Pseudonymisation in the context of GDPR-compliant medical research

Ioannis Basdekis
SPHYNX Technology Solutions AG
Zug, Switzerland
0000-0002-4873-7920

Christos Kloukinas
Dept. of Computer Science, City, University of London
London, United Kingdom
0000-0003-0424-7425

Carlos Agostinho
Center of Technology and Systems, UNINOVA
Portugal
0000-0002-2884-776X

Ioannis Vezakis
SPHYNX Analytics Ltd
Nicosia, Cyprus
0000-0003-4976-4901

Andreia Pimenta
Secretaria Regional de Saúde e Proteção Civil, SRS
Madeira-Portugal
andreia.pimenta@madeira.gov.pt

Luigi Gallo
Institute for High Performance Computing and Networking National Research Council
Italy
0000-0002-1281-404X

Georgios Spanoudakis
SPHYNX Technology Solutions AG Zug, Switzerland

***Abstract*—** Pseudonymisation is a data protection technique often used to protect the privacy of individuals when their personal data are being used for research purposes. Not only is it a key ingredient of the General Data Protection Regulation (GDPR) that requires organisations to ensure that the personal data they process is handled in a secure manner, but it is particularly important in assisting medical research given that often relies on sensitive personal data, since it reduces the risk that medical data could be misused or mishandled. For managing their medical data, it is important to ensure that such data are protected against unauthorised access, and can be reutilised in an anonymous fashion, while still authorised personnel is able to identify the study participant that some data belong to (e.g., for personalised interventions, technical alerts, technical support). In addition, the re-identification of a study participant is a pre-requisite for exercising their rights under the GDPR, since it assists organisations in meeting GDPR requirements (such as the right to access, rectify and portability of data). We argue that the application of pseudonymisation is particularly effective when considered during the early stages (Privacy by Design) of digital services implementation, as well as when defining the complementary to these organizational procedures. Aim of this paper is to present the way in which the pseudonymisation mechanism of the SMART BEAR H2020 project supports the triptych of research activities conducted within the context of an observational medical study, legal obligations arising from the regulatory framework for the protection of personal data, and reutilisation of data for research purposes. Evidence-based security and privacy assessments will be conducted on two different H2020 projects to evaluate such privacy practice.

***Keywords*—** pseudonymisation, privacy, data minimisation, GDPR, observational studies

I. Introduction

Healthcare organisations are faced with a range of different cybersecurity threats, ranging from malicious actors exploiting inherited technical vulnerabilities to accidental data breaches due to human errors [1]. In one way or the other, these threats can lead to the loss of sensitive patient information, financial losses, and even legal repercussions for an organisation [2]. Those are even more substantial within the EU area of "health research", considering that in addition to requirements imposed by the General Data Protection Regulation (GDPR), more specific safeguards must be applied whenever personal data are processed for the purposes of health research and for related matters (e.g., [4] Ireland, [5] Germany, [6] Hellas and [7] Italy). To make things worse for healthcare organisations, in addition to paying attention and coping with shortfalls within their local area of responsibility due to their participation in the context of clinical trials, observational studies and participating in large-scale studies, their information systems transmit or receive medical data of study participants by third-party systems, which might contain vulnerabilities that cannot be addressed by "least privilege" access management methods.

The anonymisation of data so that the individuals who are the subjects of the data are not identifiable, although an effective method, does not seem applicable when conducting medical studies. If applied, from a clinical perspective, individual research results which should be offered to participants cannot be associated back to the individual [2]. Especially in cases where devices (e.g., mIoT, wellness devices, smartphones, sensors) are used by patients, some sort of matching with them must be in place, an element that allows the monitoring of their good use. Equally important from an organisational and technical perspective, requests associated with GDPR rights (e.g., data transferability) cannot be answered. As such, the main recommended way [8] is to apply pseudonymisation. In this paper, we present the pseudonymisation approach implemented within the context of the EU-funded SMART BEAR (SB) H2020 project [9], which, along with compliance with legal requirements and privacy obligations laid down by GDPR to conduct personalised analysis and provide feedback to study participants about their condition, supports the exercise of all GDPR rights.

II. Background and Rationale

A recent comparison of approaches taken for medical research [23] identifies three main types of systems developed: "distributed data analyses", "secure multi-party computation protocols", and "data enclaves". The SB project is of the third type, where data are collected into a single, secure system where analyses are performed. Given the highly confidential nature of the data used in medical research, whenever attempting to design a system for storing and processing such data, designers try to balance a set of requirements. On the one hand, there are the legal requirements concerning privacy, as those described by GDPR [3]. These can be easily met through anonymisation [24][12] – removing any personally identifying data from the

Quasi-final version –Published as: https://doi.org/10.1109/DRCN57075.2023.10108370

data repository, so that the stored data cannot be linked back to the original individuals from which they were obtained. Anonymised data still allow researchers to conduct research, e.g., by computing various statistics using the collected data. However, there are often requirements for healthcare personnel to provide healthcare to the study participants, if the data identify the need for so. This requirement imposes the need for the healthcare personnel to be able to identify the study participant that some data belong to. The same need arises when it is possible for the study participant to have technical issues (sometimes unbeknownst to them) that require assistance – there needs to be a way to re-identify them so as to aid them. As such, complete anonymisation of data is not an option (unless healthcare/technical provisions are not considered at all), and pseudonymisation is required instead. While anonymisation removes all links from the data back to the individual, pseudonymisation allows for the individual to be identified through some of the data [24][11][3][12].

The most common pseudonymisation techniques used in healthcare are identifier replacement and hashing (with or without an additional key – also known as a "salt") [11][8]. The former replaces data identifying an individual with some unique identifier (e.g., a monotonic counter) that does not directly reveal the identity of the individual, while the latter replaces such data with a unique cryptographic value that is computed by a one-way (hash) function (over the data alone or over the data and an extra key, or "salt" as it is sometimes called). Other pseudonymisation techniques used in healthcare include tokenisation and encryption [11][8]. Replacement and plain hashing are deterministic methods, producing the same results when given the same inputs, while hashing with keys and encryption are generally nondeterministic methods, since for the same identifying information they can produce different outputs (the output now depends on extra input data, such as the key). Depending on the type of data that is pseudonymised, replacement/plain hashing are not considered as strong as hashing with keys and encryption, since an attacker can reverse them relatively easily (e.g., through so-called "dictionary attacks") [8][12]. Accordingly, it should be used in conjunction with other techniques, such as encryption. In a project such as SB, where one wishes to study how an individual's data evolve during time, having a deterministic pseudonymisation method is important to be able to relate the same individual's data at different time points – otherwise, these will appear as data from different individuals. Of course, it should be noted that this is not a general scenario, and non-deterministic pseudonymisation methods may be acceptable in other settings. It all depends on the type of analysis one wishes to perform and how one plans to share data (or not) with other organisations.

The literature [23][11] has considered a few different scenarios where pseudonymisation can be used, depending on whether the data controller (who decides on types of data processing), the data processor (who performs the processing), or indeed a third party are responsible for the pseudonymisation (and the relations between these three roles, e.g., whether the controller and the processor are the same real entity). In all of them, the basic task required is to (a) separate the data that can identify a study participant (Personal Identifying Information – PII) from the rest, and (b) to provide a non-directly identifying value for each participant that can be linked back to their PII, through one of the aforementioned techniques (tokenisation, hashing, etc.).

All the systems employing pseudonymisation schemes reported in the literature aim to achieve confidentiality [15][16][17][18][19][20] both at rest (through data encryption in their database) and in transit (usually through the use of a classic encrypted communication protocol like TLS). Some of the systems try to offer further properties, e.g., authorisation [15][18][20] or accountability [16][17][19] – see Kohlmayer et al. [21] for a recent survey. Authorisation itself is usually achieved through the usage of roles and the requirement of each user to log into the system using an appropriate role before being allowed to perform certain tasks, while accountability is usually achieved by keeping logs of which user performed which action at what time and with what data.

III. SMART BEAR Pseudonymisation

One of the aims of the SB project is to provide an intelligent and personalised digital solution for sustaining and extending healthy and independent living by implementing an affordable, accountably secure and privacy-preserving platform with off-the-shelf smart and medical devices, to support the healthy and independent living of elderly people with five prevalent health-related conditions (hearing loss, cardiovascular diseases, cognitive impairments, mental health issues and balance disorders), as well as Frailty [13]. From the technical perspective, this will be achieved through intelligent, evidenced-based interventions on lifestyle, medically significant risk factors, and chronic disease management, enabled by the utilisation of continuous and objective medical and environmental sensing, assistive technologies and big data analytics. To support this goal, pseudonymised data (including personal health records, questionnaires, and devices usage data) stored in the SB@Cloud are subjected to different types of analysis (e.g., statistical, ML) to obtain the evidence needed to offer personalised interventions promoting their healthy and independent living. To address heterogeneity in the data and lack of shared semantics across sources, SB project leverages widely adopted ontologies and standards, and developed extensions to model relevant knowledge in the domains of the project for which no standards exist [10]. To provide the specification of the semantically harmonised information in the data model, we leverage the HL7 FHIR standard [29], and use FHIR resource profiles to define constraints and extensions to the FHIR base model capturing the required information and semantics (Figure 1 portrays the conversion of the raw blood pressure measurements to FHIR).

The architecture of the SB@Cloud, by virtue of its design, supports privacy by adhering to the "Privacy by Design" principle [10], and consists of the following seven components: i) the SB@Dashboard responsible for the interaction between the end-users and the services in the backend; ii) the SB@Big Data Analytics Engine responsible of the execution of validated analytics workflows on the data collected by the platform; iii) the SB@Data Repository built around the HL7 FHIR standard [29], responsible for storing the data transmitted by the HomeHub, the mobile application, and other external sources (e.g., devices/services associated to HoloBalance and Smart4Health projects), structuring and disposing clinical information using FHIR standard as specification (also SNOMED CT [30] for medical terminology not fully covered by FHIR, as well as ICD-9 [31], and LOINC [32]); iv) the SB@Decision Support System that realises the process of deciding upon the adoption and execution of interventions driven by the decision models; v) the SB@Security Component allowing interactions between components and supporting, authentication, authorisation and pseudonymisation; vi) the SB@Mobile application supporting the data acquisition from the various (m)IoT devices and a HomeHub component responsible for

connecting sensors that are tethered to a participant's home and for serving as the main data-transmission point to the SB@Cloud main repository; and vii) the SB@Security and Privacy Assurance Platform via which hybrid security and privacy assessments will be conducted, to provide a comprehensive analysis of the security and privacy posture of the whole platform.

Blood pressure
OMRON M7
Raw data

Standing Blood pressure systolic/diastolic

Systolic: 119 mm[Hg]
Diastolic: 79 mm[Hg]
Creation: timestamp

Pseudo-Id1: Patient001
IMEI: 111111111111110

FHIR transformation


```
{"resourceType":"Observation","id":"BloodPressureStanding001",
 "meta":{"lastUpdated":"2021-02-20T09:29:23.356+00:00",
 "profile":["https://smart-
bear.eu/fhir/StructureDefinition/BloodPressureStandingObservation"]
 },"text":{"status":"extensions",.....},
 "extension":[{"url":"https://smart-
bear.eu/fhir/StructureDefinition/LoadedIntoApp",
   "valueDateTime":"2021-02-20T09:29:23.356+00:00"}],
 "status":"final",
 "code":{"coding":
            [{"system":"http://loinc.org",
            "code":"85354-9",
            "display":"Blood pressure panel with all children optional"}]},
 "subject":{"reference":"Patient/Patient001"},
 "effectiveDateTime":"2021-02-20T09:29:23.356+00:00",
 "bodySite":{"coding":[{
            "system":"http://snomed.info/sct",
            "code":"368209003",
            "display":"Right arm"}]},
 "device":{"reference":"Device/111111111111110"},
 "component":[{"code":
              {"coding":[{"system":"http://loinc.org",
                          "code":"8460-8",
                          "display":"Systolic blood pressure--
standing"}]
            },
            "valueQuantity":{"value":119,
            "system":"http://unitsofmeasure.org",
            "code":"mm[Hg]"}
            },
            {"code":{"coding":[{"system":"http://loinc.org",
              "code":"8454-1",
              "display":"Diastolic blood pressure--standing"}]},
            "valueQuantity":{"value":79,
              "system":"http://unitsofmeasure.org",
              "code":"mm[Hg]"}}]}
```


Figure 1: Example of raw data to FHIR transformation (blood pressure measurements generated by OMRON M7)

The SB pseudonymisation constitutes the backbone of all processes upon the data, the implementation of which fundamentally affects the way in which data is managed by all services while ensuring that all records cannot be attributed to a specific data subject without having previously been associated with additional information, a process that can be performed by specific end-user roles. The Security Component is the component which provides mechanisms and services to secure communication between all platform's components (intranet), allows access to APIs based on a policy (e.g., RBAC), assists the logging of activities when required (e.g., evidence for security and privacy assessments, for GDPR-related audits), but most importantly performs pseudonymisation by altering the identifiers associated to the study participants and valid smartphones (by checking whether the IMEI is associated to the specific patient) they are utilising, filtering out PII, and in turn to validate usage records transmitted based on specific requirements (i.e., valid pseudoid1 and IMEI pair, switching back and forth the identifiers pseudo-Id1 and pseudo-Id2). All personal data (i.e., identifiers, contact info, and PIIs) are stored encrypted in a separate repository, ensuring privacy as well as better managing their data and maintaining a better audit trail for a significantly smaller data size. By keeping personal data in a separate repository, users can ensure that their data is secure and private, is manageable, and a better audit trail can be maintained.

The SB pseudonymisation technique that governs all data transfers is as follows:

A. Ids and PII management: each participant is associated with two identifiers (i.e., Pseudo-Id1, Pseudo-Id2), as are the devices assigned to them in the context of the project's study (e.g., Pseudo-Id1 - smartphone's IMEI pairs). In addition, in cases of common participants in joint studies of other projects, which are HoloBalance (HB) and Smart4Health (S4H) in the case of the SB project, participants' identifiers with which they have been associated in their context are also collected (e.g., S4HID or HBID, as reported in Figure 2).

B. For all data transmitted to the SB@Cloud (i.e., entered manually through the Dashboard, m2m transmissions through a smartphone or an external system of a synergetic project), records containing an identifier associated with a unique study participant (i.e., Pseudo-Id1, S4HID or HBID), as well as PIIs, these identifiers will be replaced by another identifier (Pseudo-Id2), and PIIs will be removed before the data is stored in the repository. This process also varies the exact dates associated with an individual (e.g., birthday), as these could potentially lead to his/her identification (dates are replaced by the date of the Sunday that follows it, e.g., "1/01/1980" will be replaced by "6/01/1980") and can also support the deletion of some sensitive values if included in transmissions, e.g., first name.

C. In cases of micro-aggregated data pulled from an External Vendor's cloud (e.g., Garmin smartwatch), these are associated with the same Pseudo-Id1 within the scope of the smartphone (SB@Mobile Application), later to be replaced by Pseudo-Id2 via the same mechanism (as B).

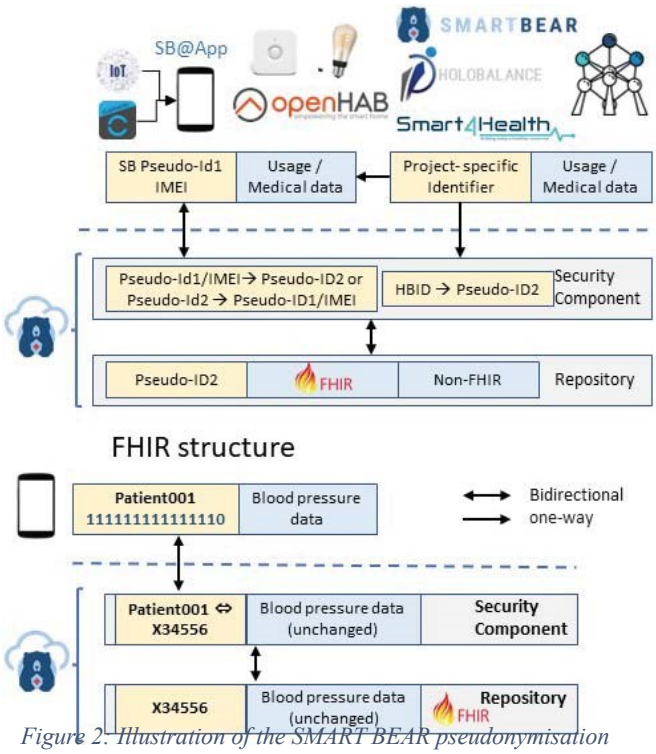

Figure 2: Illustration of the SMART BEAR pseudonymisation mechanism

D. The reverse process is performed for all data transmitted from the SB@Cloud to a smartphone for personalised notification delivery. Any personalised message generated as a result of an internal process (e.g., personalised intervention, technical alert) refers to the internally used Pseudo-Id2 identifier, which is then replaced with Pseudo-Id1 before transmitting, while at the same time, the valid smartphone identifier is included (to avoid sending to stolen or replaced smartphones).

E. Trailed access to mapping data is allowed only to authorised end-users.

The aforementioned processes A-E are performed by the Security Component (as depicted in Figure 2). It is a SB@Cloud component that, during every API REST triggering, allows the transmission of data into the SB@Cloud and its components, with the use of appropriate identifiers as the case may be.

IV. Procedures for addressing GDPR-related requests in Pilot Sites

Since the enforcement of the GDPR, data privacy and data security have become two essential components for the data protection strategy of any organisation processing personal data in the EU [25]. In most cases, organisations are no longer able to use individuals' personal data without their consent (except in cases where another legal basis is applicable). In addition, GDPR compliance requires the protection of privacy at various levels [26], introducing several compliance obligations related to the organisational context of a medical study, such as: explicit informed consents; support of the "right to be forgotten" and "data transferability"; the assignment of a Data Protection Officer (DPO); the obligation to report a data breach to GDPR Supervision authorities within a short period of time of having become aware of it, in addition to a stricter sanctioning regime for non-compliance [25]. On a more practical level, and beyond any technical implications of the application of pseudonymisation in the management of personal and medical data supported by information systems (such as the SB platform), there are other aspects in the context of tracking study participants where pseudonymisation is also a catalyst in the protection of privacy. Having entered the digital health age, a wide range of e-health devices and tools (e.g., mIoT, wellness devices/sensors) have emerged, and a growing number of hospitals and healthcare institutions are embracing information and communication technologies to support and advance their healthcare practices. Although digital health has many meaningful applications, its high dependence on sensitive information concerning patients' health can inevitably trigger data security problems even outside the confines of a hospital [27]. Therefore, in the Digital Health domain, as in the SB project case, data protection is critical due to processing highly sensitive personal data and doing so outside the protective framework of a (single) hospital.

On the one hand, there are many exposed (i.e., whose operation does not have special safety features/defences) devices in homes and organisations, which, if left unsecured and vulnerable to breaches, can produce substantial adverse personal and social impacts on patients and cause huge financial costs [2]. In healthcare, unsecured medical Internet of Things (IoT) devices might contain confidential patient information, test results, and medical images. On the other hand, taking into account that usually medical study participants do not have much experience in utilising modern devices and sensors, and even less familiarity with activating complex authentication mechanisms (where applicable), a system that digests such usage data cannot consider the remote use of such devices as absolutely guaranteed in terms of security and privacy. Even at the purely organisational level, data collection forms (e.g., consent forms, questionnaires), communications for other than medical reasons (e.g., accounting, technical support), and other non-medical scenarios should be implemented with knowledge of the least possible information about a study participant. Within this context, the use of pseudonymisation provides a balanced act, which focuses on minimising the risk of identification of the participant.

In the context of the SB project, the following organisational measures were considered for the specific case of the pilot site of Madeira [28], in Portugal, where 100 patients have been recruited and are providing data:

- The host institution appointed a DPO with advisory functions and monitoring compliance with the GDPR and with national provisions on data protection. The DPO also acts as a point of contact with both the supervisory authority and the data subjects. Through the DPO's assessment of the SB project, it became clear that some ethical requirements identified had to be safeguarded (e.g., beneficence: clinicians acting on behalf of a study participant may issue a GDPR request);

- In addition to the GDPR provisions, it is up to the respective pilot to verify whether additional ones have been established under the national law of the country where the research is carried out and take appropriate technical and organisational measures to ensure and to be able to demonstrate that the processing is performed in accordance;

- Given the nature of the SB project, the DPO considered it necessary to identify and list all envisaged data processing operations and the purpose(s) of processing, as well as the type of data subjects concerned and what measures were envisaged to address the risks inherent to such processing operations, including safeguards, security measures, and procedures to ensure the protection of personal data and to demonstrate compliance with the GDPR. Upon this recommendation, a data protection impact assessment (DPIA) was performed, where risks have been identified following the approach of the ENISA Manual [1], and the security measures to be taken for their mitigation have been proposed. These security measures are: the systematic description of the envisaged processing operations and the purpose of the processing; the assessment of the necessity and proportionality of the processing operations in relation to the purposes; the assessment of the risks to the rights and freedoms of data subjects; and the relevant measures for the mitigation of the risks identified (among which, access control, encryption techniques, pseudonymisation process, etc.).

- The provision of the information documents to the subject prior to obtaining their consent is essential for the participant to make an informed and reasoned decision, additionally ensuring the necessary clarifications to exercise the right to remove their consent at any time during the project. Due to the nature of the SB project and its activities with the elderly, that may or may not have the necessary digital literacy for a better understanding of common practices associated with digital health, subject access requests can be made by the person whose data is being processed or by an authorised third party.

- The informed consent, information sheet, data breach policy (including a data breach logbook), data subjects' rights policy (including a log of data subject requests), and the clinical protocol were presented to the local ethics committee. All documents were submitted in compliance with the formal requirements for submission

of an observational prospective study for the SB respective analysis.

- Upon completion of the recruitment process, after the successful creation of a study participant via the SB@Dashboard, an identifier visible to various end-user roles as well as to the participant him(her-)self that supports the various SB project activities, is automatically assigned to his/her personal record. In particular, this identifier is the only element on the basis of which communication (via phone) can take place for technical reasons. Notably, as demonstrated during the pilot in Madeira, knowing that their data have been pseudonymised study participants felt more comfortable in providing information, as they were not worrying that it could be used to identify them.

- All pre-recruitment, recruitment, and monitoring procedures (where data are sent to the SB platform) were carried out in accordance with the principles of confidentiality, availability, integrity, and privacy monitoring, including all components and intercomponent links in the SB platform (Dashboard, SB@App, HB and S4H systems). Since personal data is stored in a separate repository, this data segregation helps to better monitoring the personal data stored, not only reducing the risk of data breaches but, from an organisational standpoint, also reducing the complexity of data governance, as pseudonymised data is easier to manage and less likely to be subject to any regulatory or legal requirements.

- The way the data are stored (patients' personal data and PII stored encrypted, only pseudonymised big data used in analytics) allows the analysis of the anonymised data to continue even after the SB project's lifecycle, provided that all personal data stored encrypted will be deleted. Thus, after the completion of the SB project, data kept in the separated repository will no longer be needed to conduct the research (e.g., technical support, interventions) and, consequently, will be erased and not further used for any data process.

Maintaining (and safeguarding) the confidentiality of study participants' personal data while allowing, without applying particularly expensive protection controls, numerous analyses to be carried out, is of the utmost importance throughout the SB study. The project as a whole and, consequently the SB platform, ensures compliance with the legal framework of data protection and security, based on the fundamental principle of transparency and principles of fairness and legality. As presented, both on a technical and organisational level, the SB project has taken a proactive approach to privacy by ensuring that pseudonymisation is a core part of all its operations, to ensure that all operations comply with the GDPR and that the rights of EU citizens are respected.

V. Future Work

While adhering to the 'Privacy by Design' principle (right from the design stage of any product or service) is considered of fundamental importance, one must consider that data breaches are one of the most pressing and concerning issues facing healthcare organisations today, and even more alarmingly, evidence suggests that the specific sector is among the most vulnerable ones [33]. One reason for this is the fact that medical data are more sensitive than other types of data because this type of information can be highly sensitive and, if exposed, could lead to serious physical, psychological, or financial harm to the individual. Such amount of sensitive data combined with lax security measures, makes them an attractive target for hackers. As such, even the most well safeguarded systems cannot cope with any new type of vulnerability and new hacking method that may be discovered in the future. Regardless of how optimal the implemented techniques are or any new ways introduced continuously to protect these assets like a "defensive perimeter" around them, imposing in parallel an a posteriori methodology for conducting security and privacy assessments in a continuous manner is needed to combat continuously evolving security threats, human errors, technical weaknesses that can be exploited and other less known vulnerabilities derived from the increasing complexity and interconnectivity of information and communications technologies (ICTs). Within the context of SB project, this objective is materialized by adopting a continuous security and privacy assurance approach that involves the utilisation of usage metadata for continuous monitoring and dynamic testing of the platform to generate evidence that can support the continuous assessment of the security and privacy provisions of the platform, making it transparent, audible, and trustworthy.

With the knowledge derived from assessments results, as well as the audit methodology itself, will be used to improve existing privacy and security foundations of the RESPECT H2020 project, by exploring and identifying system-specific cyber-physical weaknesses posing security, privacy, and safety threats, in autonomous mobile robots operating in a healthcare environment.

VI. Conclusion

Pseudonymisation is an important tool for protecting the privacy of individuals in medical research. It helps to ensure that personal information is not directly identifiable, while still allowing the data to be used for research purposes and for providing technical and healthcare support where needed at the same time. The SMART BEAR approach is in line with the principles of the GDPR, which requires that personal data be processed in a way that ensures appropriate security and privacy controls are in place. SMART BEAR services and organisational processes are stacked in such a way as to minimise the risk of leakage, to ensure that the data collected and processed are used only for the purpose they were intended, and that the data subjects' privacy is fully respected. One lesson from this approach is that organisations that use pseudonymisation may need to update certain procedures, to ensure the effective and secure use of pseudonymised data, and to update policies related to data access and sharing, to ensure that data are not shared with unauthorised parties.

Acknowledgment

This work was supported by the European Commission's Horizon 2020 research and innovation program under the SMART BEAR project grant agreement No 857172, and the RESPECT project under grant agreement No 101007673.

References

[1] ENISA, (2021). Cloud Security for Healthcare Services, online (retrieved 2/3/2023): https://www.enisa.europa.eu/publications/cloudsecurity-for-healthcare-services

[2] Md Haris Uddin Sharif, & Mehmood Ali Mohammed. (2022). A literature review of financial losses statistics for Cyber Security and future trend. World Journal of Advanced Research and Reviews, 15(1), 138–156. https://doi.org/10.30574/wjarr.2022.15.1

[3] 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 (GDPR), Directive 95/46/EC, OJ L 119, 4.5.2016, p. 1.

[4] Data Protection Act 2018 (Section 36(2)) (Health Research) Regulations 2018 (S.I. No. 314/2018), online (retrieved 2/3/2023): https://www.irishstatutebook.ie/eli/2018/si/314/made/en/pdf

[5] German Federal Data Protection Act (Bundesdatenschutzgesetz – 'BDSG'), online (retrieved 2/3/2023): https://www.gesetze-iminternet.de/englisch_bdsg/

[6] Hellenic Law 4624/2019 enacts supplemental measures for the application of the General Data Protection Regulation ("GDPR"), online (retrieved 2/3/2023): https://www.ey.com/en_gr/tax/taxalerts/ey-law-alert-law-4624-2019-protection-of-personal-data-andmeasures-for-the-implementation-of-the-gdpr

[7] Italian Data Protection Decree Harmonizes National Law with GDPR Provisions, commentary article online (retrieved 2/3/2023): https://www.jonesday.com/en/insights/2018/10/italian-dataprotection-decree-harmonizes-national

[8] ENISA, (2022). Deploying Pseudonymisation Techniques, The case of the Health Sector, online (retrieved 2/3/2023): https://www.enisa.europa.eu/publications/deployingpseudonymisation-techniques

[9] Project "Smart Big Data Platform to Offer Evidence-based Personalised Support for Healthy and Independent Living at Home" SMART BEAR EU HORIZON 2020, H2020-SC1-FA-DTS-2018-2, online (retrieved 2/3/2023): https://www.smart-bear.eu/

[10] Peretokin, V. et al., (2022). Overview of the SMART-BEAR Technical Infrastructure. In Proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and eHealth - ICT4AWE, ISBN 978-989-758-566-1; pages 117-125. https://doi.org/10.5220/0011082700003188

[11] ENISA, (2018). Recommendations on shaping technology according to GDPR provisions - An overview on data pseudonymisation, online (retrieved 2/3/2023): https://www.enisa.europa.eu/publications/recommendations-onshaping-technology-according-to-gdpr-provision

[12] R. Tinabo, F. Mtenzi and B. O'Shea, (2009). Anonymisation vs. Pseudonymisation: Which one is most useful for both privacy protection and usefulness of e-healthcare data, International Conference for Internet Technology and Secured Transactions, (ICITST), London, UK, 2009, pp. 1-6, doi: https://doi.org/10.1109/ICITST.2009.5402501

[13] V. Bellandi et al., (2021). Engineering Continuous Monitoring of Intrinsic Capacity for Elderly People, 2021 IEEE International Conference on Digital Health (ICDH), Chicago, IL, USA, 2021, pp. 166-171, https://doi.org/10.1109/ICDH52753.2021.00030

[14] H2020 Programme Guidelines on FAIR Data Management in H2020, online (retrieved 2/3/2023): https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[15] Demiroglu, S., Skrowny, D., Quade, M. et al., (2012). Managing sensitive phenotypic data and biomaterial in large-scale collaborative psychiatric genetic research projects: practical considerations. Mol Psychiatry 17, 1180–1185 (2012), https://doi.org/10.1038/mp.2012.11

[16] Aamot, H., Kohl, C.D., Richter, D., & Knaup-Gregori, P., (2013). Pseudonymization of patient identifiers for translational research. BMC Medical Informatics and Decision Making, 13, 75 - 75. https://doi.org/10.1186/1472-6947-13-75

[17] Pommerening, K., et al., (2006). Pseudonymization Service and Data Custodians in Medical Research Networks and Biobanks. In: Hochberger, C. & Liskowsky, R. (Hrsg.), INFORMATIK 2006 – Informatik für Menschen, Band 1. Bonn: Gesellschaft für Informatik e.V.. (S. 715-721). https://dl.gi.de/handle/20.500.12116/23646

[18] Neubauer T, Heurix J., (2011). A methodology for the pseudonymization of medical data. Int J Med Inform. 2011 Mar;80(3):190-204. https://doi.org/10.1016/j.ijmedinf.2010.10.016

[19] Gulcher, J. R., Kristjansson, K., Gudbjartsson, H. & Stefansson, K., (2000). Protection of privacy by third-party encryption in genetic research in Iceland. Eur. J. Hum. Genet. 8, 739–742, https://doi.org/10.1038/sj.ejhg.5200530

[20] Spitzer, M., Ullrich, T., & Ueckert, F. (2009). Securing a web-based teleradiology platform according to German law and "best practices". Studies in health technology and informatics, 150, 730–734, https://doi.org/10.3233/978-1-60750-044-5-730

[21] Kohlmayer, F., Lautenschläger, R.R., & Prasser, F., (2019). Pseudonymization for research data collection: is the juice worth the squeeze? BMC Medical Informatics and Decision Making, 19. https://doi.org/10.1186/s12911-019-0905-x

[22] Wirth, F.N., Meurers, T., Johns, M., Prasser, F., (2021). Privacypreserving data sharing infrastructures for medical research: systematization and comparison. BMC Med Inform Decis Mak 21, 242 (2021). https://doi.org/10.1186/s12911-021-01602-x

[23] Prasser , F., (2021). Pseudonymisation in healthcare research and practice, IPEN webinar 2021 "Pseudonymous data: processing personal data while mitigating risks", online: https://edps.europa.eu/ipen-webinar-2021-pseudonymous-dataprocessing-personal-data-while-mitigating-risks_en

[24] ISO/TC 215 Health informatics. "ISO 25237:2017 Health informatics - Pseudonymization" 2017-01, https://www.iso.org/standard/63553.html

[25] Hussein R., et al., (2022). General Data Protection Regulation (GDPR) Toolkit for Digital Health, Studies in Health Technology and Informatics (2022), volume 290: MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation.

[26] Fatehi F., Hassandoust F., Ko R. K. L., Akhlaghpour S., (2020). General Data Protection Regulation (GDPR) in Healthcare: Hot Topics and Research Fronts, Stud Health Technol Inform (2020), 270: 1118-1122

[27] Yuan B., Jiannan Li Y., (2019). The Policy Effect of the General Data Protection Regulation (GDPR) on the Digital Public Health Sector in the European Union: An Empirical Investigation, Int J Environ Res Public Health (2019), 16(6): 1070

[28] Madeira Digital Health and Wellbeing, (2023). online (last visited on 09/03/2023): www.digit-madeira.pt

[29] HL7 FHIR Specification v4.0.1: R4: online (last visited on 26/03/2023): http://hl7.org/fhir/R4/

[30] SNOMED Clinical Terms: online (last visited on 26/03/2023): https://www.snomed.org/

[31] International Classification of Diseases,Ninth Revision, Clinical Modification (ICD-9-CM): online (last visited on 26/03/2023): https://apps.who.int/iris/handle/10665/40492

[32] Logical Observation Identifiers Names and Codes (LOINC): online (last visited on 26/03/2023): https://loinc.org/

[33] Seh, A. H., et al. (2020). Healthcare Data Breaches: Insights and Implications. Healthcare (Basel, Switzerland), 8(2), 133. https://doi.org/10.3390/healthcare8020133