SmartCHANGE"):

Training of machine-learning models to predict long-term health risk of children and youth. The dataset may be combined with other datasets. The exact definition of risk is not yet defined, but it will be cardio-metabolic risk. An important aspect of the research will be trustworthiness of AI, which means that the data and models will be visualised, methods to explain the models and their predictions will be applied, and advanced approaches to evaluate them will be used. The models will be integrated in risk-assessment tools, which will then be evaluated in a research context.

The Provider and the Recipients agree on the following conditions governing this Agreement:

1. Having performed duty of care and having had its Data Protection Officer and/or Legal Department review this Agreement, the Provider attests that: 1) all raw data at the origin of the dataset has been obtained in accordance with EU General Data Protection Regulation (EU 2016/679) and applicable national laws and regulations, using research conducted in accordance with sound scientific methods and principles, ii) the Provider is authorized to transfer the dataset mentioned in this Agreement and iii) that said transfer does not infringe upon the EU General Data Protection Regulation (EU 2016/679) or the relevant national legislation.

2. The Recipients will not transmit or grant access to the dataset to other parties not mentioned in this Agreement, except for processing/storing purposes within the framework of separate agreements based on EU General Data Protection Regulation (EU 2016/679) and/or strict confidentiality clauses (in case of anonymised data).
3. The Recipients will use the dataset provided by the Provider only for the research purposes listed in this Agreement and act, within the context of the uses of the pseudonymised dataset for the purposes of the SmartCHANGE project, as controller of the data and comply with the relevant norms of EU General Data Protection Regulation and all relevant national norms. Any other use will require the prior written approval of the Provider.
4. Any publication by the Recipients based on the use of the dataset provided by the Provider shall either mention the name of the Provider in the 'Acknowledgements' section of the publication or, in alternative, the names of the main researchers of the Provider, associated with the provided dataset and named by the Provider in this clause of this Agreement, may be added to the list of co-authors of the publication unless the same researchers indicate their wish not to be listed as co-authors. The main researchers associated with the dataset provided by the Provider are: i) ……., ii) ……..
5. The Recipients shall not perform any act which could lead to the re-identification of the data subjects at the origin of the dataset, including by linking different sets of information, comparing and processing information.
6. In case of a pseudonymised dataset, the Recipients will process the data on the legal basis described in article 9(2)(j)[19] of EU General Data Protection Regulation (EU 2016/679) and comply with all relevant norms of the same Regulation and relevant national laws.
7. The Recipients shall take reasonable steps to ensure the reliability of any employee, agent or contractor of any contracted processor who may have access to the dataset provided by the Provider, ensuring in each case that access is strictly limited to those individuals who need to know / access the dataset, as strictly necessary for the purposes of this Agreement, and to comply with applicable laws in the context of that individual's duties to the Recipients or contracted processor, ensuring that all such individuals are subject to confidentiality undertakings or professional or statutory obligations of confidentiality.

[19] "processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject"

Funded by the European Union

8. In case of pseudonymised data, taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the Recipients shall in relation to the dataset provided by the Provider implement appropriate technical and organizational measures to ensure a level of security appropriate to that risk, including, as appropriate, the measures referred to in Article 32(1) of the GDPR. In assessing the appropriate level of security, the Recipients shall take account in particular of the risks that are presented by processing, in particular from a personal data breach.
9. The Recipients shall notify the Provider without undue delay upon the Recipients becoming aware of a personal data breach affecting the Provider's dataset, sending the Provider sufficient information to allow the Provider to meet any obligations to report or inform data subjects of the personal data breach under EU General Data Protection Regulation and/or relevant national laws.
10. The Recipients may not transfer or authorize the transfer of the dataset to countries outside the EU, with the exception of Switzerland, without the prior written consent of the Provider. If personal data processed under this Agreement is transferred from a country within the European Economic Area and Switzerland to a country outside the European Economic Area and Switzerland, the Parties shall ensure that the personal data are adequately protected. To achieve this, the Parties shall, unless agreed otherwise, rely on EU approved standard contractual clauses for the transfer of personal data.
11. Each Party must keep this Agreement and information it receives about the other Parties and its business in connection with this Agreement confidential and must not use or disclose that confidential information without the prior written consent of the other Party except to the extent that: (a) disclosure is required by law; (b) the relevant information is already in the public domain.
12. This Agreement enters into force on the date in which, once all Parties have signed this Agreement, JSI receives, on behalf of all Recipients, the dataset from the Provider and will end on 30th April 2029, i.e. two years after the end of the SmartCHANGE project. The Agreement may be terminated by either party at any time for any reason upon thirty (30) days written notice. The Recipients commit to the deletion of the dataset at the end of the duration of this Agreement.
13. The terms of this Agreement can be changed only by a written modification by the authorised signatories of the Parties (or their designated representatives) to this Agreement or by the Parties adopting a new agreement in place of this Agreement.
14. This Agreement shall be exclusively governed by, and construed in all respects in accordance with the laws of Belgium without regard to its conflicts of laws rules. Any claims, controversies or disputes arising out of or in connection with this Agreement

which cannot be settled amicably between the Parties, shall be subject to the exclusive jurisdiction of the competent court in Brussels, Belgium.

IN WITNESS WHEREOF the duly authorized signatories of the Parties signed this amendment in five (5) originals, all in the English language, all having the same validity, for and on behalf of

[The Provider]

(Legal Representative) (Data Protection Officer) (Acknowledged name)

INSTITUT JOZEF STEFAN (JSI)

(Legal Representative) (Acknowledged name)

UNIVERSITY OF PIRAEUS RESEARCH CENTER (UPRC)

(Legal Representative) (Acknowledged name)

TECHNISCHE UNIVERSITEIT EINDHOVEN (TUE)

(Legal Representative) (Acknowledged name)

UNIVERSIDADE DO PORTO (UPORTO)

(Legal Representative) (Acknowledged name)

UNIVERSITA DELLA SVIZZERA ITALIANA (USI)

(Legal Representative) (Acknowledged name)

Appendix II. Data Transfer Agreement for 3rd countries

DATA TRANSFER AGREEMENT

for datasets from non-EU countries transferred to the European Union and Switzerland

between

…… [Full name and address of legal entity based in the non-EU country], henceforth named the "**Provider**",

and

1. **INSTITUT JOZEF STEFAN (JSI)**, established in Jamova 39, LJUBLJANA 1000, Slovenia,
2. **UNIVERSITY OF PIRAEUS RESEARCH CENTER (UPRC)**, established in Al. Papanastasiou 91, Piraeus 185 33, Greece,
3. **TECHNISCHE UNIVERSITEIT EINDHOVEN (TUE)**, established in Groene Loper 3, 5612 AE Eindhoven, The Netherlands,
4. **UNIVERSIDADE DO PORTO (UPORTO)**, established in Praca Gomes Teixeira, Porto 4099-002, Portugal,
5. **UNIVERSITA DELLA SVIZZERA ITALIANA (USI)**, established in Via G. Buffi 13, 6900 Lugano, Switzerland,

henceforth named the "**Recipients**". The Provider and the Recipients are collectively named the "**Parties**". This Agreement governs the transfer of the following dataset:

[name and description of the dataset]

from the Provider to JSI, on behalf of all Recipients,

in

[please, choose one]

□ anonymised format

□ pseudonymised format

obtained through the following procedure:

[short description of the procedure used to anonymise or pseudonymise the dataset]

for the following research purposes within the framework of the EU-funded project "AI-based long-term health risk evaluation for driving behaviour change strategies in children and youth" (a.k.a. "SmartCHANGE"):

Training of machine-learning models to predict long-term health risk of children and youth. The dataset may be combined with other datasets. The exact definition of risk is not yet defined, but it will be cardio-metabolic risk. An important aspect of the research will be trustworthiness of AI, which means that the provided dataset and models trained on it will be visualised, methods to explain the models and their predictions will be applied, and advanced approaches to evaluate them will be used. The models will be integrated in risk-assessment tools, which will then be evaluated in a research context.

The Provider and the Recipients agree on the following conditions governing this Agreement:

1. Having performed duty of care, the Provider attests that: i) all raw data at the origin of the dataset has been obtained in accordance with applicable laws and regulations, using research conducted in accordance with sound scientific methods and principles, ii) the Provider is authorized to transfer the dataset mentioned in this Agreement and

iii) that said transfer does not infringe upon local or national legislation and that, in case of pseudonymised data, either the protection afforded by the EU General Data Protection Regulation (EU 2016/679) is deemed sufficient and compatible with local or national legislation or that the relevant national or local data protection legislation has no extraterritoriality.

2. The Recipients will not transmit or grant access to the dataset to other parties not mentioned in this Agreement, except for processing/storing purposes within the framework of separate agreements based on EU General Data Protection Regulation (EU 2016/679) and/or strict confidentiality clauses (in case of anonymised data).
3. The Recipients will use the dataset provided by the Provider only for the research purposes listed in this Agreement. Any other use will require the prior approval of the Provider.
4. Any publication by the Recipients based on the use of the dataset provided by the Provider shall either mention the name of the Provider in the 'Acknowledgements' section of the publication. Alternatively, in health journal papers, up to three names (subject to limitations by the publication venue) of the main researchers of the Provider, associated with the provided dataset and named by the Provider in this clause of this Agreement, may be added to the list of co-authors of the publication according to the Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals, established by the International Committee of Medical Journal Editors, unless the same researchers indicate their wish not to be listed as co-authors. The main researchers associated with the dataset provided by the Provider are: i) ..... , ii) ..... , iii) ..... .
5. The Recipients shall not perform any act which could lead to the re-identification of the data subjects at the origin of the dataset, including by linking different sets of information, comparing and processing information.
6. In case of a pseudonymised dataset, the Recipients will process the data on the legal basis described in article 9(2)(j)[20] of EU General Data Protection Regulation (EU 2016/679) and comply with all relevant norms of the same Regulation.
7. The Recipients shall ensure that all personnel and agents, the latter contracted on the basis of article 2 of this Agreement, with access to the dataset comply with the terms of this Agreement, as well as EU General Data Protection Regulation in case of pseudonymised data, while being bound by confidentiality.

[20] "processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject"

Funded by the European Union

8. The Recipient shall take appropriate technical and organisational measures to safeguard a level of security attuned to the risk, so that the use of the provided dataset complies with the requirements under the EU General Data Protection Regulation.
9. This Agreement enters into force on the date in which, once all Parties have signed this Agreement, JSI receives, on behalf of all Recipients, the dataset from the Provider and will end on 30th April 2029, i.e. two years after the end of the SmartCHANGE project. The Agreement may be terminated by either party at any time for any reason upon thirty (30) days written notice. The Recipients commit to the deletion of the dataset at the end of the duration of this Agreement.
10. The terms of this Agreement can be changed only by a written modification by the authorised signatories of the Parties (or their designated representatives) to this Agreement or by the Parties adopting a new agreement in place of this Agreement.
11. This Agreement shall be exclusively governed by, and construed in all respects in accordance with the laws of Belgium without regard to its conflicts of laws rules. Any claims, controversies or disputes arising out of or in connection with this Agreement which cannot be settled amicably between the Parties, shall be subject to the exclusive jurisdiction of the competent court in Brussels, Belgium.

IN WITNESS WHEREOF the duly authorized signatories of the Parties signed this amendment in five (5) originals, all in the English language, all having the same validity, for and on behalf of

[The Provider]

(Legal Representative) (Acknowledged name)

Funded by the European Union

INSTITUT JOZEF STEFAN (JSI)

Director Prof. Dr. Boštjan Zalar

Funded by
the European Union

UNIVERSITY OF PIRAEUS RESEARCH CENTER (UPRC)

(Legal Representative) (Acknowledged name)

Funded by
the European Union

TECHNISCHE UNIVERSITEIT EINDHOVEN (TUE)

(Legal Representative) (Acknowledged name)

UNIVERSIDADE DO PORTO (UPORTO)

(Legal Representative) (Acknowledged name)

UNIVERSITA DELLA SVIZZERA ITALIANA (USI)

(Legal Representative) (Acknowledged name)

Appendix III. Data Protection Impact Assessment template

Data Protection Impact Assessment (DPIA)

1. Legal Framework

Article 35 GDPR

1. Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.

2. The controller shall seek the advice of the data protection officer, where designated, when carrying out a data protection impact assessment.

3. A data protection impact assessment referred to in paragraph 1 shall in particular be required in the case of:

(a) a systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;

(b) processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offences referred to in Article 10; or

(c) a systematic monitoring of a publicly accessible area on a large scale.

4. The supervisory authority shall establish and make public a list of the kind of processing operations which are subject to the requirement for a data protection impact assessment pursuant to paragraph 1. The supervisory authority shall communicate those lists to the Board referred to in Article 68.

5. The supervisory authority may also establish and make public a list of the kind of processing operations for which no data protection impact assessment is required. The supervisory authority shall communicate those lists to the Board.

Funded by the European Union

6. Prior to the adoption of the lists referred to in paragraphs 4 and 5, the competent supervisory authority shall apply the consistency mechanism referred to in Article 63 where such lists involve processing activities which are related to the offering of goods or services to data subjects or to the monitoring of their behaviour in several Member States, or may substantially affect the free movement of personal data within the Union.

7. The assessment shall contain at least:

(a) a systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;

(b) an assessment of the necessity and proportionality of the processing operations in relation to the purposes;

(c) an assessment of the risks to the rights and freedoms of data subjects referred to in paragraph 1; and

(d) the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data and to demonstrate compliance with this Regulation taking into account the rights and legitimate interests of data subjects and other persons concerned.

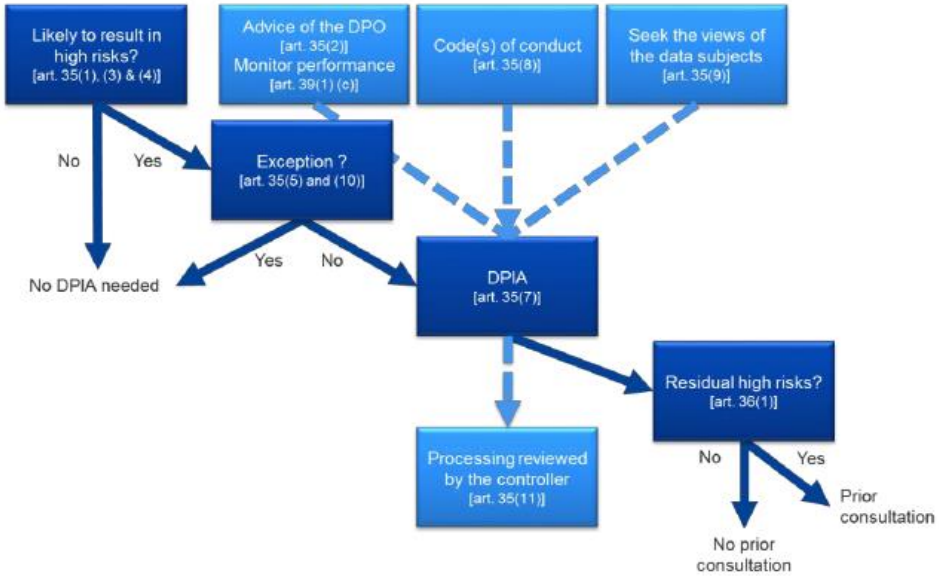
8. Compliance with approved codes of conduct referred to in Article 40 by the relevant controllers or processors shall be taken into due account in assessing the impact of the processing operations performed by such controllers or processors, in particular for the purposes of a data protection impact assessment.

9. Where appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing, without prejudice to the protection of commercial or public interests or the security of processing operations.

10. Where processing pursuant to point (c) or (e) of Article 6(1) has a legal basis in Union law or in the law of the Member State to which the controller is subject, that law regulates the specific processing operation or set of operations in question, and a data protection impact assessment has already been carried out as part of a general impact assessment in the context of the adoption of that legal basis, paragraphs 1 to 7 shall not apply unless Member States deem it to be necessary to carry out such an assessment prior to processing activities.

11. Where necessary, the controller shall carry out a review to assess if processing is performed in accordance with the data protection impact assessment at least when there is a change of the risk represented by processing operations.

In its Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is 'likely to result in a high risk" for the purposes of Regulation 2016/679, Article 29 Working Party summed up the procedure on whether a DPIA is necessary with following figure:



2. Preliminary questions

1) Date _____

2)Title and description of the application/project _____

3) Name of the owner of the application/project _____

4) DPO _____

5) Does the project process personal data? □ Yes □ No

If Yes, for which purposes?

_____
_____
_____
_____
_____
_____

6) If the answer to question 5 is Yes, the processing operations are included in the *whitelist* in which the supervisory authority excludes the performance of the DPIA?

□ Yes □ No

Funded by the European Union

If Yes, which is the operation that does not require the performance of a DPIA?

_____
_____
_____
_____
_____
_____

7) If the answer to question 6 is No, the processing operations are included in the *blacklist* in which the supervisory authority requires the performance of the DPIA?

□ Yes □ No

If Yes, which is the operation that requires the performance of a DPIA?

_____
_____
_____
_____
_____
_____

8) Has a DPIA been performed already on the application/project for the same operations?

□ Yes □ No

If Yes, which are the already performed DPIAs?

_____
_____
_____
_____
_____
_____

3. Questions that trigger the DPIA

If you have replied No in all questions from 6 to 8, please mind the application/project and proceed with the following questions.

Funded by
the European Union

1) Does the processing operation involve evaluation or scoring?[21] □ Yes
□ No

2) Does the processing operation allow automated decisions producing legal or similar significant effects on data subjects?[22] □
Yes □ No

3) Does the processing operation involve systematic monitoring?[23] □ Yes
□ No

4) Is "sensitive data" processed?[24] □ Yes
□ No

5) Is the processing operation on a "large scale"?[25] □ Yes
□ No

6) Have datasets been matched or combined?[26] □ Yes
□ No

7) Does the processing operation involve data concerning vulnerable data subjects?[27] □ Yes
□ No

[21] Evaluation or scoring, including profiling and predicting, especially from "aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behavior, location or movements" (recitals 71 and 91). Examples of this could include a bank that screens its customers against a credit reference database, or a biotechnology company offering genetic tests directly to consumers in order to assess and predict the disease/health risks, or a company building behavioural or marketing profiles based on usage or navigation on its website.
22 Automated-decision making with legal or similar significant effect means processing that aims at taking decisions on data subjects producing "legal effects concerning the natural person" or which "similarly significantly affects the natural person" (Article 35(3)(a)). For example, the processing may lead to the exclusion or discrimination against individuals. Processing with little or no effect on individuals does not match this specific criterion.
23 Systematic monitoring means processing used to observe, monitor or control data subjects, including data collected through "a systematic monitoring of a publicly accessible area" (Article 35(3)(c)). This type of monitoring is a criterion because the personal data may be collected in circumstances where data subjects may not be aware of who is collecting their data and how they will be used. Additionally, it may be impossible for individuals to avoid being subject to such processing in frequent public (or publicly accessible) space(s).
[24] "Sensitive data" includes special categories of data as defined in Article 9 (for example information about individuals' political opinions), as well as personal data relating to criminal convictions or offences. This criterion also includes data which may more generally be considered as increasing the possible risk to the rights and freedoms of individuals, such as electronic communication data, location data, financial data (that might be used for payment fraud). In this regard, whether the data has already been made publicly available by the data subject or by third parties may be relevant. The fact that personal data is publicly available may be considered as a factor in the assessment if the data was expected to be further used for certain purposes.
25 The GDPR does not define what constitutes large-scale, though recital 91 provides some guidance. In any event, the following factors, in particular, be considered when determining whether the processing is carried out on a large scale:
a. the number of data subjects concerned, either as a specific number or as a proportion of the relevant population;
b. the volume of data and/or the range of different data items being processed;
c. the duration, or permanence, of the data processing activity;
d. the geographical extent of the processing activity.
26 Datasets that have been matched or combined are those for example originating from two or more data processing operations performed for different purposes and/or by different data controllers in a way that would exceed the reasonable expectations of the data subject.
27 The processing of this type of data can require a DPIA because of the increased power imbalance between the data subject and the data controller, meaning the individual may be unable to consent to, or oppose, the processing of his or her data. For example, employees would often meet serious difficulties to oppose to the processing performed by their employer, when it is linked to human resources management. Similarly, children can be considered as not able to knowingly and thoughtfully oppose or consent to the processing of their data. This also concerns more vulnerable segment of the population requiring special protection, such as, for

8) Does the processing operation involve new technological or organisational solutions?[28] □ Yes
□ No

9) Does the processing operation prevent data subjects from exercising a right or using a service or a contract?[29] □ Yes
□ No

If you have replied Yes to at least two of the above 9 questions, proceed with the DPIA.

example, the mentally ill, asylum seekers, or the elderly, a patient, or in any case where an imbalance in the relationship between the position of the data subject and the controller can be identified.

28 Innovative use or applying technological or organizational solutions, like combining use of finger print and face recognition for improved physical access control, etc. The GDPR makes it clear (Article 35(1) and recitals 89 and 91) that the use of a new technology can trigger the need to carry out a DPIA. This is because the use of such technology can involve novel forms of data collection and usage, possibly with a high risk to individuals' rights and freedoms. Indeed, the personal and social consequences of the deployment of a new technology may be unknown. A DPIA will help the data controller to understand and to treat such risks. For example, certain "Internet of Things" applications could have a significant impact on individuals' daily lives and privacy; and therefore require a DPIA.

29 This includes processings performed in a public area that people passing by cannot avoid, or processings that aims at allowing, modifying or refusing data subjects' access to a service or entry into a contract. An example of this is where a bank screens its customers against a credit reference database in order to decide whether to offer them a loan.

Funded by
the European Union

4. The DPIA

What type of personal data is processed?

□ Personal data □ Special categories of personal data □ Judicial data

_____
_____
_____
_____
_____
_____

Which are the recipients of personal data?

□ Authorised persons □ Data processors □ Autonomous data controllers

(details)_____
_____
_____
_____
_____
_____

Indicate the period for which the personal data is stored, and the reasons for determining such period, or if that is not possible, the criteria used to determine that period

_____
_____
_____
_____
_____
_____
_____

Indicate the assets through which the application/project process personal data

□ Hardware □ Software □ Processing not by automated means

□ Personnel belonging to a specific department

Funded by the European Union

(details)_____
_____
_____
_____
_____
_____

The application/project process data under the rules approved in a code of conduct pursuant Article 40 GDPR.

□ Yes □ No

(details)_____
_____
_____

Are purposes of the processing indicated in Answer 5 specified, explicit and legitimate?

□ Yes □ No

(details)_____
_____
_____

Has the legal basis for the processing been identified? (if YES, describe the legal basis for each of the purposes listed in answer 5 in the Preliminary Questions)

□ Yes □ No

_____
_____
_____
_____
_____

Is data processing adequate, relevant and limited to what is necessary in relation to the purposes for which data is collected ('data minimization')?

□ Yes □ No

_____
_____
_____
_____
_____

Funded by the European Union

Has adequate information notice been provided to the data subject? (Artt. 13 and 14 GDPR)

□ Yes □ No

_____
_____
_____
_____
_____

Are data subject's rights respected?

Right of access □ Yes □ No

Right to data portability □ Yes □ No

Right to rectification □ Yes □ No

Right to erasure □ Yes □ No

Right to restriction □ Yes □ No

Which are the conditions for transfers of data to third countries or international organizations?

□ Data is not transferred □ Standard Model Clauses/other binding agreements

□ Adequacy Decision □ BCRs □ Other Derogations (Art. 49 GDPR)

_____
_____
_____
_____
_____

Funded by the European Union

5. The risk assessment

The following analysis takes into account the high risk that the application/project may cause to the rights and freedoms of the data subjects. For each risk, the impact and probability of occurrence shall be indicated.

1. Impact shall be a number between 1 to 4 (1 – no real impact; 2 - small impact; 3 – normal impact; 4 – high impact);
2. Probability shall be a number between 1 to 4 (1 – improbable; 2; unlikely to occur; 3- very probable; 4 – highly probable);

"Result" shall be the multiplication of "Impact" and Probability".

Id	Risk:	Impact	Probability	Result
1	Discrimination			
2	Identity theft or misuse			
3	Financial loss			
4	Loss of reputation			
5	Loss of confidentiality of personal data (in particular if protected by professional secrecy) (i.e. data breach)			
6	Loss of data (i.e. ransomware, zero day attacks)			
7	Unauthorized deciphering or compromised algorithm			
8	Other significant social or economic damages			
9	Loss of freedoms or rights			

10	Impossibility of exercising control over data			
11	Illegal access to special categories of personal data or security measures to protect that protect data			
12	Profiling mechanisms unclear or update without communication to the data subject			
13	Processing of data related to vulnerable persons (i.e. patients, workers, minors)			
14	Large scale processing			

Select the Results higher than 8, identify security and/or organizational measures to reduce the risk and report the new result based on the application of such measures. If the New Result is still higher than 8 you have to proceed with the prior consultation.

Risk	Measure to reduce the risk	New Result	Requiring prior consultation?

Funded by the European Union

Funded by
the European Union

Appendix IV. Letter from TMU to SmartCHANGE consortium and Regulations on the Production and Management of Electronic Medical Records by Medical Institutions

Funded by
the European Union

To

Mitja Lustrek
Coordinator SmartCHANGE Consortium
Institut Jozef Stefan, Ljubljana, Slovenia.

Sub: Clarifications on Research Database from Taipei Medical University

The source of the data for the TMU is from National Health Insurance Research Database (NHIRD) covering 23 million individuals across all age groups. Prior to being transferred to NHIRD, the data undergoes through anonymization by the hospitals.

The dataset from NHIRD is anonymous because it has been processed to remove or modify personal identifiers, ensuring that individuals cannot be readily identified by the data. This anonymization process typically involves several key steps:

Removal of Direct Identifiers: All direct personal identifiers, such as names, addresses, phone numbers, email addresses, social security numbers, and medical record numbers, are removed from the dataset.

De-identification of Indirect Identifiers: Indirect identifiers that could be combined with external information to identify individuals, such as rare diseases, unique treatment regimens, or specific age ranges, are either removed or aggregated to prevent re-identification.

Data Perturbation: In some cases, values of the dataset may be altered slightly (perturbation) to prevent identification. This can include adding random noise to certain data points or modifying exact values to a range (e.g., changing date of birth to an age range).

Data Aggregation: Data is typically compiled into group-based summary statistics rather than individual records. For example, comorbidities are presented as a collective aggregation of distinct illnesses.

Risk Assessment and Testing: After the above steps, a risk assessment is typically performed to evaluate he likelihood of re-identification. Additional measures might be deemed too high. For instance, all the patients with rare diseases will be eliminated from the data that is shared with NHIRD.

Compliance with Standards and Legislation: The anonymization process is conducted in accordance with relevant data protection standards and legislation, such as GDPR or HIPAA, which provide guidelines on how to handle personal data and de-identification procedures.

By implementing these measures, NHIRD ensures that the dataset can be used for research purposes without compromising the privacy and confidentiality of the patients whose information was recorded. This allows researchers to analyze the data for patterns and trends in healthcare without the risk of infringing on individual privacy.
Prof. Shabbir have been the principal investigator for several international (European Horizon 2020) projects including SmokefreeBrain (GA:681120), CrowdHEALTH (GA: 727560), CATCH-ITN (GA: 722012), and current project iHelp (GA: 101017441). He possesses extensive experience in managing and sharing of clinical datasets with European partners. I wish to highlight that TMU is fully committed to adhering to international best practices in data sharing and management.

Best regards,

Prof. Marc Hsu
Chief Data Officer,
Taipei Medical University

Marc Hsu.

臺北醫學大學
TAIPEI MEDICAL UNIVERSITY

Regulations on the Production and Management of Electronic Medical Records by Medical Institutions

Chapter 1: General Provisions

Article 1
These regulations are established in accordance with Article 69 of the Medical Care Act (hereinafter referred to as this Act).

Article 2
Medical institutions that produce and store medical records in electronic document form (hereinafter referred to as electronic medical records), which comply with the provisions of these regulations, are exempt from having to produce them in written form.

Article 3
Medical institutions implementing electronic medical records shall establish an electronic medical records information system (hereinafter referred to as the system) and have the following management mechanisms:

1. Standard operating mechanism: Standard operating procedures for system setup, maintenance, and audit.
2. Access control mechanism: Control of production, access, addition, deletion, consultation, copying, transmission, and other usage rights of electronic medical records.
3. Emergency response mechanism: Prevention, notification, response, recovery, and other emergency measures for system failures.
4. System security mechanism: Ensuring system safety, accurate timing, system redundancy, data backup, and other protective measures.
5. Transmission encryption mechanism: Encrypting the network transmission of electronic medical records using encryption mechanisms recognized by the International Organization for Standardization.
6. Security incident response mechanism: Prevention, notification, response, review, and corrective measures for system intrusions, data breaches, damage, or other security incidents.

Records shall be made for the execution of each of the management mechanisms mentioned in the preceding paragraph and kept properly for at least five years.

Article 4

Other protective measures of the system security mechanism mentioned in Item 4 of Paragraph 1 of the previous article shall include the following:

1. Verification of user identity.
2. Concealment of personal data display or other appropriate protective measures.
3. Verification and confirmation procedures for system development, launch, maintenance, and application software.
4. Monitoring measures for system use and data access.
5. Preventive measures for network intrusion systems.
6. Response measures for illegal or abnormal use.

Article 5

When a medical institution encounters a security incident as described in Item 6 of Paragraph 1 of Article 3, it shall notify the individual or their legal representative in the manner and content prescribed in Article 22 of the Enforcement Rules of the Personal Data Protection Act.

If a security incident affects the operation of the medical institution or the rights and interests of the individual, the medical institution shall report it to the competent authority of the municipality or county (city) within seventy-two hours after becoming aware of the incident.

Article 6

The system mentioned in Paragraph 1 of Article 3 may be established and managed by colleges and universities, legally registered or filed legal persons, institutions, or organizations (hereinafter collectively referred to as the entrusted institution), with the medical institution bearing the responsibility prescribed by this Act and these regulations.

The entrustment by the medical institution mentioned in the preceding paragraph shall be stipulated in a written contract. However, under any of the following circumstances, a written contract may be exempted:

1. Entrustment to other affiliated hospitals of the medical corporation or other legal persons.
2. Entrustment to other affiliated hospitals of the school.
3. Entrustment to other hospitals established by the institution.
4. The entrusted institution mentioned in Paragraph 1 shall pass the information security standard certification recognized by the central competent authority and

have documentary proof.

Article 7
The contract mentioned in Paragraph 2 of the previous article shall specify the following:

1. The scope of the entrustment.
2. The rights and obligations of the entrusted institution.
3. The entrusted institution shall comply with the provisions of Articles 3 to 4, Article 8, and Articles 13 to 16.
4. If the entrusted institution uses a system or resources not developed by itself, the source and proof of authorization must be specified.
5. Measures for the protection of patient privacy and the confidentiality and security of data.
6. The entrusted institution shall follow the standard operating procedures, risk management, internal control, and audit systems stipulated by the entrusting institution.
7. Causes for termination and dissolution of the contract and the data handling mechanism.
8. The entrusted institution agrees that the competent authority may obtain related documents, data, or reports within a designated period for the entrusted matters.
9. The entrusted institution shall immediately notify the entrusting institution upon discovering any information security anomalies or deficiencies in the entrusted matters.
10. The entrusted institution shall not further entrust a third party with the entrusted matters. However, with the consent of the medical institution and specification in the contract of the re-entrustment matters, duration, and re-entrusted party, and provided that it does not exempt the original entrusted institution from its obligations, this limitation does not apply.
11. Other matters designated by the central competent authority.

Article 8
When collecting, processing, and using system data and utilizing the database, and employing cloud services or entrusting the entrusted institution to provide cloud services, medical institutions shall proceed in accordance with the following regulations:

1. Take appropriate risk control measures.

2. Implement measures to avoid interruption of medical services.
3. Supervise cloud service providers and, as needed, entrust the entrusted institution or other professional institutions to assist in supervision.
4. Mechanisms for data transfer back to the entrusting institution or another cloud service provider when stopping or terminating cloud services.
The location for data storage of cloud services mentioned in the preceding paragraph shall be set within the territory of our country. However, in special circumstances and with the approval of the central competent authority, this limitation does not apply.
Providers of cloud services mentioned in Paragraph 1 shall pass the information security standard certification recognized by the central competent authority and have documentary proof.

Article 9
Medical institutions implementing electronic medical records shall specify the start date and scope of implementation and attach the contract mentioned in Paragraph 2 of Article 6 and the certification of passing the verification mentioned in Paragraph 3, and report to the competent authority of the municipality or county (city) for filing within fifteen days from the date of implementation; the same applies when changing the scope of implementation, the entrusted institution, or stopping the implementation.

Article 10
When medical institutions implementing electronic medical records undergo hospital accreditation or declare payments for National Health Insurance, the competent or organizing authority for hospital accreditation or National Health Insurance shall not require the provision of printed or photocopies of electronic medical records without special reasons.

Chapter 2: Production and Signature of Medical Records

Article 11
Medical institutions producing electronic medical records shall comply with the following regulations:

1. Entry of identification codes or other identification methods, after verification of identity and permissions by the computer system, may proceed.
2. When adding or deleting electronic medical records, it should be clearly distinguishable from before the addition or deletion, and personal usage records and

date data should be preserved.
3. Signing or stamping as required by Article 68, Paragraph 1 of this Act, shall be done with an electronic signature.
4. After the production of the medical record, the electronic signature should be completed within twenty-four hours.
5. After electronic signing, archiving and backup should be performed.
If medical personnel cannot complete the electronic signature within the time limit mentioned in Item 4 of the preceding paragraph due to reasons, the medical institution should use the medical institution certificate signature as a substitute; except in special circumstances, medical personnel should complete the supplement signature afterwards.

Article 12
Electronic signatures, in addition to the regulations of Paragraph 2 of the previous article, shall be made with a medical personnel certificate issued by the central competent authority. However, if the medical institution has other signature methods that comply with the regulations of the Electronic Signature Act, they may follow such methods.
The issuance, reissuance, and replacement of the medical institution certificate mentioned in Paragraph 2 of the previous article, the medical personnel certificate, and its associated cards, spare cards, shall be handled by the central competent authority itself or entrusted to private organizations, and fees may be charged; the amount of the fees shall be in accordance with the regulations of the medical certificate charging standards.

Chapter 3: Storage, Destruction, and Exchange

Article 13
During the period of preservation of medical records prescribed by Article 70, Paragraph 1 of this Act, the access, addition, deletion, consultation, copying, and other related matters of electronic medical records, as well as the personnel, time, and content involved, should be fully recorded.

Article 14
When transferring electronic medical records for preservation by a recipient according to Article 70, Paragraph 2 of this Act, the reason, target, method, time, place of the transfer, and the legal basis for the recipient to possess the electronic medical records should be recorded and preserved by the recipient for at least five

years.

Article 15
Medical institutions storing electronic medical records in computers, automated machines, or other electronic media (hereinafter collectively referred to as storage media) should take appropriate measures to ensure the complete removal or clearing of electronic medical record data without the risk of leakage when scrapping, replacing, or repurposing the storage media; if the storage media cannot be completely removed, cleared, or the data restored afterwards, physical destruction should be carried out to make it unusable.

Article 16
When destroying electronic medical records according to Article 70, Paragraph 2 and 4 of this Act, the personnel, method, time, and place of destruction should be recorded, and the record preserved for at least five years; the same applies to outsourced destruction.
The destruction mentioned in the preceding paragraph should be supervised throughout the entire process to confirm complete destruction, and photographic evidence should be taken.

Article 17
Medical institutions may convert the following data into electronic files for preservation; after conversion, the content of the electronic files should be reviewed for consistency with the originals, and sealed with the medical institution certificate signature, then reported to the competent authority of the municipality or county (city) for filing, and thus regarded as electronic medical records:

1. Documents that should be consented to in writing according to this Act or other medical regulations and preserved together with medical records.
2. Paper medical records existing before the implementation of electronic medical records by the medical institution.
3. Other documents and data that must be preserved with medical records according to legal regulations.
The originals mentioned in the preceding paragraph may be exempt from being preserved in paper form and are not subject to the preservation period limit set by Article 70, Paragraph 1 of this Act.
When destroying the original paper documents and data mentioned in Paragraph 1, the medical institution should record the details, method, time, and place of

destruction, and preserve the record for at least five years.

Chapter 4: Supplementary Provisions

Article 18

The central competent authority or institutions and public or private institutions approved by the central competent authority may establish electronic medical record exchange platforms for medical institutions to conduct cross-institutional electronic medical record exchanges or use.

When conducting electronic medical record exchanges or use, medical institutions should use the platforms mentioned in the preceding paragraph. However, medical institutions using the same system as specified in the proviso to Article 6, Paragraph 2, are not subject to this limitation.

The electronic medical record exchange format, signature, timestamp, and other related matters of Paragraph 1 shall be announced by the central competent authority.

Article 19

When medical institutions exchange or use electronic medical records through the exchange platform mentioned in Paragraph 1 of the previous article, they must obtain the patient's consent before proceeding. However, in emergency situations where consent cannot be obtained or obtained in time, this limitation does not apply.

When the patient is a fetus, the mother's consent is required; for persons with limited capacity to act or those declared to be under assistance, the consent of the person themselves and their legal representative or assistant is required; for persons without capacity to act or those declared to be under guardianship, the consent of their legal representative or guardian is required.

When the patient is an adult without the capacity to express intent and consent cannot be obtained in accordance with the provisions of the preceding paragraph, the consent of relatives or related persons should be obtained.

Article 20

Documents that should be consented to in writing according to this Act or other medical regulations and preserved together with medical records may be done electronically in accordance with the provisions of the Electronic Signature Act; and may be provided in paper form or electronically upon the request of the relevant party.

Article 21
The protection of personal data in electronic medical records not specified in these regulations shall be in accordance with the Personal Data Protection Act, the Implementation Regulations for the Security Maintenance Plan of Hospital Personal Data Files, and other relevant laws and regulations.

Article 22
Before the amendment and enforcement of these regulations on July 18 of the 111th year of the Republic of China, medical institutions that have already entrusted an entrusted institution to establish and manage the system shall, within one year from the date of the amendment and enforcement, proceed in accordance with the provisions of Article 9.
Before the amendment and enforcement of these regulations on July 18 of the 111th year of the Republic of China, institutions and public or private institutions that have already established electronic medical record exchange platforms shall, within one year from the date of the amendment and enforcement, obtain approval according to Paragraph 1 of Article 18 and comply with the provisions of Paragraph 3 of Article 18.