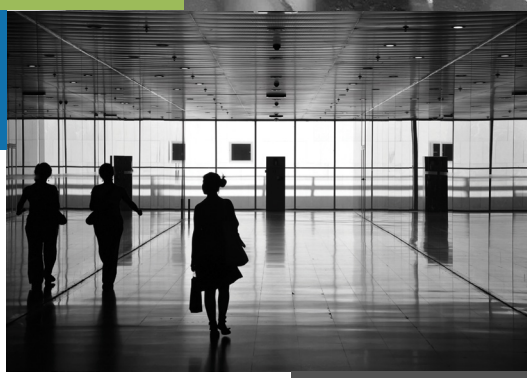
PRIVACY ANALYTICS

WHITE PAPER

A Privacy Governance Framework to Support De-identification

Organizations should establish a framework to manage re-identification risks holistically while enabling a wide range of data uses. This paper provides an outline of a governance framework specifically supporting the implementation of de-identification within an enterprise. Key to this is understanding and managing the processes, the people, and the technology required for data governance strategy.



PRIVACY ANALYTICS

a QuintilesIMS company

Many innovations in treatment methods, drugs, and devices are driven by access to detailed patient data. However, there are regulatory restrictions to gaining access to such data, as well as the practical need for a social license by patients to use that data for secondary purposes. De-identification addresses both of these issues in that regulations allow the use and disclosure of health data for secondary purposes without having to obtain patient consent. Surveys show that patients agree to their data being shared for these purposes if it is adequately de-identified.

De-identification is a process used to remove any identifying attributes from data, reducing the risk an individual can be re-identified from the data to an acceptably small level. Being recognized and prescribed across jurisdictions, de-identification should be risk-based to account for the specific disclosure context. Risk-based anonymization methods are consistent with recommendations from the Information Commissioner's Office in the UK[1], anonymization guidance from European Medicines Agency[2], the privacy commissioner of Ontario[3], the Expert Determination methods under the HIPAA Privacy Rule in the US[4], and other governmental, academic, and professional associations and groups.[5]

As an example, the risk of re-identification is higher when a dataset is released publicly than when it is released to a trusted recipient in a controlled environment under specific contractual terms. Consistently measuring the risk of re-identification to determine how the data should be de-identified is important. This risk measurement should consider the likelihood of a deliberate re-identification attempt, inadvertent re-identification (such as by an acquaintance), and a breach. In lower-risk contexts, less de-identification is required and higher data utility can be maintained, whereas higher-risk contexts require more de-identification to be applied to the data.

De-identification techniques include generalization, suppression, date shifting, and adding noise, in addition to masking and pseudonymizing. Sophisticated tools and technologies are available to support de-identification as well as to automate and integrate risk measurement. Integrating risk measurement into a technological solution enables data utility to be optimized for the disclosure context.

While technology is an important enabler of data de-identification, technology is not the end of the story. Effective de-identification at an enterprise level is as much about governance as it is about technology. Accounting for risk in a de-identification solution is critical to achieving the right level of de-identification and resulting data utility, which influences the analytic outcomes. Accordingly, to maximize outcomes, an organization must have efficient methods of measuring, monitoring and assuring the controls associated with each disclosure context. More broadly, organizations should establish a framework to manage re-identification risks holistically while enabling a wide range of data uses.

If you only apply technology to anonymize data, you miss out on a vital area of the overall strategy – the people and decisions behind the solution and the processes and procedures that instill consistency. Without these elements, you miss the tenets of governance – accountability, transparency, and applicability.

The Three Pillars of an Effective De-identification Governance Framework

De-identification strategies need to promote data use and data utility while protecting privacy and complying with requirements. Disclosure context should influence the technological methods used to de-identify. Only considering the data when prescribing de-identification specifications is insufficient, and will not address growing demands for data in a privacy-preserving manner.

When first planning a strategy for de-identification of health data, there is need for a framework comprised of three pillars: People, Process, and Technology. You can think of the Technology as being the engine of de-identification, with the Process being the oil, and the People being the drivers of that engine. This is illustrated in Figure 1.

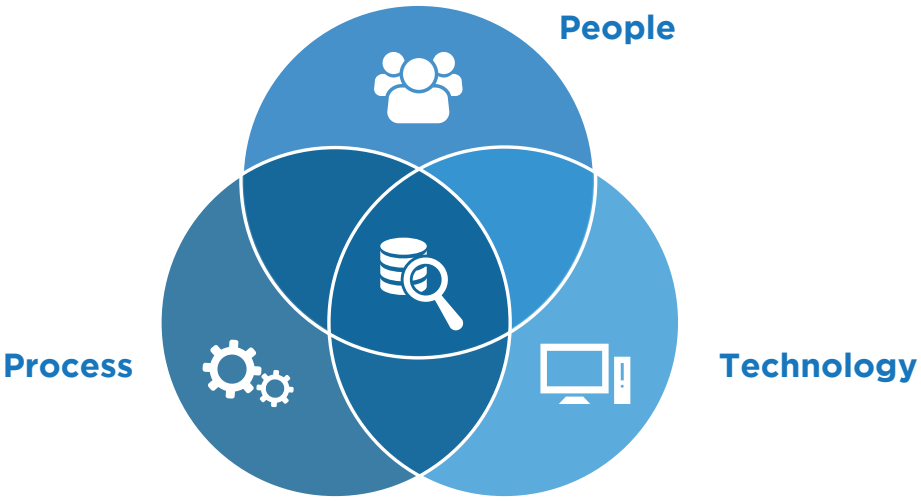
People

The people behind the framework are part of a network of decision makers, knowledge workers, and implementers that bring the framework to fruition and apply its tenets. The People Pillar encompasses:

1. **Understanding the need.** Which method of de-identification is needed, for which data set? This is the fundamental premise from which to work. An organization needs to understand the risks associated with re-identification, and how to minimize those risks. Different strategies can be applied to measure and manage risk effectively, such as grouping data consumers based on risk characteristics. Understanding the need will involve defining use cases and developing effective risk-based de-identification strategies to address these use cases.
2. **Approval mechanisms.** Knowing when de-identification is required is a key starting point. Someone, somewhere, has to understand the reasoning of 'acceptable uses of patient data'. While effective de-identification protects privacy, organizations should also consider ethical and reputational factors. For example, disclosure of attributes relating to groups of individuals that may be stigmatizing to those individuals (e.g. a data set showing that children of parents with a particular religious affiliation are not being vaccinated against HPV could result in stigmatization).[6] Policy and associated governance mechanisms should be in place to define and approve acceptable uses of data from a privacy, ethical, and reputational standpoint.
3. **Capacity to deliver.** Understanding re-identification risks and de-identification techniques is important to support an enterprise solution. De-identification requires cross-functional involvement and expertise. Data consumers, data analysts, privacy, security, legal, and IT personnel all play an important role in an effective solution. The organization needs a common vocabulary, methodology and set of operating guidelines to properly manage the risks of re-identification while making data available efficiently, consistently, and with the level of utility needed. A multi-tiered training program that accounts for all stakeholders can be important to establishing organizational capacity. On one end of the spectrum, broad

PILLARS OF DATA GOVERNANCE FRAMEWORK

Accessing and disclosing health data for secondary use depends on an effective de-identification governance framework that protects patient privacy. A holistic enterprise approach spanning people, process, technology is needed.



Process

- Standard Operating Procedures
- Data sharing templates
- Centralized disclosure reporting and tracking
- Risk assessment templates, recipient audit programs
- Workflow optimization, including process automation

People

- Understanding of data use needs and development of de-identification strategies
- Communication internally and externally
- Governance structures, escalation paths
- Multi-disciplinary approval mechanisms
- Capacity and training at various levels
- Oversight and monitoring

Technology

- Automation for risk determination, pseudonymization, and full de-identification
- Information security controls, embedded throughout
- Logging and auditing

Figure 1: The Three Pillars of a De-identification Governance Framework

training of core concepts, terminology, and policy statements is required. On the other end, focused, hands-on technical training with a range of de-identification use cases and scenarios is required. Establishing a standardized, sustainable training and knowledge management program underpins the capacity needed to de-identify data for a range of different data uses.

4. **Communication.** Related to organizational capacity is effective communication, internally and externally. Communication supports transparency, and can improve the data consumer experience while promoting patient trust. Patients, analysts and regulators need to understand how privacy is being protected across a spectrum of circumstances. Empowering the consumers of data with de-identification knowledge supports a better consumer experience as well as a strong privacy program.

5. **Oversight.** Successful de-identification delivery and execution require feedback. Your organization needs to have appropriate methods to monitor, evaluate, and audit the implementation and use of de-identification. Oversight should include mechanisms to audit recipients of de-identified data against data sharing terms and conditions.

6. **Governance models and escalation paths.** Good governance provides the tools and structures to make the right decisions. Some of these decisions are driven by regulation. Requirements can include data breach response and reporting responsibilities, which may involve notification and communication with patients.[7] Generally, if the data has been properly de-identified, it is not considered a privacy breach, whereas if the data is personal health information (PHI), the breach may have to be reported. Organizations need to be ready to respond, both with standard operating procedures and with defined escalation paths.

7. **Regulatory monitoring and change management.** Keeping ahead of the game where regulation is concerned is a must to prevent the last-minute dash to retain compliance. Managing change includes monitoring, anticipating and responding to regulatory changes, with a view of court decisions and precedents influencing risk thresholds, as well as technical and security changes (e.g. encryption protocols), and organizational changes. Effective de-identification is a function of data and context, so change management should also consider how environmental changes influence the risk of re-identification. Underpinning change management with standard operating procedures is important to maintaining control, consistency and compliance.

Process

Processes are the pivot upon which a successful framework turns. Processes are driven by people. Processes need to be defined, documented, communicated and widely understood. Effectively designed processes can optimize your de-identification program in practice: maintaining freedom of data use, while ensuring privacy. The Process Pillar encompasses:

1. **Standard Operating Procedures (SOPs).** SOPs are already an intrinsic part of

healthcare. They ensure that specific functions perform correctly. Extending SOPs to the area of de-identification – encapsulating both anonymization and pseudonymization – builds control into the de-identification governance framework. SOPs should encompass, at a minimum: re-identification risk measurement, pseudonymization, key management, risk-based de-identification (including classification of data elements as directly or indirectly identifying), and de-identified data use/access (including registration and authentication). Business guidelines should exist to define how organizations handle requests for data; training; logging, reporting and tracking of all data disclosures (including de-identification specifications); performance measurement; data use agreements, monitoring and audits; and exception handling.

2. **Data sharing agreement templates.** Effective de-identification involves understanding and managing data sharing and usage agreements. The agreement templates need to include terms of use, required security controls, and contractual mechanisms (such as rights to audit and data retraction). These agreements need to be actively managed to maintain compliance and support high-utility data disclosures to secure, well-controlled recipients. Automating the creation of agreement attachments, based on dataset specifics and risk-based transformations applied, is a best practice to promote consistency, efficiency, and compliance.
3. **Risk assessment and audit programs.** Having criteria and checklists to assess and audit against are simple methods to keep the complexities of risk assessment under control.
4. **Centralized disclosure logging.** Using a centralized logging system for tracking your disclosures enables compliance to be demonstrated, audits to be performed, and performance to be measured over time. Each disclosure and risk measurement should be reported on, logged and centrally tracked with details on the dataset as well as the context.
5. **Protocol tracking.** Keeping track of your user and recipient registration protocols will allow you to map these to your overall risk assessment integration.
6. **Workflow optimization.** Optimization of workflow is a key requirement of an efficient system. For example, integrating risk assessment into a case management solution that supports the customer experience can promote efficiency, performance, and effective control. Workflows should be optimized to support intended data outcomes. Embedding tools, controls and measures into de-identification workflows can drive efficiencies in support of these data outcomes.

Technology

The people and processes of a de-identification governance framework are supported and enabled by the technology used. Choosing a technology that has been designed to work as part of a wider framework is the key to finalizing a successful de-identification program. The Technology Pillar encompasses:

1. **De-identification software.** For enterprise solutions with large volumes of health data, automation through de-identification software is essential in practice. De-identification software needs to be part of this de-identification governance framework, but not as a distinct component. De-identification software sits within the framework as part of the overall strategy, and enables pseudonymization as well as a range of data perturbations that can be applied on a risk basis: generalization, suppression at multiple levels (e.g. row-based suppression or variable-based suppression), date shifting, additive noise, and various types of masking.
2. **Information security controls.** Essential information security measures are an integral part of the technology stack used to implement de-identification. These measures should include, as a minimum, robust authentication, key management, encryption, and identity assurance, where appropriate.
3. **Auditing events.** The technology must have extensive audit and logging built in. In addition to supporting compliance reporting, auditing is your system feedback and allows you to optimize a system once it is in production.

Applying the three pillars of de-identification governance provides the tools to create a highly successful strategy for using health data in a privacy enhanced manner.

Building a Framework from the Pillars

An effective de-identification governance framework is vital when considering the risks of re-identification as data volumes continue to grow exponentially. Implementing a successful de-identification program in practice requires a framework with all three pillars.

Core to a de-identification governance framework is a thorough understanding of the risks of re-identification. Re-identification is the ultimate failure point. There must be a deep understanding of the risks associated with each level of de-identification, from pseudonymization to full anonymization. Monitoring and oversight, as well as decision making and change management, are critical to maintaining enterprise standards, consistency, and performance.

Using the People, Process, Technology pillars as the foundation for your de-identification program allows a complete and holistic approach to managing the risks of re-identification while efficiently driving important outcomes from secondary uses of health data.

Conclusion

Having access to health data for the progress of health research is essential. Patient lives can be enhanced by the legitimate use of health data, but they can be damaged by its illegitimate use as well as by accidental breaches.

Technology alone cannot crack the complex problem of keeping data private yet accessible and rich. De-identification governance is increasingly important as we face a wide range of contexts, risks, and demands for data. With a multi-faceted stakeholder landscape and heterogeneous data, a governance framework must enable flexibility, scalability, and compliance across jurisdictions.

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. Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice' (Information Commissioner's Office 2012).
2. European Medicines Agency, 'External Guidance on the Implementation of the European Medicines Agency Policy on the Publication of Clinical Data for Medicinal Products for Human Use,' http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2017/04/WC500225880.pdf.
3. Ann Cavoukian and Daniel Castro, 'Big Data and Innovation, Setting the Record Straight: De-Identification Does Work' (Information & Privacy Commissioner of Ontario 2014); A Cavoukian and K El Emam, 'Dispelling the Myths Surrounding De-Identification: Anonymization Remains a Strong Tool for Protecting Privacy' (2011) 8 Canadian Law Review 89; A Cavoukian and K El Emam, 'A Positive-Sum Paradigm in Action in the Health Sector' (Office of the Information and Privacy Commissioner of Ontario 2010); A. Cavoukian and K. 'El Emam', 'De-Identification Protocols: Essential for Protecting Privacy' (Office of the Information and Privacy Commissioner of Ontario 2014); Information and Privacy Commissioner of Ontario, 'De-Identification Guidelines for Structured Data' (2016).
4. Office for Civil Rights, 'Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule' (Department of Health and Human Services 2012).
5. Institute of Medicine, 'Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk' (2015); The Expert Panel on Timely Access to Health and Social Data for Health Research and Health System Innovation, 'Accessing Health And Health-Related Data in Canada' (Council of Canadian Academies 2015); PhUSE De-Identification Working Group, 'De-Identification Standards for CDISC SDTM 3.2' (2015); Mark Elliot and others, Anonymisation Decision-Making Framework (UKAN Publications 2016), http://ukanon.net/ukan-resources/ukan-decision-making-framework.
6. IPC De-identification Guidelines for Structured Data, June 2016 (page 21): While de-identification techniques protect against the disclosure of individuals' identities, they do not protect against the disclosure of attributes relating to groups of individuals that may be stigmatizing to those individuals. Some inferences may be desirable insofar as they may enhance our understanding of a particular issue or topic. Others may subject groups of individuals to unjust or prejudicial treatment or would be considered offensive.
7. Generally, if the data has been properly de-identified, it is not considered a privacy breach, whereas if the data is personal health information (PHI), the data and the possible breach must be analyzed.