# Assessing the Evolution and Influence of Medical Open Databases on Biomedical Research and Healthcare Innovation: A 25-Year Perspective with a Focus on Privacy and Privacy-Enhancing Technologies

Albert Yang, Mei-Lien Pan, Henry Horng-Shing Lu, Chung-Yueh Lien, Da-Wei Wang, Chih-Hsiung Chen, Der-Cherng Tarng, Dau-Ming Niu, Shih-Hwa Chiou, Chun-Ying Wu, Ying - Chou Sun, Shih-Ann Chen, Shuu-Jiun Wang, Wayne Huey-Herng Sheu, Chi-Hung Lin

# *Table of Contents*

# Assessing the Evolution and Influence of Medical Open Databases on Biomedical Research and Healthcare Innovation: A 25-Year Perspective with a Focus on Privacy and Privacy-Enhancing Technologies

Albert Yang[1] MD, PhD; Mei-Lien Pan[2]; Henry Horng-Shing Lu[3]; Chung-Yueh Lien[4]; Da-Wei Wang[5]; Chih-Hsiung Chen[6]; Der-Cherng Tarng[7]; Dau-Ming Niu[8]; Shih-Hwa Chiou[9]; Chun-Ying Wu[10]; Ying - Chou Sun[11]; Shih-Ann Chen[12]; Shuu-Jiun Wang[13]; Wayne Huey-Herng Sheu[14]; Chi-Hung Lin[15]

[1]National Yang Ming Chiao Tung University Taipei City, TW
[2]Institute of Hospital and Health Care Administration, National Yang Ming Chiao Tung University, Taipei TW
[3]Institute of Statistics, National Yang Ming Chiao Tung University Taipei TW
[4]Dept. of Information Management, of National Taipei University of Nursing and Health Sciences Taipei TW
[5]Institute of Information Science, Academia Sinica Taipei TW
[6]Institute of Technology Law, National Yang Ming Chiao Tung University Taipei TW
[7]Department of Medicine, Taipei Veterans General Hospital Taipei TW
[8]Department of Pediatric, Taipei Veterans General Hospital Taipei TW
[9]Department of Medical Research, Taipei Veterans General Hospital Taipei TW
[10]Institute of Biomedical Informatics, National Yang Ming Chiao Tung University Taipei TW
[11]Department of Radiology, Taipei Veterans General Hospital Taipei TW
[12]Department of Internal Medicine, Taichung Veterans General Hospital Taichung TW
[13]Department of Neurology, Neurological Institute, Taipei Veterans General Hospital Taipei TW
[14]Institute of Molecular and Genomic Medicine, National Health Research Institutes Taipei TW
[15]Department of Biological Science & Technology, National Chiao Tung University Taipei TW

**Corresponding Author:**
Albert Yang MD, PhD
National Yang Ming Chiao Tung University
No. 155 Sec. 2 Linong St., Beitou Dist.
Taipei City,
TW

## *Abstract*

The integration of medical open databases with artificial intelligence (AI) technologies marks a transformative era in biomedical research and healthcare innovation. Over the past 25 years, initiatives like PhysioNet have revolutionized data access, fostering unprecedented levels of collaboration and accelerating medical discoveries. This rise of medical open databases presents challenges, particularly in harmonizing research enablement with patient confidentiality. In response, privacy laws such as Health Insurance Portability and Accountability Act have been established, and privacy-enhancing technologies (PETs) have been adopted to maintain this delicate balance. PETs, including differential privacy, secure multiparty computation, and notably, federated learning, have become instrumental in safeguarding personal health information. Federated learning, in particular, represents a significant advancement by enabling the development and training of AI models on decentralized data. In Taiwan, significant strides have been made in aligning with these global data-sharing and privacy standards. We have actively promoted the sharing of medical data through the development of dynamic consent systems. These systems enable individuals to control and adjust their data-sharing preferences, ensuring transparency and continuity of consent in the ever-evolving landscape of digital health. Despite the challenges associated with privacy protections, the benefits, including improved diagnostics and treatment, are substantial. The availability of open databases has notably accelerated AI research, leading to significant advancements in medical diagnostics and treatments. As the landscape of healthcare research continues to evolve with open science and federated learning, the role of medical open databases remains crucial in shaping the future of medicine, promising enhanced patient outcomes and fostering a global research community committed to ethical integrity and privacy.

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
   Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
   Only make the preprint title and abstract visible.
   No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
   Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
   Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in  <a href="http

# Original Manuscript

**Title:** Assessing the Evolution and Influence of Medical Open Databases on Biomedical Research and Healthcare Innovation: A 25-Year Perspective with a Focus on Privacy and Privacy-Enhancing Technologies

**Running Title**: Medical Open Databases and Privacy over 25 Years

**Authors and Affiliations:**

Albert C. Yang[1,2,3*], Mei-Lien Pan[4], Henry Horng-Shing Lu[5], Chung-Yueh Lien[6], Da-Wei Wang[7], Chih-Hsiung Chen[8], Der-Cherng Tarng[9,10,11], Dau-Ming Niu[10,12], Shih-Hwa Chiou[3,13], Chun-Ying Wu[14,15], Ying-Chou Sun[16,17], Shih-Ann Chen[18], Shuu-Jiun Wang[16,19,20], Wayne Huey-Herng Sheu[21], Chi-Hung Lin[22]

[1] Digital Medicine and Smart Healthcare Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

[2] Institute of Brain Science, National Yang Ming Chiao Tung University, Taipei, Taiwan

[3] Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan

[4] Institute of Hospital and Health Care Administration, National Yang Ming Chiao Tung University, Taiwan

[5] Institute of Statistics, National Yang Ming Chiao Tung University, Taipei, Taiwan

[6] Department of Information Management, of National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

[7] Institute of Information Science, Academia Sinica, Taipei, Taiwan

[8] Institute of Technology Law, National Yang Ming Chiao Tung University, Taipei, Taiwan

[9] Department of Medicine, Taipei Veterans General Hospital, Taipei, Taiwan

[10] Institute of Clinical Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[11] Department and Institute of Physiology, National Yang Ming Chiao Tung University, Taipei, Taiwan

[12] Department of Pediatric, Taipei Veterans General Hospital, Taipei, Taiwan

[13] Institute of Pharmacology, College of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[14] Institute of Biomedical Informatics, National Yang Ming Chiao Tung University, Taipei, Taiwan

[15] Health Innovation Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

[16] School of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan

[17] Department of Radiology, Taipei Veterans General Hospital, Taipei, Taiwan

[18] Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan

[19] Brain Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

[20] Department of Neurology, Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan

[21] Institute of Molecular and Genomic Medicine, National Health Research Institutes, Taiwan

[22] Department of Biological Science & Technology, National Chiao Tung University, Taiwan

Institute of Microbiology & Immunology, National Chiao Tung University, Taiwan

**\*Corresponding Author:** Dr. Albert C. Yang, M.D., Ph.D.

Digital Medicine and Smart Healthcare Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

No. 155 Sec. 2 Linong St., Beitou Dist., Taipei City, Taiwan

Email: accyang@nycu.edu.tw

**Manuscript Information:** 1 table, 2 figures.

**Word Counts:** 260 words in abstract; 3,883 words in main text

**Abstract**

The integration of medical open databases with artificial intelligence (AI) technologies marks a transformative era in biomedical research and healthcare innovation. Over the past 25 years, initiatives like PhysioNet have revolutionized data access, fostering unprecedented levels of collaboration and accelerating medical discoveries. This rise of medical open databases presents challenges, particularly in harmonizing research enablement with patient confidentiality. In response, privacy laws such as Health Insurance Portability and Accountability Act have been established, and privacy-enhancing technologies (PETs) have been adopted to maintain this delicate balance. PETs, including differential privacy, secure multiparty computation, and notably, federated learning, have become instrumental in safeguarding personal health information. Federated learning, in particular, represents a significant advancement by enabling the development and training of AI models on decentralized data. In Taiwan, significant strides have been made in aligning with these global data-sharing and privacy standards. We have actively promoted the sharing of medical data through the development of dynamic consent systems. These systems enable individuals to control and adjust their data-sharing preferences, ensuring transparency and continuity of consent in the ever-evolving landscape of digital health. Despite the challenges associated with privacy protections, the benefits, including improved diagnostics and treatment, are substantial. The availability of open databases has notably accelerated AI research, leading to significant advancements in medical diagnostics and treatments. As the landscape of healthcare research continues to evolve with open science and federated learning, the role of medical open databases remains crucial in shaping the future of medicine, promising enhanced patient outcomes and fostering a global research community committed to ethical integrity and privacy.

**Keywords**: Medical Open Databases; Artificial Intelligence; Privacy-Enhancing Technologies; Federated Learning; Dynamic Consent Systems

## 1. Introduction

The emergence of medical open databases, coupled with advances in artificial intelligence (AI), heralds a significant change in biomedical research and healthcare innovation, facilitating an era of enhanced accessibility and data sharing [1–3]. This movement towards open data science, augmented by AI technologies, enables researchers worldwide to access a wealth of physiological [4,5], biomedical [6], and healthcare information [7], fostering collaboration and speeding up the pace of medical discoveries. AI's role in analyzing vast datasets has been instrumental in uncovering patterns and insights that would be impossible for humans to detect unaided, leading to breakthroughs in understanding diseases and patient care. Initiatives like annual challenges and shared toolboxes have spurred the development of novel algorithms and techniques, leveraging AI to address complex biomedical challenges and advance medical diagnostics and treatments. This synergy between open medical databases and AI is transforming the landscape of healthcare, promising a future of more accurate, efficient, and personalized medicine.

Simultaneously, this rise in open data repositories brings to the forefront crucial privacy concerns [6,8]. The necessity to balance the imperative of research enablement with the protection of patient confidentiality has never been more pronounced. In this context, laws such as the Health Insurance Portability and Accountability Act (HIPAA) play a pivotal role in shaping the landscape of data de-identification and anonymization processes, ensuring that shared data comply with strict privacy standards [9]. Moreover, the introduction of privacy-enhancing technologies (PETs) such as differential privacy [10], synthetic data [11], homomorphic encryption [12], Secure Multiparty Computation [13], and Federated Learning [14] represents a proactive approach to safeguarding personal health information. These technologies provide the means to conduct meaningful research while upholding the principles of data privacy and security.

In the country like Taiwan, strides in medical data sharing suggest the global shift towards interconnected health systems, highlighting both advancements and ongoing challenges in securing

patient data. The implementation of dynamic consent frameworks reflects a growing recognition of the need for more flexible approaches to data privacy, particularly in an era of personalized medicine and digital health records [15]. As the landscape of medical research evolves with these developments, the interplay of data sharing, privacy, and technology continues to reshape the boundaries of what is possible in healthcare innovation, marking a critical junction in the journey towards more open, collaborative, and ethically responsible research environments.

## 2. The rise of medical open database

The rise of medical open databases represents a transformative shift in the landscape of biomedical research and healthcare innovation. Among the pioneers in this movement is PhysioNet. Established in 1999, PhysioNet is a pioneering open database that provides free access to a wide range of physiologic signals and related open-source software for research in medicine, physiology, and biomedical engineering [4,16]. It was initiated by a collaborative project involving researchers from Boston's Beth Israel Deaconess Medical Center/Harvard Medical School, Boston University, McGill University, and Massachusetts Institute of Technology [4,17]. The database contains a diverse collection of physiological datasets, including those related to cardiovascular and other complex biomedical signals [4,18]. PhysioNet has had a significant impact on the development of medical open databases, serving as a model for the establishment of similar resources. It has also played a key role in promoting the dissemination and exchange of medical resources.

A significant contribution of PhysioNet to the scientific community is its annual PhysioNet Challenge, which has markedly influenced the field by promoting innovation and collaboration among researchers and clinicians. These challenges stimulate the creation of novel algorithms and methods aimed at solving complex biomedical problems, thus expanding the limits of what can be achieved in medical data analysis and application. For instance, the challenges have catalyzed the development of innovative algorithms capable of detecting obstructive sleep apnea from

electrocardiograms [19], illustrating the practical impact of these competitions on advancing medical diagnostics and treatment strategies [20].

Since the success of PhysioNet, numerous other medical open databases have emerged globally, fostering a more cooperative and transparent research atmosphere. One notable example is the UK Biobank [21], launched in 2006, which provides a vast repository of genetic and health information from half a million UK participants. This database has become an essential tool for unraveling the complex interplay between genetics, lifestyle, and disease, thereby enhancing our understanding of the factors influencing human health [22–25]. Similarly, the Cancer Imaging Archive, inaugurated in 2011 in the United States, offers a dedicated platform for the cancer research community, enabling access to a comprehensive array of imaging datasets [26]. These initiatives, along with PhysioNet, suggest the accelerating shift towards open science within the medical sector. They display how shared resources can drive forward innovation and improve patient care worldwide, illustrating the critical role of collaborative environments in the advancement of healthcare research and application.

Additionally, Kaggle, an online platform for data science and machine learning competitions, has emerged as a pivotal player in the field of machine learning analysis and data sharing [27]. Launched in 2010, Kaggle facilitates collaboration and competition among data scientists and researchers by hosting challenges in various domains, including healthcare. These competitions often involve complex medical datasets, encouraging participants to develop innovative solutions and algorithms for disease prediction, medical imaging analysis [28], and other health-related issues. Kaggle has not only democratized access to large medical datasets but has also fostered a global community where knowledge and techniques are openly shared. This environment has led to significant breakthroughs and advancements in medical research and analytics, further suggesting the importance of open data and collaborative problem-solving in improving healthcare outcomes and accelerating medical innovation.

## 3. Balancing privacy and the need of medical open databases

The success of open medical databases such as PhysioNet poses the challenge of balancing patient confidentiality with research enablement on open platforms [29,30]. Research datasets in healthcare often contain protected health information (PHI), and the process of removing this information—a process known as de-identification or anonymization—can be challenging and prone to errors [31]. Despite the utility of these datasets, the need for de-identification introduces a significant barrier to data sharing due to the effort and cost involved.

The HIPAA, established in the United States in 1996, plays a vital role in safeguarding patients' medical information. In response to the HIPAA mandate, U.S. Department of Health and Human Service published a final regulation in the form of the Privacy Rule in December 2000, which became effective on April 14, 2001. Central to this rule is the designation of 18 specific categories of PHI that, if disclosed, could be used to identify an individual (Table 1). These categories encompass a broad spectrum of personal data, including, but not limited to, names, geographic details smaller than a state, various identifiers like social security numbers, medical record numbers, and contact information, as well as certain biometric and photographic images [32].

Additionally, HIPAA mandates that covered entities must ensure they do not possess knowledge that the remaining information could be used, whether alone or in conjunction with other data, to identify the subject. By strictly adhering to these guidelines, entities can share de-identified health information for broader uses, such as public health and research, without infringing on individual privacy rights, thus striking a balance between privacy protection and the beneficial use of health data [33]. Despite these efforts, the tension between the promise of big data and patient privacy in healthcare research remains a challenge [34].

PhysioNet is a pioneer in medical public databases, ensuring that the datasets it provides do not compromise individual privacy. This involves ensuring that any data shared does not contain PHI or has been sufficiently anonymized to prevent the identification of individuals. The challenges posed

by the HIPAA Privacy Rule are not insignificant; they include the need for informed consent from data subjects and potential limitations on access to health information which can hinder clinical research [35,36]. Furthermore, the rule's interaction with other regulations, like the Common Rule, adds complexity to privacy concerns in research, leading to inconsistencies and additional burdens for researchers.

Despite these challenges, the Privacy Rule does allow for certain disclosures without patient authorization, particularly for public health purposes. This is intended to facilitate the use of medical data in important public health endeavors without undermining individual privacy protections [37]. The balance sought by the HIPAA Privacy Rule between protecting privacy and facilitating research is a critical aspect of its implementation, particularly in the context of medical open databases. By navigating these regulations successfully, repositories can contribute to the advancement of medical research while ensuring compliance with privacy standards [38].

The emergence of medical databases, such as the PhysioNet, UK Biobank and the Cancer Imaging Archive, has significantly advanced collaborative research in healthcare [39,40]. These databases have the potential to transform cancer research and improve patient outcomes [39]. However, the collection, linking, and use of data in biomedical research raise ethical concerns, particularly regarding privacy and security [30,41,42]. Despite these concerns, the benefits of open data in healthcare, including improved diagnostics and treatment, are substantial [42]. The push for data sharing in cancer trials by pharmaceutical companies further underscores the importance of open medical databases in driving innovation and improving patient care [43].

## 4. Privacy enhancing technology

A range of studies have been conducted to explore the increasing frequency and impact of healthcare data breaches, highlighting the rising number of incidents and their detrimental effects on patient privacy and healthcare providers [44–47]. These breaches are often caused by a combination

of technical, organizational, and human factors [44–46]. Human vulnerabilities, such as lack of

awareness and training, play a significant role in these breaches [45]. The use of the Swiss Cheese

Model can help assess vulnerabilities and risks [44]. Cloud computing breaches are a particular

concern, highlighting the need for digital forensic readiness [48]. Hacking and unauthorized internal

disclosures are the most prevalent forms of attack [47]. Further studies may examine specific cases

and the implications for digital forensic readiness, emphasizing the importance of adhering to

regulations.

Below, we reviewed several PETs and their applications in enhancing data privacy and security

in healthcare settings. PETs, such as encryption, anonymization techniques, and secure multi-party

computation, offer powerful mechanisms to protect sensitive health data. Implementing these

technologies, alongside robust privacy policies and employee training, can significantly reduce the

likelihood of data breaches and bolster the trust between patients and healthcare providers.

Differential privacy, a method for protecting individual privacy in data analysis, has been

increasingly applied in the healthcare sector. It involves adding noise to the data to prevent re-

identification of individuals. This approach has been used in various areas of health research,

including genomics, neuroimaging, and health surveillance [49]. However, there are challenges in its

practical application, such as the theoretical nature of the privacy parameter epsilon [50]. To address

these challenges, researchers have proposed differentially private data release strategies and noise

mechanisms, such as the Laplace and exponential mechanisms [51]. The application of differential

privacy in medical questionnaires has also been explored, with the Randomized Response

mechanism showing promise in improving privacy while retaining data utility [52]. Furthermore, the

use of differential privacy in geospatial analyses of standardized healthcare data has been

demonstrated, with the development of geodatabase functions for privacy-aware analysis [53].

Lastly, the combination of differential privacy and decision tree approach has been proposed for data

publishing, and the differentially private mini-batch gradient descent algorithm for model publishing

of medical data [54].

Synthetic data, generated through simulators, is increasingly used in healthcare to address the challenges of data availability and privacy [55]. Privacy-enhancing technologies, such as differential privacy, are combined with synthetic data generators to create private synthetic data, preserving statistical properties while ensuring privacy [56]. These technologies have been applied in various use cases, including clinical risk prediction [57] and medical research [58]. However, the evaluation of synthetic data's privacy and utility metrics remains a challenge, with a lack of consensus on standard approaches [59]. Despite these challenges, the potential of synthetic data in preserving data utility and patient privacy in electronic healthcare data is being explored [60].

Homomorphic encryption, a powerful tool for preserving privacy in medical data, allows for computations to be performed on encrypted data without the need for decryption. It has been successfully applied in various medical data scenarios, including machine learning models for classification and training, secure genomic algorithms, and predictive analysis tasks [61–63]. For example, it has been used to securely manage personal health metrics data, process medical images [64,65], and enable secure medical computation [66]. The use of homomorphic encryption in these applications ensures that sensitive medical data remains private and secure.

Secure Multiparty Computation is a cryptographic technique that enables data analytics without sharing the underlying data, making it a valuable tool for preserving privacy in medical data analysis [67]. It has been applied in various healthcare scenarios, including collaborative systems [68], statistical analysis of health data [69], and electronic medical record data [69]. Secure Multiparty Computation has also been used in healthcare Internet of Things systems to handle privacy issues [70], prevent data disclosure in sensor networks [71], and enable the reuse of distributed electronic health data [69]. Furthermore, it has been applied to enable privacy-preserving query processing on electronic medical records [72].

Federated learning, a decentralized machine learning approach, is increasingly being applied in

the medical field due to the sensitive and fragmented nature of healthcare data [14,73]. It allows for the collaborative development of machine learning models without sharing raw data, thus preserving privacy [74,75]. This approach has been used in various medical domains, including oncology and radiology, for tasks such as image analysis and disease prediction [75,76]. However, there are challenges to be addressed, such as data homogeneity and transparency [75]. Despite these challenges, federated learning shows promise in improving the efficiency and privacy of medical data processing [77–79].

We have proposed a method of Dynamically Synthetic Images method for Federated Learning (DSIFL), significantly improving the conventional federated learning framework by integrating local information from local multiple institutions with heterogeneous data types [80]. DSIFL's key technique is to dynamically adjust the number of synthetic images based on local data misclassified by the current model. Incorporating these synthetic images into training enables the global model to better handle the diversity of heterogeneous data types collected by local medical institutions. Experimental results showed that under different learning frameworks, DSIFL's performance can match or exceed traditional centralized learning methods [80].

## 5. Taiwan medical AI and data portal and dynamic consents system

Taiwan has made significant strides in medical data sharing, particularly in the areas of privacy protection and electronic health records exchange. The country's comprehensive embedded integrated circuit-based health insurance card system, implemented by the Bureau of National Health Insurance, Taiwan, allows for the secure sharing of health information [81]. The use of blockchain technology has been proposed as a means to further enhance the security and privacy of medical data sharing [82,83]. The Taiwan Electronic Medical Record Template and the National Electronic Medical Record Exchange System have been developed to facilitate the exchange of electronic medical records [84,85]. However, concerns about unauthorized access and secondary use of

electronic medical records persist, particularly among highly educated individuals [86]. The country has also established guidelines for the security and privacy protection of health information, drawing on international best practices [87].'

In the past two years, funded by the National Science and Technology Council of Taiwan, we have assembled teams from National Yang Ming Chiao Tung University, Taipei Veterans General Hospital, Academia Sinica, and National Taipei University of Nursing and Health Sciences to form a data repository task force known as the Smart Medical AI and Repository Taskforce Center. We launched a medical AI and data-sharing platform aimed at advancing the field of medical AI research in October 2023 [88]. This platform not only provides the public and researchers with access to a multitude of shared datasets but also ensures a meticulous evaluation process (Fig. 1). Researchers can apply for access to the data by providing an abstract of their research proposal. Dataset managers assess applications based on their intended use, specific needs, and detailed research plans.

Currently, seventeen datasets have been released on the platform, and we plan to ultimately integrate more than 40 comprehensive clinical databases. This effort aims to advance research across seven crucial clinical areas that greatly benefit from AI technology, including heart disease, neurological disorders, mental illness, diabetes, cancer, genetic predispositions to complex diseases, and medical imaging. Moreover, the platform underpins collaboration between distinct teams specializing in AI methodology, science and law, and data governance, jointly fostering a robust data governance framework that emphasizes federated learning, cloud-based AI solutions, and trusted AI practices. Importantly, the system is designed to streamline the research process while maintaining a focus on ethical standards and participant privacy. In line with this, the platform incorporates a dynamic informed consent mechanism, especially for datasets that are anonymized but cannot be completely separated from their sources. This approach ensures that participants' privacy is safeguarded while also enabling their informed and ongoing consent, reflecting our commitment to ethical research practices and the dynamic nature of consent in medical studies.

Dynamic informed consent, a concept that has been explored in various contexts, is a personalized, digital communication interface that allows participants to manage their consent preferences [89]. It has been proposed as a solution to improve patient confidence and trust in the use of electronic patient records in medical research [90]. In the context of privacy-aware pervasive health and well-being, dynamic consent enables granular data consent and management [91]. It has also been suggested as a potential solution to challenges in modern biomedical research, including participant recruitment, informed consent, and consent management [92]. The use of blockchain technology has been proposed to enhance the privacy-preserving aspects of dynamic consent in genomic data sharing [93]. The concept of dynamic informed consent has been further explored in the context of personalized medicine, emphasizing the need for a more dynamic and enriched consent model [94].

To enhance privacy of participants contributing to the data in our platform and increase participants' engagement, we have developed a dynamic informed consent system, named the Health Data Authorization Service Platform (Fig. 2). This collaborative effort involves our data governance, humanities, science and law, and clinical teams, aiming to facilitate scalable, dynamic consent operations suitable for complex data environments. The platform supports flexible data governance, allowing data owners to dynamically express their consent preferences, thereby making dynamic consent practical and sustainable.

The platform's architecture encompasses four key roles: the Resource Owner, usually the participants, who owns the data; the Resource Server, which stores and manages medical data; the Requesting Party, typically researchers seeking data access; and the Authorization Server, the backbone of our dynamic consent system, connecting the three roles. This system ensures that data is accessed only with the owner's express consent, respecting their preferences and enhancing data use transparency. Resource owners can modify their consent settings at any time, reflecting changes in their willingness to share different data types, like electronic medical records or medical imaging, for

specific purposes such as academic research. This flexibility ensures that data use aligns with the owners' current preferences. The platform seamlessly integrates with our shared data framework, maintaining each citizen's consent history and enabling swift updates to consent forms as needed. By streamlining the consent process and ensuring data is shared according to owner permissions, our platform respects individual preferences while promoting responsible data science development. It exemplifies a forward-thinking approach to data governance, enabling real-time adjustments in consent and fostering a culture of trust and transparency between data owners and users.

## 6. Acceleration of medical AI research through open databases

The advent of medical open databases has significantly accelerated the field of medical AI research, fostering an environment of innovation and rapid development [95]. By providing researchers with access to vast amounts of health-related data, these databases have become a cornerstone for advancements in predictive analytics, diagnostic algorithms, and personalized medicine.

One of the most notable contributions of open medical databases to AI research is the democratization of data [96]. Historically, the scarcity and inaccessibility of medical data posed substantial hurdles to AI development. However, platforms like PhysioNet, established in 1999, have bridged this gap by offering a plethora of datasets ranging from physiological signals to clinical outcomes [97]. However, challenges remain, including the need for large datasets and the lack of external validation in perioperative medicine [98]. The use of open science approaches, including data liberation and crowdsourcing, can help address these challenges [99]. The integration of networked medical devices and clinical repositories based on open standards can further enhance AI research in high acuity medical environments [100]. This enhanced availability allows researchers from diverse backgrounds and institutions to engage in healthcare innovation, leveling the playing field and stimulating a surge in AI-based solutions.

Open medical databases encompass a wide variety of data types, including electronic medical records, imaging, genomic sequences, and more. This diversity enables AI researchers to explore multifaceted healthcare questions, from predicting disease trajectories to optimizing treatment plans. Moreover, the rich, varied datasets facilitate the training of more robust and generalizable AI models, capable of addressing complex medical scenarios across different populations and settings. The shared nature of open databases fosters collaboration across the global research community [101]. Through platforms that offer shared data, researchers can combine their expertise to tackle larger and more complex problems than they could individually. This collaborative approach has led to significant breakthroughs in AI, such as algorithms that can detect diseases from images with accuracy rivaling that of trained professionals [102,103].

Open databases also streamline the validation and implementation phases of AI development [104]. Access to diverse datasets enables researchers to rigorously test their algorithms under various conditions and patient demographics, ensuring their reliability and effectiveness. The expansion of these databases has significantly propelled medical AI research forward, marking a new phase of healthcare innovation with faster discoveries, collaborative efforts, and a commitment to ethical data use. As the field evolves, the role of open databases in shaping the future of medicine remains pivotal.

7. **Conclusions**

In conclusion, the evolution towards medical open databases, exemplified by the inception of platforms like PhysioNet in 1999 and their progression over the past 25 years, alongside the integration of privacy-enhancing technologies, marks a significant milestone in the domain of biomedical research and healthcare innovation. This journey not only fosters an unprecedented level of collaboration and accessibility but also emphasizes the crucial need to address privacy concerns and ethical considerations diligently. The ongoing efforts to balance data sharing with individual

privacy protection are underscored by the adaptation of legal frameworks and the implementation of cryptographic and data management solutions. The introduction and growth of medical open databases have been pivotal, providing a wealth of data that has propelled research and innovation while highlighting the challenges and responsibilities of managing sensitive information. Specifically, the availability of open medical databases has significantly accelerated AI research, leading to breakthroughs in disease prediction, diagnostics, and personalized medicine. As we continue to explore the vast potential of open science and federated learning, the landscape of healthcare research is on the brink of remarkable transformations. These advancements promise enhanced patient outcomes, faster medical discoveries, and a more inclusive global research community, all achieved by adhering to the highest standards of privacy and ethical integrity.

**Acknowledgements**

**Contributions**

A.C.Y wrote the manuscript. All authors discussed the content and approved the manuscript.

**Conflict of Interest**

On behalf of all authors, the corresponding author states that there is no conflict of interest.

# References

1.  Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D. Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. BioMed Res Int 2014;2014:1–13. doi: 10.1155/2014/134023

2.  Roski J, Bo-Linn GW, Andrews TA. Creating Value In Health Care Through Big Data: Opportunities And Policy Implications. Health Aff (Millwood) 2014 Jul;33(7):1115–1122. doi: 10.1377/hlthaff.2014.0147

3.  Wang Y, Kung L, Byrd TA. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technol Forecast Soc Change 2018 Jan;126:3–13. doi: 10.1016/j.techfore.2015.12.019

4.  Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000 Jun 13;101(23):E215-20.

5.  Dean DA, Goldberger AL, Mueller R, Kim M, Rueschman M, Mobley D, Sahoo SS, Jayapandian CP, Cui L, Morrical MG, Surovec S, Zhang G-Q, Redline S. Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource. Sleep 2016 May;39(5):1151–1164. doi: 10.5665/sleep.5774

6.  Wylie JE, Mineau GP. Biomedical databases: protecting privacy and promoting research. Trends Biotechnol 2003 Mar;21(3):113–116. doi: 10.1016/S0167-7799(02)00039-2

7.  Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010 Mar 1;17(2):124–130. doi: 10.1136/jamia.2009.000893

8.  Shahin MH, Bhattacharya S, Silva D, Kim S, Burton J, Podichetty J, Romero K, Conrado DJ. Open Data Revolution in Clinical Research: Opportunities and Challenges. Clin Transl Sci 2020 Jul;13(4):665–674. doi: 10.1111/cts.12756

9.  Ness RB. Influence of the HIPAA Privacy Rule on Health Research. JAMA 2007 Nov 14;298(18):2164. doi: 10.1001/jama.298.18.2164

10. Dwork C, Roth A. The Algorithmic Foundations of Differential Privacy. Found Trends® Theor Comput Sci 2013;9(3–4):211–407. doi: 10.1561/0400000042

11. Rubin DB. Discussion: Statistical Disclosure Limitation. J Off Stat 9:461–468.

12. Yi X, Paulet R, Bertino E. Homomorphic Encryption. Homomorphic Encryption Appl Cham: Springer International Publishing; 2014. p. 27–46. doi: 10.1007/978-3-319-12229-8_2

13. Cramer RJF, Damgård IB, Nielsen JB. Secure multiparty computation and secret sharing. Cambrdige, United Kingdom New York: Cambridge University Press; 2015. ISBN:978-1-107-04305-3

14. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated Learning for Healthcare Informatics. J Healthc Inform Res 2021 Mar;5(1):1–19. doi: 10.1007/s41666-020-00082-4

15.  Packer M. Data sharing in medical research. BMJ 2018 Feb 14;k510. doi: 10.1136/bmj.k510

16.  Moody GB, Mark RG, Goldberger AL. PhysioNet: Physiologic signals, time series and related open source software for basic, clinical, and applied research. 2011 Annu Int Conf IEEE Eng Med Biol Soc Boston, MA: IEEE; 2011. p. 8327–8330. doi: 10.1109/IEMBS.2011.6092053

17.  Moody GB, Mark RG, Goldberger AL. PhysioNet: a research resource for studies of complex physiologic and biomedical signals. Comput Cardiol 2000 Vol27 Cat 00CH37163 Cambridge, MA, USA: IEEE; 2000. p. 179–182. doi: 10.1109/CIC.2000.898485

18.  Henry IC, Goldberger AL, Moody GB, Mark RG. PhysioNet: an NIH research resource for physiologic datasets and open-source software. Proc 14th IEEE Symp Comput-Based Med Syst CBMS 2001 Bethesda, MD, USA: IEEE Comput. Soc; 2001. p. 245–250. doi: 10.1109/CBMS.2001.941728

19.  Thomas RJ, Mietus JE, Peng C-K, Goldberger AL. An Electrocardiogram-Based Technique to Assess Cardiopulmonary Coupling During Sleep. Sleep 2005 Sep;28(9):1151–1161. doi: 10.1093/sleep/28.9.1151

20.  Thomas RJ, Mietus JE, Peng C-K, Gilmartin G, Daly RW, Goldberger AL, Gottlieb DJ. Differentiating Obstructive from Central and Complex Sleep Apnea Using an Automated Electrocardiogram-Based Method. Sleep 2007 Dec;30(12):1756–1769. doi: 10.1093/sleep/30.12.1756

21.  Collins R. What makes UK Biobank special? The Lancet 2012 Mar;379(9822):1173–1174. doi: 10.1016/S0140-6736(12)60404-8

22.  Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018 Oct;562(7726):203–209. doi: 10.1038/s41586-018-0579-z

23.  Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, Gallacher J, Green J, Matthews P, Pell J, Sprosen T, Collins R. UK Biobank: Current status and what it means for epidemiology. Health Policy Technol 2012 Sep;1(3):123–126. doi: 10.1016/j.hlpt.2012.07.003

24.  Littlejohns TJ, Holliday J, Gibson LM, Garratt S, Oesingmann N, Alfaro-Almagro F, Bell JD, Boultwood C, Collins R, Conroy MC, Crabtree N, Doherty N, Frangi AF, Harvey NC, Leeson P, Miller KL, Neubauer S, Petersen SE, Sellors J, Sheard S, Smith SM, Sudlow CLM, Matthews PM, Allen NE. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nat Commun 2020 May 26;11(1):2624. doi: 10.1038/s41467-020-15948-9

25.  Littlejohns TJ, Sudlow C, Allen NE, Collins R. UK Biobank: opportunities for cardiovascular research. Eur Heart J 2019 Apr 7;40(14):1158–1166. doi: 10.1093/eurheartj/ehx254

26.  Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. J Digit Imaging 2013 Dec;26(6):1045–1057. doi: 10.1007/s10278-013-9622-7

27. Kaggle. Kaggle. Available from: https://www.kaggle.com/

28. Yang X, Zeng Z, Teo SG, Wang L, Chandrasekhar V, Hoi S. Deep Learning for Practical Image Recognition: Case Study on Kaggle Competitions. Proc 24th ACM SIGKDD Int Conf Knowl Discov Data Min London United Kingdom: ACM; 2018. p. 923–931. doi: 10.1145/3219819.3219907

29. Krishna R, Kelleher K, Stahlberg E. Patient Confidentiality in the Research Use of Clinical Medical Databases. Am J Public Health 2007 Apr;97(4):654–658. doi: 10.2105/AJPH.2006.090902

30. Kobayashi S, Kane T, Paton C. The Privacy and Security Implications of Open Data in Healthcare: A Contribution from the IMIA Open Source Working Group. Yearb Med Inform 2018 Aug;27(01):041–047. doi: 10.1055/s-0038-1641201

31. Eze B, Peyton L. Systematic Literature Review on the Anonymization of High Dimensional Streaming Datasets for Health Data Sharing. Procedia Comput Sci 2015;63:348–355. doi: 10.1016/j.procs.2015.08.353

32. Thacker S. HIPAA privacy rule and public health. Guidance from CDC and the U.S. Department of Health and Human Services. MMWR Supplements; 2003.

33. Prasser F, Kohlmayer F, Spengler H, Kuhn KA. A Scalable and Pragmatic Method for the Safe Sharing of High-Quality Health Data. IEEE J Biomed Health Inform 2018 Mar;22(2):611–622. doi: 10.1109/JBHI.2017.2676880

34. Sarpatwari A, Gagne JJ. Balancing benefits and harms: privacy protection policies: Privacy Protection Policies. Pharmacoepidemiol Drug Saf 2016 Aug;25(8):969–971. doi: 10.1002/pds.4048

35. Kulynych J, Korn D. The New HIPAA (Health Insurance Portability and Accountability Act of 1996) Medical Privacy Rule: Help or Hindrance for Clinical Research? Circulation 2003 Aug 26;108(8):912–914. doi: 10.1161/01.CIR.0000080642.35380.50

36. Nosowsky R, Giordano TJ. The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule: Implications for Clinical Research. Annu Rev Med 2006 Feb 1;57(1):575–590. doi: 10.1146/annurev.med.57.121304.131257

37. Gostin LO, Nass S. Reforming the HIPAA Privacy Rule: Safeguarding Privacy and Promoting Research. JAMA 2009 Apr 1;301(13):1373. doi: 10.1001/jama.2009.424

38. Wartenberg D, Thompson WD. Privacy Versus Public Health: The Impact of Current Confidentiality Rules. Am J Public Health 2010 Mar;100(3):407–412. doi: 10.2105/AJPH.2009.166249

39. Green AK, Reeder-Hayes KE, Corty RW, Basch E, Milowsky MI, Dusetzina SB, Bennett AV, Wood WA. The Project Data Sphere Initiative: Accelerating Cancer Research by Sharing Data. The Oncologist 2015 May 1;20(5):464-e20. doi: 10.1634/theoncologist.2014-0431

40. Conroy M, Sellors J, Effingham M, Littlejohns TJ, Boultwood C, Gillions L, Sudlow CLM, Collins R, Allen NE. The advantages of UK Biobank's open-access strategy for health research.

J Intern Med 2019 Oct;286(4):389–397. doi: 10.1111/joim.12955

41. Anderson R. The collection, linking and use of data in biomedical research and health care: ethical issues. [object Object]; 2015 Feb. doi: 10.17863/CAM.31760

42. Kostkova P, Brewer H, De Lusignan S, Fottrell E, Goldacre B, Hart G, Koczan P, Knight P, Marsolier C, McKendry RA, Ross E, Sasse A, Sullivan R, Chaytor S, Stevenson O, Velho R, Tooke J. Who Owns the Data? Open Data for Healthcare. Front Public Health 2016 Feb 17;4. doi: 10.3389/fpubh.2016.00007

43. Bhattacharjee Y. Pharma Firms Push for Sharing of Cancer Trial Data. Science 2012 Oct 5;338(6103):29–29. doi: 10.1126/science.338.6103.29

44. Kamoun F, Nicho M. Human and Organizational Factors of Healthcare Data Breaches: The Swiss Cheese Model of Data Breach Causation And Prevention. Int J Healthc Inf Syst Inform 2014 Jan 1;9(1):42–60. doi: 10.4018/ijhisi.2014010103

45. Nifakos S, Chandramouli K, Nikolaou CK, Papachristou P, Koch S, Panaousis E, Bonacina S. Influence of Human Factors on Cyber Security within Healthcare Organisations: A Systematic Review. Sensors 2021 Jul 28;21(15):5119. doi: 10.3390/s21155119

46. McLeod A, Dolezel D. Cyber-analytics: Modeling factors associated with healthcare data breaches. Decis Support Syst 2018 Apr;108:57–68. doi: 10.1016/j.dss.2018.02.007

47. Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, Ahmad Khan R. Healthcare Data Breaches: Insights and Implications. Healthcare 2020 May 13;8(2):133. doi: 10.3390/healthcare8020133

48. Chernyshev M, Zeadally S, Baig Z. Healthcare Data Breaches: Implications for Digital Forensic Readiness. J Med Syst 2019 Jan;43(1):7. doi: 10.1007/s10916-018-1123-2

49. Ficek J, Wang W, Chen H, Dagne G, Daley E. Differential privacy in health research: A scoping review. J Am Med Inform Assoc 2021 Sep 18;28(10):2269–2276. doi: 10.1093/jamia/ocab135

50. Dankar F, Emam K. Practicing Differential Privacy in Health Care: A Review. Transactions on Data Privacy; 2013.

51. Tamane S, Solanki VK, Dey N, editors. Privacy and Security Policies in Big Data: IGI Global; 2017. doi: 10.4018/978-1-5225-2486-1ISBN:978-1-5225-2486-1

52. Appenzeller A, Terzer N, Philipp P, Beyerer J. Applying Differential Privacy to Medical Questionnaires. 2023 IEEE Int Conf Pervasive Comput Commun Workshop Affil Events PerCom Workshop Atlanta, GA, USA: IEEE; 2023. p. 608–613. doi: 10.1109/PerComWorkshops56833.2023.10150373

53. Harris DR. Leveraging Differential Privacy in Geospatial Analyses of Standardized Healthcare Data. 2020 IEEE Int Conf Big Data Big Data Atlanta, GA, USA: IEEE; 2020. p. 3119–3122. doi: 10.1109/BigData50022.2020.9378390

54. Sun Z, Wang Y, Shu M, Liu R, Zhao H. Differential Privacy for Data and Model Publishing of Medical Data. IEEE Access 2019;7:152103–152114. doi: 10.1109/ACCESS.2019.2947295

55. McDuff D, Curran T, Kadambi A. Synthetic Data in Healthcare. [object Object]; 2023; doi: 10.48550/ARXIV.2304.03243

56. Appenzeller A, Leitner M, Philipp P, Krempel E, Beyerer J. Privacy and Utility of Private Synthetic Data for Medical Data Analyses. Appl Sci 2022 Dec 1;12(23):12320. doi: 10.3390/app122312320

57. Qian Z, Callender T, Cebere B, Janes SM, Navani N, Van Der Schaar M. Synthetic data for privacy-preserving clinical risk prediction. Epidemiology; 2023 May. doi: 10.1101/2023.05.18.23290114

58. Kokosi T, Harron K. Synthetic data in medical research. BMJ Med 2022 Sep;1(1):e000167. doi: 10.1136/bmjmed-2022-000167

59. Kaabachi B, Despraz J, Meurers T, Otte K, Halilovic M, Prasser F, Raisaro JL. Can We Trust Synthetic Data in Medicine? A Scoping Review of Privacy and Utility Metrics. Health Informatics; 2023 Nov. doi: 10.1101/2023.11.28.23299124

60. Wang Z, Myles P, Tucker A. Generating and evaluating cross-sectional synthetic electronic healthcare data: Preserving data utility and patient privacy. Comput Intell 2021 May;37(2):819–851. doi: 10.1111/coin.12427

61. Wood A, Najarian K, Kahrobaei D. Homomorphic Encryption for Machine Learning in Medicine and Bioinformatics. ACM Comput Surv 2021 Jul 31;53(4):1–35. doi: 10.1145/3394658

62. Bos JW, Lauter K, Naehrig M. Private predictive analysis on encrypted medical data. J Biomed Inform 2014 Aug;50:234–243. doi: 10.1016/j.jbi.2014.04.003

63. Bocu R, Costache C. A homomorphic encryption-based system for securely managing personal health metrics data. IBM J Res Dev 2018 Jan 1;62(1):1:1-1:10. doi: 10.1147/JRD.2017.2755524

64. Krishnegowda P, M. Boregowda A. Efficient matrix key homomorphic encryption of medical images. Indones J Electr Eng Comput Sci 2023 Jul 1;31(1):406. doi: 10.11591/ijeecs.v31.i1.pp406-416

65. Kartit A. New Approach Based on Homomorphic Encryption to Secure Medical Images in Cloud Computing. Trends Sci 2022 Apr 30;19(9):3970. doi: 10.48048/tis.2022.3970

66. Khedr A, Gulak G. SecureMed: Secure Medical Computation Using GPU-Accelerated Homomorphic Encryption Scheme. IEEE J Biomed Health Inform 2018 Mar;22(2):597–606. doi: 10.1109/JBHI.2017.2657458

67. Veeningen M, Chatterjea S, Horváth AZ, Spindler G, Boersma E, van der Spek P, van der Galiën O, Gutteling J, Kraaij W, Veugen T. Enabling Analytics on Sensitive Medical Data with Secure Multi-Party Computation. Stud Health Technol Inform 2018;247:76–80. PMID:29677926

68. Marwan M, Kartit A, Ouahmane H. Applying secure multi-party computation to improve collaboration in healthcare cloud. 2016 Third Int Conf Syst Collab SysCo Casablanca,

Morocco: IEEE; 2016. p. 1–6. doi: 10.1109/SYSCO.2016.7831325

69.  Yigzaw KY, Bellika JG. Evaluation of secure multi-party computation for reuse of distributed electronic health data. IEEE-EMBS Int Conf Biomed Health Inform BHI Valencia, Spain: IEEE; 2014. p. 219–222. doi: 10.1109/BHI.2014.6864343

70.  Sahinbas K, Catak FO. Secure Multi-party Computation-Based Privacy-Preserving Data Analysis in Healthcare IoT Systems. In: Kose U, Gupta D, Khanna A, Rodrigues JJPC, editors. Interpret Cogn Internet Things Healthc Cham: Springer International Publishing; 2023. p. 57–72. doi: 10.1007/978-3-031-08637-3_3

71.  Tso R, Alelaiwi A, Mizanur Rahman SM, Wu M-E, Hossain MS. Privacy-Preserving Data Communication Through Secure Multi-Party Computation in Healthcare Sensor Cloud. J Signal Process Syst 2017 Oct;89(1):51–59. doi: 10.1007/s11265-016-1198-2

72.  Tawfik AM, Sabbeh SF, EL-Shishtawy T. Privacy-Preserving Secure Multiparty Computation on Electronic Medical Records for Star Exchange Topology. Arab J Sci Eng 2018 Dec;43(12):7747–7756. doi: 10.1007/s13369-018-3122-5

73.  Li T, Sahu AK, Talwalkar A, Smith V. Federated Learning: Challenges, Methods, and Future Directions. IEEE Signal Process Mag 2020 May;37(3):50–60. doi: 10.1109/MSP.2020.2975749

74.  Nguyen DC, Pham Q-V, Pathirana PN, Ding M, Seneviratne A, Lin Z, Dobre O, Hwang W-J. Federated Learning for Smart Healthcare: A Survey. ACM Comput Surv 2023 Mar 31;55(3):1–37. doi: 10.1145/3501296

75.  Crowson MG, Moukheiber D, Arévalo AR, Lam BD, Mantena S, Rana A, Goss D, Bates DW, Celi LA. A systematic review of federated learning applications for biomedical data. Mordaunt DA, editor. PLOS Digit Health 2022 May 19;1(5):e0000033. doi: 10.1371/journal.pdig.0000033

76.  Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis ICh, Shi W. Federated learning of predictive models from federated Electronic Health Records. Int J Med Inf 2018 Apr;112:59–67. doi: 10.1016/j.ijmedinf.2018.01.007

77.  Antunes RS, André Da Costa C, Küderle A, Yari IA, Eskofier B. Federated Learning for Healthcare: Systematic Review and Architecture Proposal. ACM Trans Intell Syst Technol 2022 Aug 31;13(4):1–23. doi: 10.1145/3501813

78.  Guo K, Chen T, Ren S, Li N, Hu M, Kang J. Federated Learning Empowered Real-Time Medical Data Processing Method for Smart Healthcare. IEEE/ACM Trans Comput Biol Bioinform 2024;1–12. doi: 10.1109/TCBB.2022.3185395

79.  Pfitzner B, Steckhan N, Arnrich B. Federated Learning in a Medical Context: A Systematic Literature Review. ACM Trans Internet Technol 2021 Jun 23;21(2):1–31. doi: 10.1145/3412357

80.  Wu JC-H, Yu H-W, Tsai T-H, Lu HH-S. Dynamically Synthetic Images for Federated Learning of medical images. Comput Methods Programs Biomed 2023 Dec;242:107845. doi: 10.1016/j.cmpb.2023.107845

81.  Liu H-H. Use and disclosure of health information and protection of patient privacy in Taiwan.

Med Law 2010 Mar;29(1):87–101. PMID:22458000

82.  Jin H, Luo Y, Li P, Mathew J. A Review of Secure and Privacy-Preserving Medical Data Sharing. IEEE Access 2019;7:61656–61669. doi: 10.1109/ACCESS.2019.2916503

83.  Liang X, Zhao J, Shetty S, Liu J, Li D. Integrating blockchain for data sharing and collaboration in mobile healthcare applications. 2017 IEEE 28th Annu Int Symp Pers Indoor Mob Radio Commun PIMRC Montreal, QC: IEEE; 2017. p. 1–5. doi: 10.1109/PIMRC.2017.8292361

84.  Rau H-H, Hsu C-Y, Lee Y-L, Chen W, Jian W-S. Developing Electronic Health Records in Taiwan. IT Prof 2010 Mar;12(2):17–25. doi: 10.1109/MITP.2010.53

85.  Li Y-C (Jack), Yen J-C, Chiu W-T, Jian W-S, Syed-Abdul S, Hsu M-H. Building a National Electronic Medical Record Exchange System – Experiences in Taiwan. Comput Methods Programs Biomed 2015 Aug;121(1):14–20. doi: 10.1016/j.cmpb.2015.04.013

86.  Hwang H-G, Han H-E, Kuo K-M, Liu C-F. The Differing Privacy Concerns Regarding Exchanging Electronic Medical Records of Internet Users in Taiwan. J Med Syst 2012 Dec;36(6):3783–3793. doi: 10.1007/s10916-012-9851-1

87.  Yang C-M, Lin H-C, Chang P, Jian W-S. Taiwan's perspective on electronic medical records' security and privacy protection: Lessons learned from HIPAA. Comput Methods Programs Biomed 2006 Jun;82(3):277–282. doi: 10.1016/j.cmpb.2006.04.002

88.  Smart Medical AI and Repository Taskforce Center. Taiwan Medical AI and Data Portal. 2023. Available from: https://data.dmc.nycu.edu.tw/

89.  Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015 Feb;23(2):141–146. doi: 10.1038/ejhg.2014.71

90.  Williams H, Spencer K, Sanders C, Lund D, Whitley EA, Kaye J, Dixon WG. Dynamic Consent: A Possible Solution to Improve Patient Confidence and Trust in How Electronic Patient Records Are Used in Medical Research. JMIR Med Inform 2015 Jan 13;3(1):e3. doi: 10.2196/medinform.3525

91.  Lee H, Lee U. Toward Dynamic Consent for Privacy-Aware Pervasive Health and Well-being: A Scoping Review and Research Directions. IEEE Pervasive Comput 2022 Oct 1;21(4):25–32. doi: 10.1109/MPRV.2022.3210747

92.  Budin-Ljøsne I, Teare HJA, Kaye J, Beck S, Bentzen HB, Caenazzo L, Collett C, D'Abramo F, Felzmann H, Finlay T, Javaid MK, Jones E, Katić V, Simpson A, Mascalzoni D. Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. BMC Med Ethics 2017 Dec;18(1):4. doi: 10.1186/s12910-016-0162-9

93.  Albalwy F, Brass A, Davies A. A Blockchain-Based Dynamic Consent Architecture to Support Clinical Genomic Data Sharing (ConsentChain): Proof-of-Concept Study. JMIR Med Inform 2021 Nov 3;9(11):e27816. PMID:34730538

94.  Goncharov L, Suominen H, Cook M. Dynamic consent and personalised medicine. Med J Aust

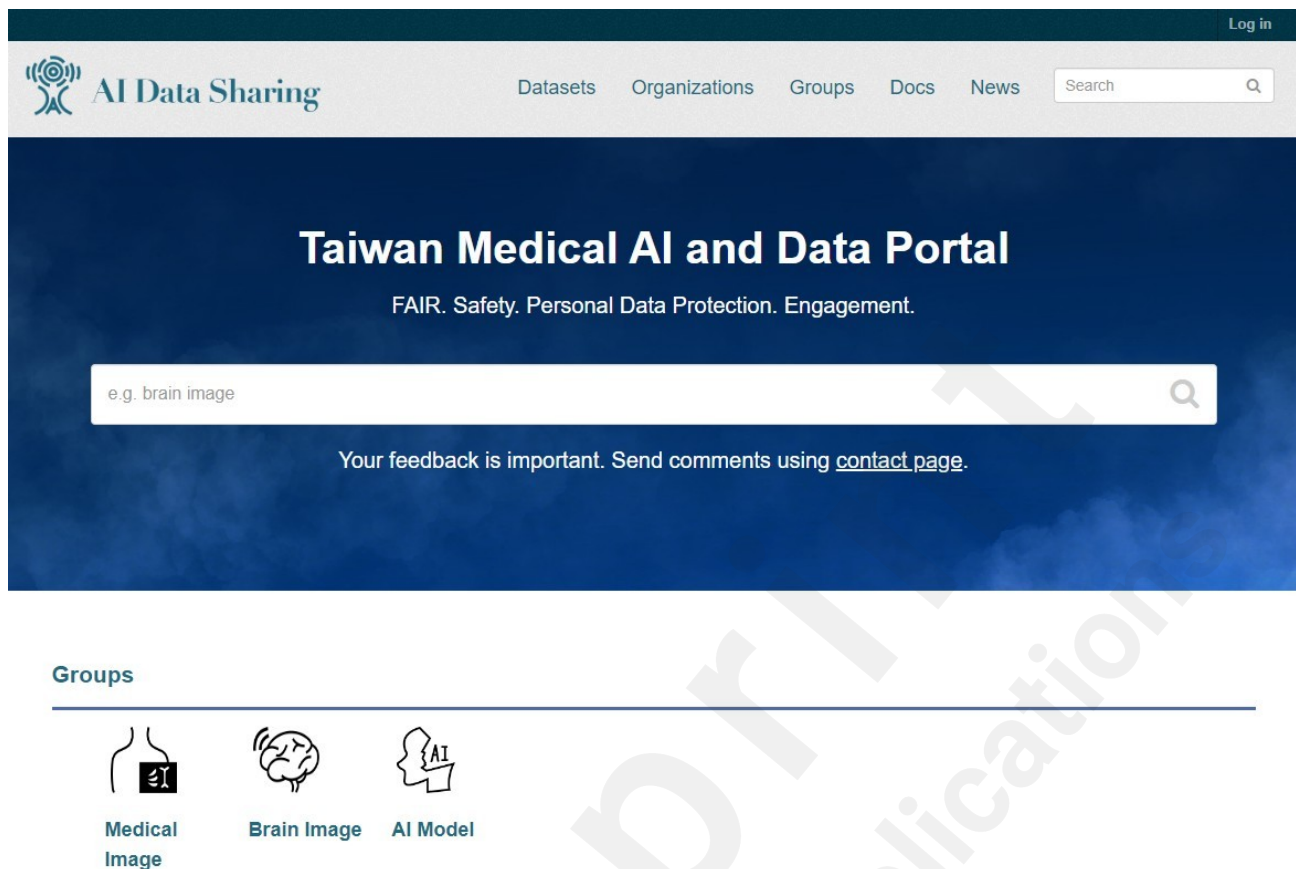2022 Jun 20;216(11):547–549. PMID:35611469

95.  Dilsizian SE, Siegel EL. Artificial Intelligence in Medicine and Cardiac Imaging: Harnessing Big Data and Advanced Computing to Provide Personalized Medical Diagnosis and Treatment. Curr Cardiol Rep 2014 Jan;16(1):441. doi: 10.1007/s11886-013-0441-8

96.  Pereira T, Morgado J, Silva F, Pelter MM, Dias VR, Barros R, Freitas C, Negrão E, Flor De Lima B, Correia Da Silva M, Madureira AJ, Ramos I, Hespanhol V, Costa JL, Cunha A, Oliveira HP. Sharing Biomedical Data: Strengthening AI Development in Healthcare. Healthcare 2021 Jun 30;9(7):827. doi: 10.3390/healthcare9070827

97.  Paton C, Kobayashi S. An Open Science Approach to Artificial Intelligence in Healthcare: A Contribution from the International Medical Informatics Association Open Source Working Group. Yearb Med Inform 2019 Aug;28(01):047–051. doi: 10.1055/s-0039-1677898

98.  Lim L, Lee H-C. Open datasets in perioperative medicine: a narrative review. Anesth Pain Med 2023 Jul 31;18(3):213–219. doi: 10.17085/apm.23076

99.  Wilson JR, Prevedello LM, Witiw CD, Flanders AE, Colak E. Data Liberation and Crowdsourcing in Medical Research: The Intersection of Collective and Artificial Intelligence. Radiol Artif Intell 2024 Jan 1;6(1):e230006. doi: 10.1148/ryai.230006

100. Kasparick M, Andersen B, Franke S, Rockstroh M, Golatowski F, Timmermann D, Ingenerf J, Neumuth T. Enabling artificial intelligence in high acuity medical environments. Minim Invasive Ther Allied Technol 2019 Mar 4;28(2):120–126. doi: 10.1080/13645706.2019.1599957

101. Milham MP, Craddock RC, Son JJ, Fleischmann M, Clucas J, Xu H, Koo B, Krishnakumar A, Biswal BB, Castellanos FX, Colcombe S, Di Martino A, Zuo X-N, Klein A. Assessment of the impact of shared brain imaging data on the scientific literature. Nat Commun 2018 Jul 19;9(1):2818. doi: 10.1038/s41467-018-04976-1

102. Patterson E, McBurney R, Schmidt H, Baldini I, Mojsilovic A, Varshney KR. Dataflow representation of data analyses: Toward a platform for collaborative data science. IBM J Res Dev 2017 Nov 1;61(6):9:1-9:13. doi: 10.1147/JRD.2017.2736278

103. El Naqa I, Haider MA, Giger ML, Ten Haken RK. Artificial Intelligence: reshaping the practice of radiological sciences in the 21st century. Br J Radiol 2020 Feb 1;93(1106):20190855. doi: 10.1259/bjr.20190855

104. Kras A, Celi LA, Miller JB. Accelerating ophthalmic artificial intelligence research: the role of an open access data repository. Curr Opin Ophthalmol 2020 Sep;31(5):337–350. doi: 10.1097/ICU.0000000000000678
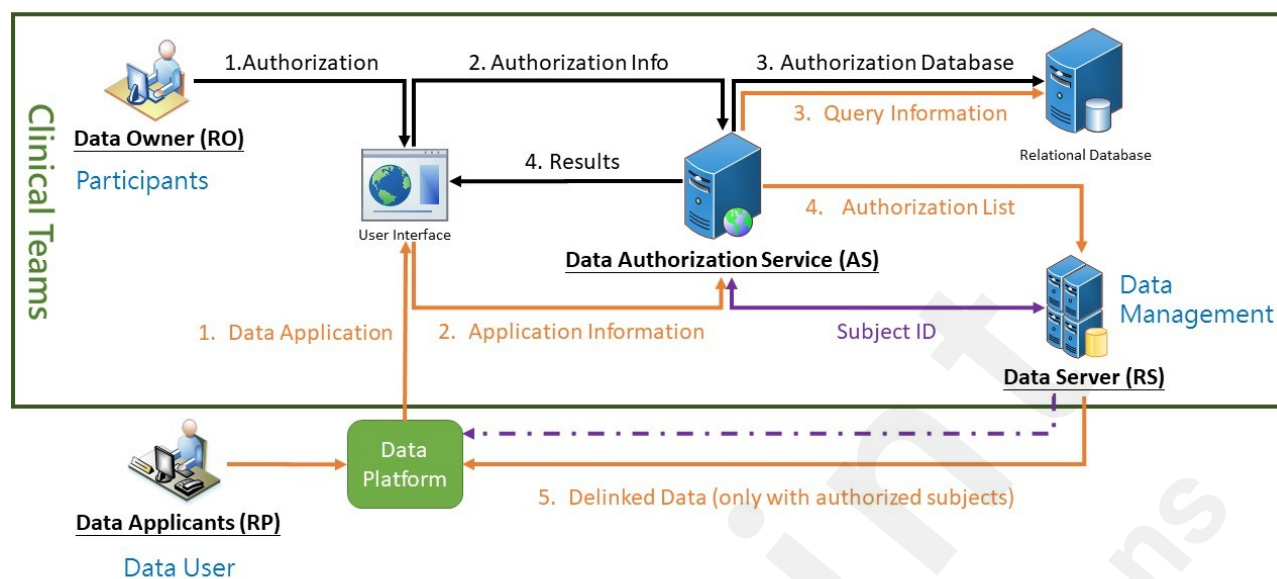
**Table 1** 18 categories of protected health information

1. Names
2. All geographic subdivisions smaller than a State
3. All elements of dates (except year) for dates directly related to an individual
4. Telephone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers
13. Device identifiers and serial numbers
14. Web Universal Resource Locators (URLs)
15. Internet Protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code

**Figure and Figure Legends**



**Fig. 1** Taiwan Medical AI and Data Portal (https://data.dmc.nycu.edu.tw/)

**Fig. 2** Health Data Authorization Service Platform. This diagram depicts the architecture of a medical data-sharing platform that integrates dynamic consent for secure and transparent data sharing. The system centers on four key roles: the Resource Owner (RO), who owns the data; the Resource Server (RS), where medical data is stored and managed; the Requesting Party (RP), typically researchers seeking data access; and the Authorization Server (AS), the core component facilitating connections among the roles. This framework ensures data access is based on explicit consent from the data owner, allowing real-time adjustments to consent settings for different data types and uses, such as academic research. The platform's primary tasks are to establish individual preferences, maintain consent history, and ensure trust and transparency between data owners and users, thereby advancing responsible data science development.

# Supplementary Files

# Figures

Taiwan medical AI and data portal.

Health Data Authorization Service Platform. This diagram depicts the architecture of a medical data-sharing platform that integrates dynamic consent for secure and transparent data sharing. The system centers on four key roles: the Resource Owner (RO), who owns the data; the Resource Server (RS), where medical data is stored and managed; the Requesting Party (RP), typically researchers seeking data access; and the Authorization Server (AS), the core component facilitating connections among the roles. This framework ensures data access is based on explicit consent from the data owner, allowing real-time adjustments to consent settings for different data types and uses, such as academic research. The platform's primary tasks are to establish individual preferences, maintain consent history, and ensure trust and transparency between data owners and users, thereby advancing responsible data science development.