Journal of Biomedical Informatics 156 (2024) 104670

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Original Research

Toward the European Health Data Space: The IMPaCT-Data secure infrastructure for EHR-based precision medicine research

Silvia Rodríguez-Mejías [a], Sara Degli-Esposti [b,*], Sara González-García [a], Carlos Luis Parra-Calderón [a]

[a] *Computational Health Informatics Group, Institute of Biomedicine of Seville, IBiS/Virgen del Rocio University Hospital/CSIC/University of Seville, Avenue Manuel Siurot S/N, Seville, 41013, Spain*
[b] *IFS-CSIC, Albasanz 26, Madrid, 28036, Spain*

ARTICLE INFO

Keywords:
European Health Data Space
IMPaCT-Data
GDPR
Health research data infrastructure
Spanish National Security Scheme
Information security

ABSTRACT

Background: Art. 50 of the proposal for a Regulation on the European Health Data Space (EHDS) states that "health data access bodies shall provide access to electronic health data only through a secure processing environment, with technical and organizational measures and security and interoperability requirements".
Objective: To identify specific security measures that nodes participating in health data spaces shall implement based on the results of the IMPaCT-Data project, whose goal is to facilitate the exchange of electronic health records (EHR) between public entities based in Spain and the secondary use of this information for precision medicine research in compliance with the General Data Protection Regulation (GDPR).
Data and methods: This article presents an analysis of 24 out of a list of 72 security measures identified in the Spanish National Security Scheme (ENS) and adopted by members of the federated data infrastructure developed during the IMPaCT-Data project.
Results: The IMPaCT-Data case helps clarify roles and responsibilities of entities willing to participate in the EHDS by reconciling technical system notions with the legal terminology. Most relevant security measures for Data Space Gatekeepers, Enablers and Prosumers are identified and explained.
Conclusion: The EHDS can only be viable as long as the fiduciary duty of care of public health authorities is preserved; this implies that the secondary use of personal data shall contribute to the public interest and/or to protect the vital interests of the data subjects. This condition can only be met if all nodes participating in a health data space adopt the appropriate organizational and technical security measures necessary to fulfill their role.

1. Introduction

This article presents a framework to reconcile roles and functions of entities responsible for the adequate processing of personal data in a health data space such as the one envisioned in the Proposal for a regulation of the European Parliament and of the Council on the European Health Data Space (EHDS). Recently it has been argued that facilitating medical data sharing is the best strategy to ensure fairness and inclusion in medical research [1]. In other words, to reduce the risk of algorithmic bias in bioinformatics research, it is necessary to increase the data visibility of historically disadvantaged populations. Thus, as already claimed in the case of patients affected by rare diseases [2], healthcare professionals need technical and organizational solutions to strike a balance between privacy risks and patients' health benefits.

In this study we present a set of security measures, which includes robust and consistent data access procedures that can help minimize privacy risks, while also fostering best research practices in healthcare. In order to truly promote fairness and inclusion in biomedical informatics research, we argue that the capacity of computational systems to find, access, interoperate, and reuse EHR (FAIR principles) shall go hand in hand with ideas of data stewardship such as the recognition and respect of the legitimate rights and interests of all relevant stakeholders [3]. These rights and interests should be articulated in a data value governance agreement designed to safeguard the trust relationship between data subjects (patients) and all entities involved in the health data space.

An important part of the *data value governance agreement* includes the organizational and technical measures taken to comply with the

[*] Corresponding author.
E-mail address: sara.degli.esposti@csic.es (S. Degli-Esposti).

https://doi.org/10.1016/j.jbi.2024.104670
Received 14 November 2023; Received in revised form 4 June 2024; Accepted 5 June 2024
Available online 14 June 2024

legal obligations of the General Data Protection Regulation (GDPR). The framework here presented draws insights from the design and implementation of IMPaCT-Data [4]: an health data space created in Spain to boost precision medicine research by promoting data sharing of clinical and molecular information between clinics, hospitals and research centers.

IMPaCT-Data is one of three programs (Grant no. IMP/00019) of the Infrastructure for Precision Medicine associated with Science and Technology (IMPaCT) [5] funded by the Instituto de Salud Carlos III and co-financed by the European Regional Development Fund (ERDF - “A way to make Europe”). During the three years program (January 2021–December 2023), the team has developed and validated an environment for the integration and analysis of electronic health records (EHR) including clinical, molecular and genetic data. In this article we present lesson learned by the team responsible for ensuring compliance of IMPaCT-Data with the Spanish National Security Scheme (Royal Act 311/2022), GDPR and the Spanish Personal Data Protection and digital rights law (Statutory Law 3/2018) and other relevant national and regional laws determining the appropriate use of EHR data.

Statement of significance

Summary	Description
Problem or issue	The European Health Data Space (EHDS) shall facilitate the secondary use of electronic health records for scientific research.
What is already known	GDPR demands the implementation of technical and organizational security measures to protect personal data.
What are the expected results	Security requirements of the Spanish National Security Scheme are applied to the IMPaCT-Data Precision Medicine Infrastructure to create a minimum model for the EHDS useful for others entering this space.

2. The European Health Data Space

The proposal for a regulation on the European Health Data Space — EHDS (COM(2022) 197 final) presented in March 2022, distinguishes between primary and secondary use of EHD. ‘Primary use of electronic health data’ means the processing of personal electronic health data for the provision of healthcare services to assess, maintain or restore the state of health of the natural person to whom that data relates. ‘Secondary use of electronic health data’ means the processing of electronic health data for reasons of public interest in the area of public and occupational health, such as protection against serious cross-border threats to health, to produce national, multi-national and Union level official statistics, education or teaching activities, scientific research related to health or care sectors, development of products or services contributing to public health or social security, training, testing and evaluating of algorithms, including in medical devices, AI systems and digital health applications, providing personalized healthcare consisting in assessing, maintaining or restoring the state of health of natural persons, based on the health data of other natural persons. Here we focus on chapter IV: Secondary use of electronic health data where specific types of data are identified as well as generic roles of EHDS entities. These are: data holders (art. 33); health data access bodies, public sector bodies, supervisory authorities, stakeholders, other national competent bodies (art. 36). Incidental findings must be communicated by the health data access body to natural persons (EHDS art. 38.3), but data subjects cannot obtain information about where their data came from (derogation of GDPR art. 14). The Body shall publish an annual activity report with information about applications for EHD access, data altruism permits, fulfillment of regulatory and contractual commitments by data users and data holders, audits, handling of requests from natural persons on the exercise of their data protection rights, revenues, number of scientific publications and of digital health products and services, including AI applications, developed using data accessed via EHDS. To summarize, in the EHDS we find three main actors: data holders, data users and health data access bodies. In the rest of these section we present the way these roles can be associated with roles identified in the GDPR and in the Spanish National Security Scheme to come up with a synthesis that can help identify type of users and associated data access privileges in a federated big data infrastructures like the one developed in IMPaCT-Data.

3. Legal liabilities under GDPR in the EHDS

The purpose of the GDPR is to provide a consistent framework for the protection of Europeans’ fundamental right to personal data protection [6]. It identifies Data Controllers as those who determine the purpose and essential means of processing. As Data Processors, who process data on behalf of Controllers, all these entities are responsible for respecting data subjects’ rights and for protecting their personal data from unauthorized access and use. Personal data must be processed adequately only for legitimate purposes, Further processing is only legitimate if the purpose is consistent with the original one. Otherwise the purpose limitation principle restricts further processing. The problem is that the interpretation of what constitute legitimate further processing is left open to interpretation, leading to considerable interpretative ambiguities as to whether “secondary use” of data by researchers constitutes “further processing” under the GDPR, which exposes researchers to the risk of not compliance [7]. Furthermore, it is not clear if data providers who let data users access remotely data in a secure computing environment shall be considered processors or joint controllers [8]. These kinds of terminological disputes in the domain of data protection law have important consequences for the design and implementation of a data space as they determine which entity is liable if something goes wrong. The proliferation of cybersecurity threats and the adoption of emerging AI tools in bioinformatics increase dramatically privacy risks and the risk of suffering a data breach and being sanctioned by a data protection authority [9]. From this standpoint, the data governance framework and security/data protection by design methodology adopted within IMPaCT-Data can be very useful and highly informative for anyone willing to join the EHDS.

4. The spanish national security scheme

In Spain, the National Security Scheme (NSS) demands the adoption of measures to ensure the confidentiality, integrity, traceability, authenticity, availability and preservation of data held by the public administration. Originally adopted in 2010, NSS was updated in may 2022 (Royal Act 311/2022) to align with ISO 27001 standard and extend its requirements to any public sector entity [10]. The updated NSS includes in its Annex II a total of 72 specific security measures that public entities or private entities in a contractual relationship with public entities must adopt.

To understand how public hospitals and healthcare providers would adapt to the updated NSS, IT professionals in charge of the infrastructure of nodes participating in the IMPaCT-Data project were asked to assess the viability of applying each of these security measures. With the use of a questionnaire (included in the appendix) testing the level of adoption with NSS a group of highly relevant security measures to be implemented in an health data space was identified. Based on this feasibility analysis of security measures actually adopted within the IMPaCT-Data federated personalized medicine infrastructure, in this article we try to create a common terminology that would help associate security measures with GDPR obligations within the context of the proposed EHDS.

5. Toward a framework to identify key roles in an health data space in line with GDPR

In may 2023 the Spanish Data Protection Authority published a guide titled "An approximation to Data Spaces from the standpoint of the GDPR" to help Data Protection Officers (DPO) comply with the data protection laws in the context of data spaces. Starting from the consideration that data spaces are populated with both non personal and personal data, the guide focuses on privacy and security measures to be adopted by data controllers and processors. Here we take a step forward and try to identify key roles. The EHDS identifies as overarching Data Controller the Health Data Access Body. Below this Body we find Data Holders and Data Users. The problem with this formulation is that in the reality of biomedical research it is not straightforward to distinguish Data Holders from Data Users. The following hybrid roles help us better describe the reality of HDS.

1. *Data Space Prosumers*: natural or legal persons who wish to carry out a processing operation involving personal data within the framework of the Data Space and/or that provide data to the space (prosumer stands for consumer/provider).
2. *Data Space Enablers*: those persons or entities that give support to all intervening parties and help implement the organizational and technical measures needed to comply with the law and to avoid data mismanagement or loss.
3. *Data Space Gatekeepers*: Health Data Access Bodies responsible for granting access to EHD for secondary use. Gatekeepers facilitate the smooth interaction between Data Prosumers and between Data Prosumers and Data Subjects. They are Joint Controllers (GDPR art. 26) with prosumers who are Data Holders. They are responsible for implementing all the organizational and technical measures needed to comply with established legal obligations (please note that this definition is not related to what qualifies as gatekeeper under the Digital Markets Act).

Now let us look into GDPR terminology (art. 4) which identifies again three main types of roles:

1. 'Controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.
2. 'Processor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.
3. 'Data subject' means an identified or identifiable natural person.

A couple of examples can help us make up our minds. Clinical researchers are prosumers when they conduct a clinical trial and collect and analyze the data. The DPO of the clinics where they work is the Data Controller. The biomedical informatics service appointed by the researcher to test the efficacy of a new AI techniques on the data is a Data Space Enabler and in GDPR terms a Data Processor.

Finally, if we move on to the analysis of the big data infrastructure we can identify three types of nodes:

1. Nodes that produces data.
2. Nodes that consumes data.
3. Nodes that produces and consumes data.

These definitions reflect the meaning of the term Prosumer we suggest to use in the context of the EHDS. Overall the big data infrastructure would be managed and maintained by Data Space Enablers and would be owned and administered by Data Space Gatekeepers. In the rest of the article we will see how this terminology matches the one adopted in IMPaCT-Data.

5.1. A spanish approximation to the European Health Data Space initiative

IMPaCT's mission is to establish and transmit the necessary knowledge to support and facilitate the effective deployment of Personalized Precision Medicine. This will represent a paradigm shift and a new reality for the healthcare system, enabling a more effective and safer personalized preventive, diagnostic and therapeutic approach for each patient. IMPaCT's Strategic Plan is structured around three strategic axes and two transversal lines. The strategic axes correspond to each of the three programs: Cohort, Data and Genomics, and are developed in specific actions and work packages, with compliance indicators to verify the effectiveness of the deployment of the IMPaCT infrastructure. In addition, the two strategic lines transversal to the three programs provide internal coherence in aspects such as data ethics and internationalization of the platform, which are common to the three strategic axes.

As members of IMPaCT-Data are based in different autonomous communities in Spain, they must comply with regional laws and national laws such as the Spanish National Security Scheme or NSS (Royal Decree 311/2022), the National Security Law (36/2015), the Law for the Protection of Personal Data and Guarantee of Digital Rights (3/2018), the Law on Trusted Electronic Services (6/2020), among others and current standards, such as ISO 27001 [11] to ensure information security.

Here we focus specifically on NSS, its risk-based approach and list of security measures to be adopted by Spanish public authorities and other organizations that handle sensitive and classified information. The supervisory authority for the implementation of NSS in the public sector is the National Cryptologic Center (CCN), which is responsible for offering guidance and verify compliance. CCN-CERT yearly publishes the (CCN STIC Guides) to help organizations implement the recommendations included in the NSS and to offer advice on data life cycle security measures. The severity of the security measures (low, medium, high) to be adopted depends on the degree of sensitivity of the information handled and on the potential negative impact that a potential breach of data confidentiality, integrity or availability could produce. In the case of EHR, which are special categories of data on a massive scale, we shall assume the highest security level. The list of security measures listed in the NSS are included in Annex A. These measures ensure data security and privacy by using measures such as encrypted data storage and strict access control and authentication mechanisms that prevent unauthorized access.

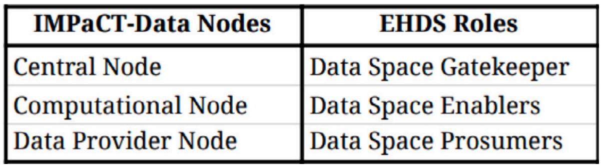
As the goal of IMPaCT-Data is to provide a secure, privacy-preserving environment for health research performed by public entities in Spain, it also important to understand why Spain represents an interesting case in the path toward the EHDS. Spanish autonomous communities enjoy different degrees of fiscal and administrative autonomy reflecting social and cultural differences [12]. Thus, Spain is an interesting case to understand how governance mechanisms anchored to public sector entities can ensure collaboration while overcoming boundaries and constraints stemming from different regulatory frameworks and jurisdictions [13].

5.2. The IMPaCT-Data infrastructure

In the IMPaCT-Data infrastructure, three types of nodes can be found:

1. *Central node* (IMPaCT-Central): includes the general infrastructure coordination services and constitutes the access point to the system. It is responsible for offering the transversal services that allow the system to be operated in an integrated manner.
2. *Computational node*: it is in charge of providing computing resources to the infrastructure by executing analysis processes according to the requests received from the central node analysis environment.

Table 1
Correspondence between IMPaCT-Data terminology and suggested EHDS roles.



IMPaCT-Data Nodes	EHDS Roles
Central Node	Data Space Gatekeeper
Computational Node	Data Space Enablers
Data Provider Node	Data Space Prosumers

3. *Data provider node*: corresponds to the provider of protected data services, it contains functional modules that ensure its interconnection with the infrastructure. These nodes can also provide computing services.

Members of IMPaCT-Data represent nodes that implement one or more of the following types of services:

1. *Core services*: general services of the IMPaCT-Data Cloud, which include: Identification and Authorization Server, Platform Access Portal, Data Catalog (access and management), Public Shared Data Server, Data Access Management Portal, Log and Tools Server, and Distributed Computing Manager. As it is a core network architecture, there is a single Central Node, although it can also assume other roles if it provides services to the central node.
2. *Computing services*: computing resources that are integrated into a distributed computing model.
3. *Sensitive data provision*: data repositories of special categories of personal data that require the implementation of additional access requirements, security measures and privacy-preserving tools.

The treatment of special categories of data is particularly relevant in the context of the EHDS. As written in GDPR recital 53: “Special categories of personal data which merit higher protection should be processed for health-related purposes only where necessary to achieve those purposes for the benefit of natural persons and society as a whole”. According to GDPR art. 6(e), secondary use of EHR for research purposes can be considered lawful when processing is necessary for the performance of a task carried out in the public interest. Of course, as already demonstrated in the case of biobanks [14], technical and organizational measures need to be implemented to ensure Data Controllers process personal data lawfully, adequately and in a transparent manner in relation to the data subject. Among these “technical and organizational measures”, security measure are of utmost importance to guarantee the respect of the principles of security and data protection by design and by default. If we now compare the terminology used in IMPaCT-Data with the one suggested for the EHDS, as shown in Table 2, we will be able to identify in the next section specific the most important security requirements the each type of node – or role – needs to implement (see Table 1).

6. Data and methods

The Social Security Health System (SGSSS) is a national insurance system for receiving healthcare services based on the principles of universality, free access, equity and fairness. Mostly funded by the central government with taxpayers’ money, SGSSS reflects the administrative and regional structure of the country. Healthcare competences are transferred to the 17 autonomous communities, with the national government and the Inter-territorial Council of the SGSSS responsible of strategic direction and for the overall coordination of the healthcare system, and supervision of performance of the overall healthcare system. Here we focus specifically on the application of national security scheme or NSS in the contest of public healthcare institutions willing to exchange EHR for research purposes.

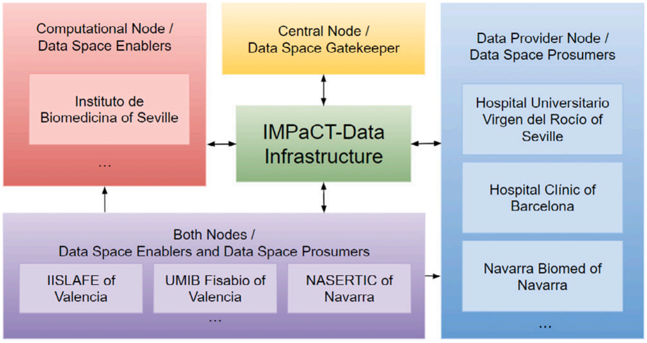

Computational Node /
Data Space Enablers
Instituto de Biomedicina of Seville
...
Central Node /
Data Space Gatekeeper
IMPaCT-Data Infrastructure
Data Provider Node /
Data Space Prosumers
Hospital Universitario Virgen del Rocío of Seville
Hospital Clínic of Barcelona
Navarra Biomed of Navarra
...
Both Nodes /
Data Space Enablers and Data Space Prosumers
IISLAFE of Valencia
UMIB Fisabio of Valencia
NASERTIC of Navarra
...

Fig. 1. The IMPaCT-Data infrastructure nodes.

The NSS includes in total 72 security measures. Of these 72 security measures, to draft the questionnaire to be sent to IMPaCT-Data members a total of 24 measures were selected by experts of the Computational Health Informatics Research Group of the Virgen del Rocio University Hospital and the Institute of Biomedicine of Seville. This reduced set of measures included minimum requirements associated with the specificity of the project nodes participating in IMPaCT-Data, which are sensitive data provider node, computation provider node, and both provider node. The reduction in the number of items also helped increase response rate and reduce. As we write the IMPaCT-Data infrastructure is still under development, but the 24 measures here analyzed are taken into account to comply with the “security by design” principle present in the GDPR.

Specifically, for sensitive data provider node and computer provider node, a total of 15 security measures were surveyed, of which six corresponds to the measures belonging to the NSS group of access control measures and nine additional security measures specific to each node that were considered the most relevant ones according to the type of node. A third questionnaire, designed for those participants who met the profile of both nodes, consisted of a total of 24 security measures: 6 associated with access control (also included in the other questionnaires); nine measures that were considered most appropriate according to the profile of the sensitive data provider node; other nine measures selected for the computing provider node.

The survey was sent to all 48 participating nodes of the IMPaCT-Data project at the beginning of May 2023. Despite all IMPaCT-Data nodes received the survey on security measures and several reminders, response rate was only 14 percent. The reason why only 7 nodes answered the entire questionnaire was high workload and lack of personnel. Nonetheless, complete questionnaires came from a sample of sufficiently heterogeneous nodes. Of the seven questionnaires obtained, three came from the sensitive data provider nodes, one from the computation provider node and three from nodes performing both computation and data provision. The 7 nodes/organizations are: (i, ii) Navarra Biomed and NASERTIC based in Navarra, (iii, iv) UMIB Fisabio and IISLAFE based in Valencia, (v) Hospital Clínic de Barcelona based in Catalonia, and (vi, vii) Hospital Universitario Virgen del Rocío and Instituto de Biomedicina de Sevilla, both based in Andalusia (see Fig. 1). During the analysis of the data, answers belonging to computation and data provision nodes were compared with those of the sensitive data provider node.

7. Results

Here we present the results obtained from the experience of the IMPaCT-Data federated personalized medicine infrastructure based on the responses gathered after interviewing IMPaCT-Data participating nodes about the most viable and effective security measures defined in the NSS taking into account the mission, organizational specificity and technical infrastructure of each node.

Table 2
NSS security measures corresponding to each type of IMPaCT-Data node.

IMPaCT-Data/suggested EHDS roles	NSS security measures
– Central node – Data Space Gatekeeper	**Access control measures:** - Access requirements - Segregation of duties and functions - Access rights management - Authentication process (for external and internal users).
– Computational node – Data Space Enablers	**Planning measures:** - Capacity sizing and management. **Data protection:** - Perimeter securitization - Data integrity and authenticity - Information flow segregation.
– Data provider node – Data Space Prosumers	**Organizational measures:** - Security policy - Security rules - Security procedures. **Planning measures:** - Risk analysis - Security architecture **System monitoring measures:** - Intrusion detection procedure - Surveillance. **Data protection:** - Custody. - Protection of personal data.

Table 2 shows a list of security measures derived from the analysis of survey responses. These security measures are those demanded by the NSS and considered viable and, thus, adopted by entities participating in IMPaCT-Data. For the sake of clarity, the first column on the left includes the terminology used in IMPaCT-Data with the associated hybrid roles we have previously identified in line with the EHDS framework. To ensure data protection by design and security by default of EHR, in the right-hand side column specific security measures are identified for type of node participating in an health data space. These measures focus on the identity management scheme and corresponding authentication and authorization procedures as well as on data accuracy and minimization, data integrity and confidentiality.

8. Discussion

At the beginning of the article we have argued that, in the context of secondary processing of special categories of data for scientific research purposes, implementing FAIR principles is a necessary but insufficient condition [15] for ensuring fairness and inclusion, even when methods to make FAIR by design during the data collection phase are adopted [16]. GDPR principles such as data integrity and quality or data portability can complement FAIR principles and help improve biomedical informatics research. Unfortunately, it is still controversial how to fully apply GDPR in practice, and navigate its legal complexities and loopholes, or how embracing security measures demanded by laws such as the Spanish Security Scheme (NSS) can enable data exchange and collaboration.

To move from a vision of GDPR in the books to its practical application, legal debates, such as those about the scope of permissible transfers of personal data and the appropriate legal and technical mechanisms to safeguard such transfers [17], can only be resolved by clarifying the terms of the data governance agreement between members of a health data space. As an important part of that agreement is about assigning responsibilities for key security functions (such as authentication, authorization, or auditing), the IMPaCT-Data case study has offered an opportunity to reflect on the usefulness of having security specific legislation, such as the NSS, to support organizations in their effort to apply data protection and security by design and by default principles.

GDPR besides establishing principles such as lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, integrity and confidentiality, demands to implement technical and organizational procedures to protect personal data, while also promoting the free flow of data. Various certification schemes exist to help data controllers and processors identify the most appropriate security policies, technologies and procedures. ISO/IEC 27001 provides a model for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an Information Security Management System (ISMS).

More specifically, the U.S. NIST SP 800-53 provides a catalog of security and privacy controls adopted by U.S. federal agencies and also by private companies. These controls are organized within categories such as access management, audit and accountability, and risk management. This framework is aligned with others, such as ISO/IEC 27001 and the NIST Cybersecurity Framework.

Similarly, the National Security Scheme (NSS) in Spain aims to guarantee the security of information handled by public entities and their suppliers. Its scope includes both public administrations and ICT service providers that handle public information. It indicates specific set of necessary measures already present in standards such as ISO/IEC 27001 according to the level of risk (low, medium, high) assigned in each case.

Similar to NIST SP 800-53, which applies to U.S. federal agencies, the National Security Scheme (NSS) applies to Spanish public administrations and their suppliers. ISO/IEC 27001 has a global scope and is applicable to any organization seeking certification. GDPR is more focused on personal data protection, while NSS, ISO/IEC 27001 and NIST SP 800-53 address information security more broadly. NSS and NIST SP 800-53 are aligned with ISO/IEC 27001 in terms of best practices and security controls. ISO/IEC 27001 allows formal certification of the ISMS, while NSS may require internal and external audits to ensure compliance. Despite their differences, all these regulations and legislation share common principles and practical recommendations and measures that are relevant and can be applied to the future European Health Data Space to ensure data security and privacy.

8.1. Limitations

This study is a preliminary attempt to identify which node shall be mostly responsible for taking each security and data protection measures within a future EHDS. Of course this study is not conclusive and has some limitations. First of all, the limited response rate that reduce the generalizability of our results, which are based on the experience of a limited number of organizations. Furthermore, NSS has a total of 72 security measures, of which only a subsample of 24 measures was evaluated. However, being this study the first one establishing a clear connection between the usefulness and viability of these security measures and the role different nodes play in the context of biomedical research, the results here presented can help members of the EHDS assign tasks and responsibilities to protect personal data, while also promoting collaboration and data exchange in biomedical research.

In a future work, all 72 NSS security measures could be surveyed and, in case of full compliance by the surveyed institutions, proceed to the complete certification of the NSS. Another limitation is that we only focused on the exchange of EHR between public entities and authorities and public research centers. The exploitation of HER for commercial purposes introduces further complexities from the perspective of data protection law that cannot be addressed here.

9. Conclusions

The proposed EU regulation on the European Health Data Space (EHDS) promises to revolutionize access to health data, while fostering transparency, accountability and security. As the relationship between Spanish autonomous communities and the Spanish government resembles somehow the relationship EU member States have with the European Union, Spain is an interesting case study for the EHDS because Spanish Autonomous Communities are responsible for drafting and implementing healthcare laws and policies.

The examination of the Spanish IMPaCT-Data infrastructure helps highlight the fundamental role that security measures play in the future EHDS. In this article in Table 1, column one, based on our initial conceptualization, we associated the terminology used in IMPaCT-Data with the roles identified in the EHDS. Then, we identify a list of the most important security measures that each type of node has decided to implement to achieve its mission, while complying with the NSS, GDPR and associated laws and guidelines. This summary may help members of an heath data space identify duplication of tasks, identify synergies and prioritize investments and efforts.

The analysis of the security measures adopted by IMPaCT-Data nodes and a better understanding of the roles these nodes can play within the EHDS offer practical guidance to biomedical informatics professionals, data protection officers, biomedical researchers and other stakeholders willing to join the EHDS. In researching the technical and governance aspects of data spaces, there is no single approach that can be applied to their configuration [18]. The empirical analysis adds additional value by showing the viability of adopting specific security measures associated to specific roles [19].

The framework here presented, associating EHDS roles with GDPR terminology and node functionalities, can help clarify responsibilities and so help safeguard the trust relationship established between patients, who consent to the processing of their personal data for diagnostic and treatment purposes, and public healthcare authorities. Instruments such as the Data Protection Impact Assessment (DPIA) described in the GDPR shall be used by Data Space Gatekeepers to define, and periodically revise, the data value governance agreement.

CRediT authorship contribution statement

Silvia Rodríguez-Mejías: Data curation, Investigation, Resources, Writing – original draft, Visualization. **Sara Degli-Esposti:** Conceptualization, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Sara González-García:** Data curation, Investigation, Methodology. **Carlos Luis Parra-Calderón:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Carlos Parra Calderon reports financial support was provided by Carlos III Health Institute.

The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the IMPaCT-Data Project (Exp. IMP/00019) and partially funded by the project “HALO: A zero-trust data governance model for altruistic health data sharing” (202410E127).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jbi.2024.104670.

References

[1] K.P. Seastedt, P. Schwab, Z. O’Brien, E. Wakida, K. Herrera, P.G.F. Marcelo, L. Agha-Mir-Salim, X.B. Frigola, E.B. Ndulue, A. Marcelo, et al., Global healthcare fairness: We should be sharing more, not less, data, PLOS Digit. Health 1 (10) (2022) e0000102.

[2] M.G. Hansson, H. Lochmüller, O. Riess, F. Schaefer, M. Orth, Y. Rubinstein, C. Molster, H. Dawkins, D. Taruscio, M. Posada, et al., The risk of re-identification versus the need to identify individuals in rare disease research, Eur. J. Hum. Genet. 24 (11) (2016) 1553–1558.

[3] C. Wendelborn, M. Anger, C. Schickhardt, What is data stewardship? Towards a comprehensive understanding, J. Biomed. Inform. 140 (2023) 104337.

[4] ISCIII, Infraestructura de medicina de precisión asociada a la ciencia y la tecnología - IMPaCT, 2023, URL https://www.isciii.es/QueHacemos/Financiacion/IMPaCT/Paginas/default.aspx.

[5] IMPaCT-Data, Impact-data. Infraestructura de medicina de precisión asociada a la ciencia y la tecnología, 2023, URL https://impact-data.bsc.es/about/impact-data/.

[6] AEPD, Aproximación a los espacios de datos desde la perspectiva del RGPD, 2023, URL https://www.aepd.es/es/documento/aproximacion-espacios-datos-rgpd.pdf.

[7] R. Becker, D. Chokoshvili, G. Comandé, E.S. Dove, A. Hall, C. Mitchell, F. Molnár-Gábor, P. Nicolàs, S. Tervo, A. Thorogood, Secondary use of personal health data: When is it “further processing” under the GDPR, and what are the implications for data controllers? Eur. J. Health Law 1 (aop) (2022) 1–29.

[8] R. Becker, A. Thorogood, J. Bovenberg, C. Mitchell, A. Hall, Applying GDPR roles and responsibilities to scientific data sharing, Int. Data Privacy Law 12 (2022) 207–219.

[9] S. Degli-Esposti, E.M. Ferrándiz, After the GDPR: Cybersecurity is the elephant in the artificial intelligence room, Eur. Bus. Law Rev. 32 (1) (2021).

[10] G. España, Real decreto 311/2022 de 3 de mayo por el que SE regula el esquema nacional de seguridad, Bol. Oficial Estado 106 (2020) 61715–61804.

[11] ISO, ISO 27001 - Certificado ISO 27001 punto por punto - Presupuesto online. ISO 27001, 2022, URL https://normaiso27001.es/.

[12] H.F. Do Vale, Cuatro décadas de distribución del poder territorial en España, REIS: Rev. Esp. Investig. Sociol. (173) (2021) 3–26.

[13] C. Del-Real, A.M. Díaz-Fernández, Understanding the plural landscape of cybersecurity governance in Spain: a matter of capital exchange, Int. Cybersecur. Law Rev. 3 (2) (2022) 313–343.

[14] C. Staunton, S. Slokenberga, D. Mascalzoni, The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks, Eur. J. Hum. Genet. 27 (8) (2019) 1159–1167.

[15] M. Boeckhout, G.A. Zielhuis, A.L. Bredenoord, The FAIR guiding principles for data stewardship: fair enough? Eur. J. Hum. Genet. 26 (7) (2018) 931–936.

[16] M.G. Kersloot, A. Jacobsen, K.H. Groenen, B. dos Santos Vieira, R. Kaliyaperumal, A. Abu-Hanna, R. Cornet, P. AC‘t Hoen, M. Roos, L.S. Kool, et al., De-novo FAIRification via an electronic data capture system by automated transformation of filled electronic case report forms into machine-readable data, J. Biomed. Inform. 122 (2021) 103897.

[17] A. Bernier, F. Molnár-Gábor, B.M. Knoppers, P. Borry, P.M. Cesar, T. Devriendt, M. Goisauf, M. Murtagh, P.N. Jiménez, M. Recuero, et al., Reconciling the biomedical data commons and the GDPR: three lessons from the EUCAN ELSI collaboratory, Eur. J. Hum. Genet. (2023) 1–8.

[18] J.R.C.E. Commission, European data spaces. Scientific insights into data sharing and utilisation at scale, 2023, URL https://op.europa.eu/en/publication-detail/-/publication/dcac6aee-0e7a-11ee-b12e-01aa75ed71a1/language-en.

[19] E. Bernal-Delgado, S. García-Armesto, J. Oliva, F.I. Sánchez Martínez, J.R. Repullo, L.M. Peña-Longobardo, M. Ridao-López, C. Hernández-Quevedo, W.H. Organization, et al., Spain: Health system review, 2018.