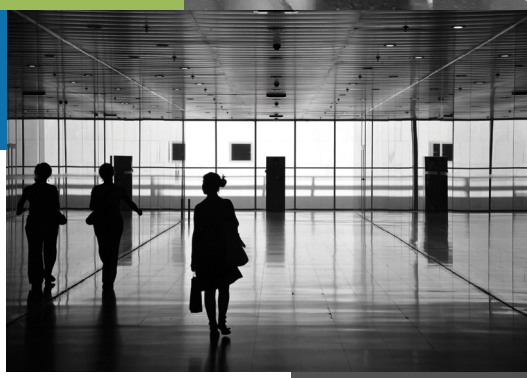
PRIVACY ANALYTICS

WHITE PAPER

What's the Risk - Sharing Data for Secondary Use

Privacy Officers in healthcare settings have a critical role to play. They need to guide their organizations in monitoring emerging risks and establishing responsible data sharing practices. Privacy Officers can help their organizations effectively limit risk when releasing data by: locating sources of PHI in their data; using effective de-identification; benchmarking against best practices; assessing security incidents; basing data sharing agreements on context; and, engaging with experts.



PRIVACY ANALYTICS

a QuintilesIMS company

Data, Health and the Risk of Non-Compliance

The technology, finance and retail industries have long relied on big data and analytics to generate revenues and lower the cost of doing business. The healthcare industry, however, has been slower to capitalize on the wealth of information collected by hospitals, health insurers and drug and medical device companies. These organizations are recognizing the potential of unlocking their data for secondary purposes. More and more healthcare organizations are now using de-identified health data to support medical research, perform post-market drug surveillance, monitor the quality of care, identify and treat diseases earlier, and deliver better clinical outcomes.

When it comes to using data for secondary purposes, privacy implications, legal implications and public relations' ramifications are all major concerns for providers, payers and the pharma industry[1]. Privacy Officers know that leveraging protected health information (PHI) or personally identifiable information (PII) requires them to tread carefully. Safeguarding patient privacy is of paramount importance and the repercussions for a breach can be costly, both in dollars and reputation.

The adoption of electronic medical records (EMRs) and increased use of medical monitoring devices, patient imaging, and mobile technologies means that the amount of patient data is growing exponentially. Consequently, the demand to share this data with other groups is growing along with projections that the healthcare analytics market will reach between $18 billion and $21 billion[2] by 2020, up from $5.8 billion in 2015[3].

The regulatory environment for health information is a complex one. Long-standing legislation like HIPAA has been modified and updated by the HITECH Act and other changes, like the proposed 21st Century Cures Act, are waiting in the wings. When you add the various national and international standards and guidelines (like HITRUST, PhUSE and the EU General Data Protection Regulation, it can become challenging to determine if your data sharing practices meet regulatory compliance.

These recent changes have made the role of Privacy Officers in healthcare organizations more important than ever before. In addition to being privacy champions, these individuals must now help their organizations navigate the regulatory landscape and manage risk to minimize financial and reputational costs.

This paper offers up six ways that Privacy Officers can limit risk when releasing data for secondary uses. By doing so, Privacy Officers can be confident that their organization's data sharing practices effectively protect privacy, comply with current legislation, and are defensible should a breach occur.

Value and Vulnerability in Sharing Data for Secondary Purposes

Government investments made through the Health Information Technology for Economic and Clinical Health Act (HITECH Act) fostered widespread adoption of EMRs and healthcare IT over the past five years[4]. The result has been a bounty of health information that can be analyzed for a broad range of purposes – from clinical

studies validation to quality monitoring to cost containment. The ability to integrate and study data from disparate systems holds the potential to speed delivery of new drugs to the market, uncover previously undetected causes of illness and find innovative therapies and treatments for disease. As a result, many healthcare organizations are recognizing that their data holdings are strategic assets and are seeking ways to monetize them.

New businesses are also being created that leverage big data in healthcare to help patients. A 2013 review by McKinsey & Company showed more than 200 companies created since 2010 offer tools that make use of available healthcare data[5].

One such organization is Flatiron Health, a New York-based company that enables researchers to find useful patterns in medical records to help in the treatment of cancer. Flatiron's cloud-based platform aggregates oncology data from millions of cancer patients. This provides clinicians and researchers with a vast and rich data source that incorporates data from a much broader group than the three to four percent of patients who participate in clinical trials. But making this information available requires rigorous privacy practices to remove PHI so that the data is HIPAA-compliant.

To be effective, companies like Flatiron rely on information that retains a high level of quality and granularity; demands that push the boundaries of standard forms of data protection like masking and Safe Harbor. When data contains personal information that can be used to re-identify an individual, it opens up an organization to the risk of a re-identification attack – a breach of patient privacy that can be devastating to the individual and pose serious consequences for the organization.

In addition to making a multi-billion dollar government investment in health IT, the HITECH Act also introduced extended protections for patients by amending the existing HIPAA legislation. The HITECH Act made reporting of data breaches to the Department of Health and Human Services (HHS) affecting more than 500 people mandatory and requires notification go to those affected. Furthermore, it increased the fines that could be levied against organizations that fail to comply, with a maximum penalty of up to $1.5 million per violation[6]. These additional reporting requirements have driven up the average cost of a healthcare data breach to more than $2.1 million[7].

> Unfortunately, data masking does little to defend against re-identification attacks.

From 2009 to 2015, 1286 data breach incidents have been reported to HHS. While theft (47%) and hacking (11%) are reasons often cited for a breach, breaches may also be the result of loss (10%) from employee negligence or unauthorized access (19%)[8]. These inadvertent causes of data breach, driven by internal processes, also pose a major risk to patient privacy. Even if steps have been taken to anonymize the data, it may be subjected to a re-identification attack if it falls into the wrong hands[9].

Unfortunately, data masking does little to defend against re-identification attacks, making real the risk of a privacy breach from data loss or unauthorized access. To minimize the likelihood of a successful re-identification, organizations need to look at using a risk-based approach to data de-identification as recommended by HITRUST, PhUSE, the Institute of Medicine and other respected groups.

In this environment of more demand, persistent threats, heavier requirements and bigger fines, Privacy Officers have a crucial role to play in educating their organizations about the risks associated with PHI, secondary use of data, and how to implement best practices to protect privacy.

Six Ways to Effectively Limit Risk When Releasing Data

Privacy Officers know that sharing data for secondary use is inherently an exercise in risk management. By effectively assessing the data's exposure to risk, proper measures can be taken to safeguard individual privacy. It's about achieving the right mix. Maximum security could be achieved by simply not sharing the data with anyone; however, this would defeat the loftier aims of secondary use. Maximum data quality can be met by leaving all of the data elements intact but, of course, this leaves potentially sensitive health information accessible to individuals who could, in turn, leak that information further.

Effectively safeguarding patient privacy while maintaining the quality of the data is achieved through two broad measures: securing the data from unauthorized access and applying a risk-based approach to data de-identification to limit the exposure of PHI. The following list provides approaches to address both of these areas and can help organizations establish practices that meet regulatory compliance.

1. Locate Sources of PHI in Your Data

In the U.S., the disclosure of PHI is governed under HIPAA legislation. To provide data for secondary use, HIPAA requires that PHI be removed from the data. Covered Entities and Business Associates have traditionally employed two approaches to achieve this – data masking and the Safe Harbor method set out under HIPAA's Privacy Rule.

As healthcare has evolved from paper charts to digital files to data analytics, the extent and complexity of the health data that is captured on a patient has increased. While data masking and Safe Harbor offer straightforward and prescriptive approaches to the removal of patient identifiers, their focus leans heavily on structured data elements, leaving any PHI that resides in free-form text untouched.

Organizations are finding PHI in their data where they didn't expect it to be, particularly in unstructured data. As a result, these organizations may be inadvertently "leaking" PHI to data recipients without being aware of it. Gaining greater visibility into your data by locating hidden sources of PHI can help to avoid a costly mistake.

Adequately assessing privacy risks requires the use of tools that can comprehensively examine structured and unstructured data elements to dig out sources of PHI. Only then can effective data de-identification be performed.

2. Use Effective De-Identification to Limit Exposure

While data breaches resulting from external attacks and hackers garner significant media attention, breaches due to re-identification attacks, although often less publicized, also present real risks to patient privacy.

The removal of basic demographic data from a database does not ensure anonymity for the individuals represented within it. Healthcare organizations are starting to share data widely with researchers and, in some cases, make information publicly available. Unfortunately, their de-identification practices may not have caught up with their sharing practices. Successful re-identification has been demonstrated by matching data that has been masked to public sources of information, like newspaper reports[10]. Meaningful de-identification requires more than the simple removal of names and addresses; it involves taking a risk-based approach to data de-identification.

In recent years, many groups and organizations have advocated for the use of a risk-based approach to de-identification, including the Institute of Medicine, the Health Information Trust Alliance (HITRUST), PhUSE, TransCelerate BioPharma, and the Council of Canadian Academies. A number of these groups have developed frameworks for the responsible sharing of clinical data.

Implementing risk-based de-identification can be facilitated by consultants with expertise in the field. Ongoing practices can be supported through the use of automated software tools. These tools run typical threat scenarios against your data, letting you proactively assess its vulnerability to various threats. Results indicate how easily re-identification can be achieved using different methods should that data fall into the wrong hands. Effective de-identification can then be applied to the dataset to limit the likelihood of successful re-identification. This limits the data's usefulness to criminals, making it a less tempting target and containing the negative consequences of a breach.

3. Benchmark Risk Exposure Against Best Practices

One of the difficult decisions to be made when sharing data is how much de-identification to apply; finding the right balance between optimal data security and data quality is imperative. Too much de-identification unnecessarily reduces the data's usefulness. Not enough de-identification potentially exposes the organization – and the subjects of its data – to a breach of privacy.

Determining the "right" amount of de-identification to meet both of these needs can be difficult since we are dealing with probabilities, not absolutes. It is impossible to say that there is a zero chance that a patient could be re-identified from their data. However, it is possible to confidently know that the risk is close to zero. This is achieved by following the lead of reputable organizations that have extensive experience in releasing data. The Centers for Disease Control, for example, has established precedents for what is deemed to be an acceptable risk of re-identification for public release. These precedents provide for good data quality while also ensuring strong security.

By benchmarking your data against precedents established by respected entities, Privacy Officers can be confident that, in the event of a privacy breach, they have applied a defensible process to de-identifying their data assets.

4. Work with IT to Assess Security Incidents

While negligent or careless employees are still a major cause of data breaches, losing or leaving behind devices storing PHI or PII, criminal attacks on healthcare organizations have grown by 125% since 2010 and, for the first time, have become the number one cause of data breaches[11].

Despite this, many Covered Entities and Business Associates say they lack sufficient resources and budget to protect against new threats to patient data. Limited bandwidth leaves organizations with only enough capacity to react to incidents once they occur, focusing on the last threat rather than preparing for the next one. Findings from a recent Ponemon study show that the assessment of security incidents is most often an ad hoc process[12]. For both types of organizations, the majority indicate that they do not perform risk assessments for all security incidents involving electronic documents even though there is a federal mandate to do so[13].

While most organizations have an incident response process in place, many say they lack the funding and resources needed to make it effective. Organizations will come under greater pressure to allocate sufficient budget towards technologies and training for security as criminal attacks on health care data become more prevalent.

5. Base Data Sharing Agreements on Context

The sensitive nature of health data means that exceptional care is needed when data is to be shared for secondary purposes. But who will have access to the data, how it will be used and where it will be stored differ from one case to the next. Some sharing agreements require widespread access to the data while others restrict access to a trusted few. To quantify your organization's vulnerability in these situations, you need to look at both the data and the data recipient.

Data sharing agreements are contracts with the data recipient that help to clearly set out from the beginning of a data sharing arrangement the limitations on the use and disclosure of the data. Assessing the context in which data will be used allows the highest quality data to be provided while ensuring the lowest privacy risk.

Established frameworks exist that allow you to objectively assess the context in which the data will be shared. These frameworks look at factors such as the motives to re-identify the data and the security controls in place at the recipient site. A scoring mechanism is used to quantify the level of risk. This lets the organization determine whether further de-identification of the data is necessary for the given context.

6. Engage with Experts

In an attempt to protect privacy, many organizations employ data masking tools or other in-house de-identification solutions to remove PHI. Unfortunately these rudimentary tools, while being cost-effective and simple to use, have significant drawbacks. Many of the commonly-used masking techniques reduce the usefulness of data, diminishing granularity and destroying the utility of the masked fields. Furthermore, masking doesn't provide guarantees for low privacy risk. These techniques do not use metrics to measure the risk of re-identification so it is not always possible to know whether the data transformations performed were sufficient and defensible in the event of a lawsuit.

CONTACT US

251 Laurier Ave W
Suite 200
Ottawa, Ontario, Canada
K1P 5J6

Phone: 613.369.4313

www.privacy-analytics.com

sales@privacy-analytics.com

Operating in the current legislative environment requires Privacy Officers to have confidence that their data is compliant. Engaging with experts in health data de-identification will help organizations effectively protect patient privacy and enable them to unlock the value of their data. In addition to providing guidance on how to address regulatory issues to achieve compliance, experts can also provide a thorough assessment of your organization's data, train staff in risk assessment and de-identification techniques, and help in establishing best practices in incident response.

Conclusion

Healthcare organizations are under increasing demand to share their data – both from internal executives looking to achieve cost efficiencies and find new sources of revenues as well as external groups who grasp the potential to conduct innovative research and improve health outcomes. Privacy Officers are caught between the need to protect patient privacy and the pressure to permit new uses of the data.

For healthcare organizations, like many companies, data is an invaluable asset that can unlock a multitude of opportunities when used responsibly. Enabling data analytics provides hope for addressing some of healthcare's greatest challenges. However, a privacy breach can have devastating consequences, compromising an individual's privacy and eroding public trust in the organization.

By applying the six strategies above, Privacy Officers can learn how to mitigate risk and prevent a costly mistake. By taking steps to locate PHI and de-identify data using a risk-based approach that is consistent with best practices, you will be able to confidently provide privacy guarantees while still allowing high-quality data to be released by your organization.

CONTACT US

251 Laurier Ave W
Suite 200
Ottawa, Ontario, Canada
K1P 5J6

Phone: 613.369.4313

www.privacy-analytics.com

sales@privacy-analytics.com

Sources:

1. PriceWaterhouseCoopers LC (2009). Transforming healthcare through secondary use of health data.

2. Healthcare Analytics Market worth $21.3 Billion by 2020. (2013, Dec. 13). PR Newswire.

3. Healthcare Analytics Market worth $18.7 Billion by 2020. MarketsandMarkets.

4. Hanrahan, Donna. (2013, Nov. 15). Data Mining, Meaningful Use, Secondary Use, & Potential Misuse of Electronic Health Records. Institute for Ethics & Emerging Technologies.

5. Foley & Lardner LLP (2015, March 2). Tapping Into the Big Value of Health Care Big Data: Top Legal and Regulatory Considerations on the Path to Monetization. Lexology.

6. U.S Department of Health and Human Services. (2013, January 17). New rule protects patient privacy, secures health information.

7. Ponemon Institute (May 2015). Fifth Annual Benchmark Study on Privacy & Security of Healthcare Data.

8. US Department of Health & Human Services, Office of Civil Rights. Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information. Breaches Affecting 500 or More Individuals. (Retrieved on Aug. 17, 2015).

9. For an explanation of re-identification attacks, see Privacy Analytics' white paper, De-Identification 301: Three Adversaries Who Could Attack Your Data.

10. Robertson, Jordan. (2013, June 5). States' Hospital Data for Sale Puts Privacy in Jeopardy. BloombergBusiness.

11. Ponemon Institute, *Ibid*.

12. *Ibid*.

13. *Ibid.*