MDDX
RESEARCH & INFORMATICS

WHITE PAPER

DATA PRIVACY AND DE-IDENTIFICATION

MDDX RESEARCH & INFORMATICS
580 California St, Floor 16
San Francisco, California 94104

T (800) 441-MDDX
F (866) 382-4696
info@mddx.com
www.mddx.com

A PROVEN STRATEGY TO ELIMINATE LONG TERM LIABILITY

It is commonly understood by most clinicians and researchers that the unauthorized disclosure of Protected Health Information (PHI) is a major violation of Health Authority regulations in most countries. However, in the race to diagnose and treat patients or collect data for clinical trials, the potential implications of disclosing PHI 5, 10 or 20 years after the initial data collection are rarely considered.

A standard tenet of Good Clinical Practice for clinical trials requires that all medical data be retained for the length of the trial, plus for a specified-period after the trial concludes. As most multicenter trials are now global in nature, the treatment of protected health information must satisfy the health authorities of each country where aspects of the trial are conducted.

A proven solution for minimizing the risks associated with the long-term storage of medical data is de-identification, the removal of all identifiers linking the data to an individual. However, to be done properly and completely, de-identification must be performed by exacting method. The safest way to store data for the long-term is in a de-identified state, but of equal importance is the ability to confirm that the de-identification has completely scrubbed all patient identifiers without distorting, corrupting or mixing up source data. This is a critical step in minimizing the long-term liability for any organization.

These preventive measures become more salient as regulatory efforts increase. In the United States, the Office of Civil Rights is ramping up their audit rate significantly, improper or incomplete de-identification is an easy violation for regulators to target, but it's also an addressable problem that can be prevented by employing proper policies and methods.

DEFINITIONS OF INFORMATION REQUIRING PROTECTION

IIHI: Individually Identifiable Health Information, held or maintained by a covered entity or its business associates acting for the covered entity, that is transmitted or maintained in any form or medium (including the individually identifiable health information of non-U.S. citizens). This includes identifiable demographic and other information relating to the past, present, or future physical or mental health condition of an individual, or the provision or payment of health care to an individual that is created or received by a health care provider, health plan, employer, or health care clearinghouse. For purposes of the Privacy Rule, genetic information is considered to be health information.

PHI: Protected health information (PHI) is all "individually identifiable health information" held or transmitted by a covered entity or its business associate, in any form or media, whether electronic, paper, or oral. The Privacy Rule calls this information "protected health information (PHI). HIPAA details 18 individual identifiers, included below.

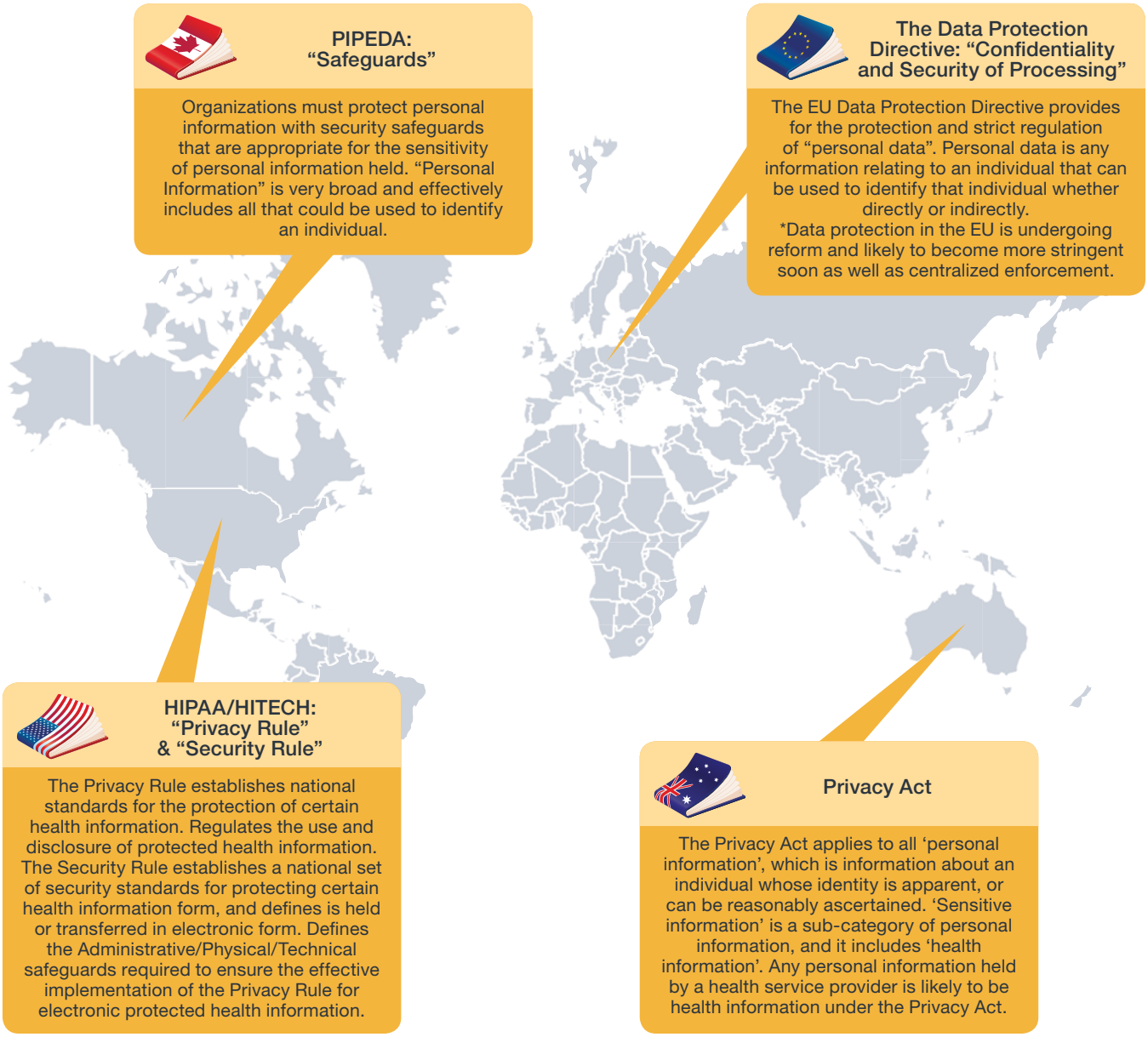
Personal Data: The EU Data Protection Directive defines personal data is any information relating to an individual that can be used to identify that individual whether directly or indirectly.

Personal Information: The Australian Privacy Act 1988 defines Personal Information as information about an individual whose identity is apparent, or can be reasonably ascertained using the information.

PRIVACY REGULATIONS OF IDENTIFIABLE MEDICAL DATA

With many of today's multicenter studies being conducted internationally, it is common for trial data to fall under the governance of multiple countries with data privacy regulations of varying restrictions. The international laws on data privacy come with many names and under the control of just as many regulatory bodies. A sampling of this complexity is provided below:

Figure 1: Medical Data Privacy Laws Around the World



Given the variations in regulations the most sensible data privacy policy requires a risk averse approach. This is readily achievable by assuring that your organization is in compliance with the most stringent privacy standards for regions in which you operate. This conservative approach not only ameliorates risk for surprise audits from Health Authorities, but also works to anticipate changing rules in International standards.

DEFINING WHAT INFORMATION HAS TO BE PROTECTED

The most restrictive regulations are broadly worded with the aim of protecting as much personal information as possible. Working from this protective approach the best answer to "what information needs to be protected?" adopts this standard:

WHAT INFORMATION NEEDS TO BE PROTECTED?

Any information contained on a record generated by (or referring to) any medical transaction which could be used to identify or lead to the identification of the individual connected with that record, their relatives, roommates, co-inhabitants, or employers.

INCREASING PENALTIES and INCREASING ENFORCEMENT

INDIVIDUAL HEALTH IDENTIFIERS[2]

Orange text is commonly found in medical images

- **Names**
- Initials
- **Location (expand w/tag link)**
- **Birth Date**
- Admission Date
- Discharge Date
- **Telephone numbers**
- Fax numbers
- E-mail addresses
- Social security numbers
- **Medical record numbers**
- **Health plan beneficiary numbers**
- **Account numbers**
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- **Device identifiers and serial numbers**
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images and any comparable images
- Government Issued ID Numbers
- **Any other unique identifying number, characteristic, or code**

With the rapid advance of technology has come an equal increase in security risk, the threat of breach is constant. While all stored data is at risk of breach, the failure to properly de-identify data leaves that data more vulnerable in case of a breach because the owner believes the data is safe. Improper or incomplete de-identification creates a substantial risk of hefty fines if a breach occurs. The current penalties in the U.S. are shown in **Table 1**. It is important to note that even the unintentional disclosure of a single medical image when exercising reasonable diligence can result in a fine of $50,000.

The denominator for any fine is the number of times the violation occurs. In other words, the fine amount can be multiplied "per violation". A singular data breach or unauthorized disclosure results in an assessed penalty for each individual medical record compromised. In the standard 5-year multicenter clinical trial, it is common to have 9 imaging endpoints per subject. If the trial enrolls 1000 patients, that represents 9000 opportunities for violations of privacy laws even when exercising reasonable diligence.

Table 1: Financial Penalties for Disclosure of PHI in US

CIVIL MONEY PENALTIES

Violation Category	Penalty Per Violation	Max Penalty for Identical Violations in a Calendar Year
Did not know and could not have known exercising reasonable diligence	$100 – $50,000	$1,500,000
Reasonable Cause	$1,000 – $50,000	$1,500,000
Willful Neglect – Corrected	$10,000 – $50,000	$1,500,000
Willful Neglect – Not Corrected	At least $50,000	$1,500,000

Reasonable diligence means the business care and prudence expected from a person seeking to satisfy a legal requirement under similar circumstances. Note that even when exercising reasonable diligence the minimum fine is $100 per violation. If 500 images containing "hidden" protected information are improperly accessed the fine will be at least $50,000. Even if the protected information was never viewed. Alternatively a breach of 500 properly de-identified images results in no fine.

Reasonable cause means an act or omission in which a covered entity or business associate knew, or by exercising reasonable diligence would have known, that the act or omission violated an administrative simplification provision, but in which the covered entity or business associate did not act with willful neglect. Under this definition which is more likely for improperly de-identified medical images the fine is a minimum of $500,000.

Willful neglect means conscious, intentional failure or reckless indifference to the obligation to comply with the administrative simplification provision violated. If willful neglect is found the minimum penalties cannot be waived by the OCR.

Penalties are not limited to civil money fines alone. In the U.S. knowingly disclosing IIHI can be prosecuted as a criminal act resulting in imprisonment as well as additional fines. Once again the failure to properly de-identify data leaves that data vulnerable and both organizations and individuals responsible.

Criminal Penalties. A person who knowingly obtains or discloses individually identifiable health information in violation of the Privacy Rule may face a criminal penalty of up to $50,000 and up to one-year imprisonment. The criminal penalties increase to $100,000 and up to five years imprisonment if the wrongful conduct involves false pretenses, and to $250,000 and up to 10 years imprisonment if the wrongful conduct involves the intent to sell, transfer, or use identifiable health information for commercial advantage, personal gain or malicious harm. The Department of Justice is responsible for criminal prosecutions under the Privacy Rule.[1]

ENFORCEMENT

Enforcement efforts are ramping up significantly in the US. After recently granting State attorney's general the authority to investigate and enforce violations under HITECH, the Office of Civil Rights has explicitly stated its intention to greatly increase audits:

> **"The OCR's last 12 months of enforcement activity will "…pale in comparison to the next 12 months.""**
>
> *Jerome B. Meites,*
> *Chief Regional Civil Rights Counsel Region V*
> *- Speaking to ABA in 2014*

By working with identified medical images, an organization is working their faith that everyone that has access to those images, including staff, consultants and their sub-contractors is going to handle the images within the constraints of the law at all times.

All it takes is one stolen laptop, disgruntled employee, hacked server or absent minded employee for an organizations medical files to be breached. De-identification can't stop the hack, but it prevents the violation before it starts.

Investigations conducted by the Office of Civil Rights typically uncover additional unknown violations resulting in greater fines and requiring immediate and costly remedial measures.

The most occurring additional findings are:

- Inadequate policies or procedures
- Inadequate staff training
- No individual identified as responsible security or privacy official
- No outside compliance auditor
- Insufficient risk analysis and risk management systems[3]

The status of regulatory audits has shifted from “if” to “when” in clinical trials. It is now imperative for organizations to anticipate an audit and begin implementing preventive measures well in advance. However, no data privacy regulations or health authorities claim to have jurisdiction or authority over privacy regulations for properly de-identified medical images, making them the safest alternative for long-term storage. In the US, if your medical image storage system only contains de-identified images then there are no reporting requirements, no regulatory investigations, no additional findings, and no fines. Furthermore, if you are storing your data in duplicate or triplicate you would simply deactivate the compromised storage device and use your other source. Properly de-identified and encoded data is generally useless to thieves, hackers, even your competitors, and it is outside the regulatory compliance scope.

Quote from the Department of Health and Human Services:

> **“There are no restrictions on the use or disclosure of de-identified health information… De-identified health information neither identifies nor provides a reasonable basis to identify an individual… the removal of specified identifiers of the individual and of the individual’s relatives, household members, and employers is required.”**

BUSINESS ASSOCIATE AGREEMENTS

Under the HIPAA/HITECH Omnibus rule BAA's are now mandatory for many business relationships in medicine. These new regulations expanded the requirement of who is required to have them and effectively eliminated the boilerplate agreements that were commonplace in the past. To meet today's regulatory standards BAA's must be customized to reflect the specific nature of the work and relationship of the parties.

If an organization meets this relationship tests having a BAA is required:

Any organization that creates, receives, maintains, or transmits protected health information for an organization that is regulated by HIPAA is considered a Business Associate.

Any organization that creates, receives, maintains, or transmits protected health information for a Business Associate is considered a Subcontractor and equally subject to the Business Associate regulations.

The relationship test has been met

Business Associates and Subcontractors are now directly liable under HIPAA regardless of whether a BAA exists. Simply put, if the relationship test described above has been met then the following applies:

1. HIPAA requires a current, executed BAA reflecting the relationship and responsibilities of both parties.
2. The absence of #1 results in a violation to both parties.
3. This extends to subcontractors and to their subcontractors.

For example: If your organization contracts with Company X to destroy old hard drives containing ePHI you must have a BAA with them. If you do not have a BAA you are both in direct violation of HIPAA. If you do have a BAA with them it might state:

"Company X is aware that all hard drives contain ePHI and will destroy all hard drives in a manner consistent with HIPAA and deploy security measures to ensure the hard drives remain secure until they are destroyed."

If Company X fails in their agreed upon duties to protect PHI and a hard drive goes missing, you might presume you are covered by the BAA — but you would be incorrect.

The only way to protect your organization is to have a customized BAA specifying each parties, roles and responsibilities but also defining oversight and creating audit trails. The BAA should be followed by inter-vendor audits to ensure that each party is meeting its regulatory obligations to uphold privacy regulations. If you can't monitor what your vendors are doing then you don't know they are doing it.

DE-IDENTIFIED DATA IS SAFE DATA

Performing the complete and accurate de-identification of any standard medical data can be challenging. When dealing with medical image data (i.e. DICOM datasets), it becomes substantially more complex. The majority of images that users believe they have correctly de-identified have actually been done improperly, leaving patient identifiers hidden below the surface. These “hidden identifiers” pose a substantial problem, hidden does not equal de-identified. What started as a common data privacy problem (improper de-identification) is greatly exacerbated by ignoring the problem and this distinction can increase a six-figure fine by ten-fold.

A medical image containing even a single patient identifier is instantly subjected to privacy regulations worldwide (such as HIPAA/HITECH/PIPEDA). Using, disclosing, or storing improperly de-identified DICOM like it was properly de-identified is a ticking time bomb. While the regulations worldwide vary in enforcement and requirements, they all require the full protection and limited use and access of identifiable medical data forever. However, the singular act of de-identifying image data using a validated process or provider eliminates all health authority oversight of that data, no restrictions, no oversight authority, no fines, no risk.

CONCLUSION

In summary, medical images used in clinical trials can be used to derive powerful endpoints, but they have many potential pitfalls. When storing the image data for your required retention period, inaccurate de-identification or faulty methodologies can result in image corruption or incorrect labeling of data sets (i.e. mixing up subjects). Incomplete de-identification can result in PHI violations if a breach occurs. While fines and penalties vary around the world, the safest approach to ensure you are compliant is to only store de-identified images that were created using a validated methodology.

SELF-ASSESSMENT

Use the self-assessment below to review the critical steps in closing gaps and achieving total compliance for your organization.

ASSESS YOUR ORGANIZATION'S DE-IDENTIFICATION READINESS:

1. At any point in the process is your team exposed to Protect Health Information?
2. Are you certain that all of your trial data is completely de-identified for all the potential identifiers in the Individual Health Identifier table?
3. Do you have a method of disposing of PHI after de-identification?
4. Are you fully confident that your staff or consultants handle PHI in a safe and compliant manner that doesn't create liability for your organization?
5. Have your internal systems been audited for security risks, including all personal computers that might be exposed to PHI?

CORRECTIVE STEPS TO CONSIDER

CRITICAL STEPS TO RESOLVING THESE ISSUES:

1. Develop and document thorough policies and procedures containing inherent privacy, security, and monitoring features.
2. Develop and deploy sufficient staff training to ensure consistent implementation of you policies and procedures.
3. Establish periodic monitoring and audit trails for critical systems and name an individual staff member or consultant responsible for continuous monitoring.
4. Evaluate all business relationships to determine if a BAA is required and then make sure your BAA's are customized, clear, and establish oversight.
5. Use a qualified vendor for de-identification of your trial's medical images, including cleaning up any legacy data.

[1] http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/
http://www.hhs.gov/ocr/privacy/hipaa/understanding/srsummary.html
http://ec.europa.eu/health/data_collection/data_protection/in_eu/index_en.htm
http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm
https://www.priv.gc.ca/leg_c/p_principle_e.asp
http://www.oaic.gov.au/privacy/privacy-act/australian-privacy-principles
http://www.oaic.gov.au/privacy/privacy-act/privacy-law-reform
http://www.oaic.gov.au/privacy/privacy-resources/privacy-fact-sheets/health-and-ehealth/

[2] http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/

[3] http://www.hhs.gov/ocr/privacy/hipaa/enforcement/examples/index.html