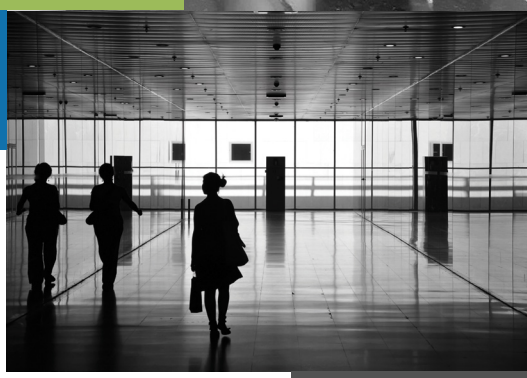
PRIVACY ANALYTICS

WHITE PAPER

GDPR and Anonymization

Pushing the idea of anonymization to the forefront is the European Parliament with the new regulation – General Data Protection Regulation (GDPR). The GDPR is a four-year development to improve the existing data security and privacy regulatory framework which encompasses the protection of personal data across the European Union member states and beyond.

To achieve a more seamless adherence to GDPR, it is best to create privacy-centric policies and procedures around sensitive data storage and sharing. This white paper explores anonymization under the new regulation.



PRIVACY ANALYTICS

a QuintilesIMS company

In our ever-connected technologic age, privacy is still important. In a study into the attitudes of early adopters of health technology towards the privacy implications of health data use, they found:

"..our findings indicate that despite an enthusiasm for contributing personal data to science, early adopters maintain concerns similar to those of 'privacy fundamentalists' and 'privacy pragmatists.'"[1]

The survey found that respondents wanted to be altruistic in the sharing of their personal health data for the benefit of science but also wanted to retain a degree of privacy and anonymity. Data sharing for health matters is on the rise.

Studies, such as a recent one carried out by Ipsos Mori/Wellcome Trust, have shown that patients will be willing to share personal health data as long as safeguards are in place[2]. Amongst these safeguards, anonymization of health data was seen as crucial to 52% of respondents. In fact, 94% of respondents thought that retention of privacy within medical records was essential.

In an environment where health data is seen as an extension of the individual, anonymity is arguably a right. In an environment where health data is found for sale, in vast quantities on the dark web, anonymity is a vital layer of protection. f we want to use health data to advance medical science while giving the patient peace of mind, anonymous data sets are a fundamental way to achieve this.

Pushing the idea of anonymization to the forefront is the European Parliament with the new regulation – General Data Protection Regulation (GDPR)[3]. The GDPR was approved on April 14, 2016, and will be enforced from May 28, 2018.

The GDPR is a four-year development to improve the existing data security and privacy regulatory framework which encompasses the protection of personal data across the European Union member states and beyond.

GDPR Fundamentals

The forerunner to the GDPR was the Data Protection Directive 95/46/EC[4], which was developed to protect the data of the individual within the EU states. The directive was originally outlined in 1995, pre-dating the current explosion of hyper-connected devices. As our environment of data creation and sharing has changed exponentially, the EU decided that this original directive needed to be updated to accommodate the many new ways in which Personally Identifiable Information (PII) is being generated, disseminated, and accessed.

The GDPR view on personal data begins with the concept of a 'data subject' and a data controller. In the case of health data, the data subject is the patient or individual sharing information about themselves within the context of healthcare. The controller is the individual or team within the organization who uses this data. The GDPR sets out a rack of expectations of the rights of that data subject to privacy protection.

In terms of privacy of an individual, there are three broad concepts used by the GDPR:

- **Data subject:** The individual
- **Personal identifier:** The GDPR is clear about what a personal identifier comprises. From a health data perspective, these identifiers include genetic, physiological, mental, and physical data.
- **Identifiability:** GDPR looks at methods of

'singling out' an individual by using their data, either directly, or indirectly.

The geographic scope of the GDPR is not confined to EU states. If data processing occurs outside of the EU, but will in some manner monitor EU residents' behavior, the GDPR rules will apply – no matter where that organization is based. For example, if a company manages a disease registry, or provides genome sequencing that allows EU citizens to take part, the company would need to abide by the GDPR.

Considering how an organization complies with the privacy expectations of the GDPR, then the whole area of 'linkage' is one of important consideration. Encouraging Privacy by Design (PbD) and gathering user consent to use and share data was previously enough to satisfy regulatory expectations. Now, the requirements are much more resolute. The language is less of 'Should' and 'May' and more of 'Must' and 'Shall.'

In terms of privacy and the GDPR, the key nuance to understand is around 'identity linkage.' This is making sure individual user attributes cannot be directly linked to an individual in a way that compromises their privacy. In this way, the GDPR is about respect for individual privacy. Respecting the rights of the individual entails the right to decide on the who, how, and why of their personal data usage. Organizations need to show they offer controls to the individuals whose data they are utilizing.

The requirements of the GDPR are all-encompassing and potentially onerous in terms of data and customer management. The rights of the data subject that need to be upheld by an organization include:

Research:

Registries and healthcare organizations (e.g., hospitals, clinics) currently submit patient data to registries. Data contained in these registries can be used for research and policy/administrative needs (such as a stroke or cancer registry). Often data is disclosed from a registry without patient consent under the assumption that it is de-identified.

Information Obligation (Articles 13 and 14):

This involves an organization giving extensive information about the use of the data to the data subject.

Right of Access (Article 15):

The data subject must be given right of access to their data at any time and have rights to query these data, with redress if these data is incorrect or requires changes.

Right to Rectification and Erasure (Article 16 and 17):

The data subject must be allowed to remove consent at any time. They must also have any inconsistencies or possible issues rectified. This is in line with the right to be forgotten, which has potential major knock-on effects across search engines management – the data controller being obliged to ensure search engines also 'forget' these data[5].

Right to Data Portability (Article 20):

This right of a data subject is one of the more complicated and far-reaching. It allows for an individual to request that the processed data is provided to them by the organization's controller, in a 'structured and commonly used and machine-readable format.' The data subject can then share this data with another controller if they so wish.

Right to Object (Article 21):

This article covers the use of automated processing of data. In terms of profiling, the GDPR includes the use of automation to predict health in this category. This article is strict in terms of getting explicit consent to perform automated processing.

Another important aspect of the GDPR is that the new regulations now apply to organizations processing data, not just those controlling or hosting it.This makes the GDPR much wider in scope. In terms of health data, this means any organization accesses health data to use for research or other purposes must comply with the dictates of the GDPR.

Achieving all of the expectations of the GDPR in terms of privacy for health data can add a massive administrative overhead to any program that requires access to, or the use of, individual health data. Everything from front-end UX/UI design to data management and archival is impacted by the expectations of the GDPR. Projects which utilize personal health identifiers are likely to find that the complex matrix of privacy requirements will touch many aspects of the business, including sub-contractors, who are also expected to adhere to GDPR privacy requirements.

The GDPR 'Hack': De-identification of Health Data

Although there is no avoiding the GDPR, there are ways to manage the expectations. The pivot upon which GDPR privacy stands is being able to identify an individual from their personal data, or being able to link them indirectly to this data. By preventing linking, the various requirements of the GDPR would be met. Recital 26[6] of the GDPR is the 'hack' that helps organizations comply with the GDPR. This section states:

> "The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.
>
> This Regulation does not, therefore, concern the processing of such anonymous information, including for statistical or research purposes."

In the case of health data, this 'get out clause' within the GDPR has clear implications. If you can de-identify the owner of that health data, in other words, anonymize the data, then you have met the requirements of GDPR by default.

This means that if you can anonymize health data, while ensuring the quality of the data for scientific purposes, you can meet GDPR compliance without impacting your processes. Removing the need to comply with GDPR, removes the overhead and spend of time, money and continual monitoring of compliance measures that ensure health data privacy.

What Happens if You Don't Comply with the GDPR?

The GDPR is now set in law. This is a regulation and not a directive. This gives the EU stronger powers to enforce the regulations it sets out. This makes anonymization of health data a worthwhile exercise because non-compliance with the GDPR carries onerous sanctions. The GDPR attempts to enforce compliance with as strong a mandate as possible. This includes much lower thresholds of adherence and much higher fines.

The current version of the regulation sets two levels of fines depending on the area where the noncompliance occurs:

1. Fines of 4% of revenue or 20 million Euros whichever is the higher, will be applied to noncompliance, including:
   - **Consent:** Consent is a major point of law under the GDPR. Fines associated with not explicitly gaining consent across all aspects of data usage.
   - **Upholding consumer rights.** This is a gray area and one which could result in compensation claims as well as fines.
2. Fines of 2% of revenue or 10 million Euros whichever is the higher, will be applied for non-compliance, including:
   - **Data Protection Officers.** The GDPR rules specify that an organization has to employ data protection officers to deal with GDPR requirements.
   - **Data breaches.** Various aspects of data breaches from time to respond to the size of breach need to be adhered to. The GDPR expects breaches notifications to be made to the relevant authority and, in some cases, the individuals affected.
   - **Privacy Impact Assessments (PIAs).** The GDPR requires that PIAs are carried out across all relevant systems and when systems are changed.

What is and What is Not Anonymization?

Anonymization or 'de-identification' can be used to avoid some of the more onerous aspects of meeting GDPR requirements. But what exactly is anonymization in the eyes of the GDPR?

The GDPR in recital 26[7] defines anonymization as:

"…personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."

To understand how anonymization can be achieved, we need to look at how, in a digital world, human beings are identified. There are a number of direct and indirect manners in which you can tie data to an individual:

- **Name:** Names by themselves are not unique. But as part of a wider data set, a name can help you focus in on an individual. A UK case, Edem vs. ICO and FSA[8], brought by the Information Commissioner's Office (ICO) sets a precedence for the use of a name as being part of a person's identifiable information.
- **Location:** Location can be used with other identifying data to pinpoint an individual. A Carnegie Mellon University study showed that you could easily identify 87% of US citizens using just three variables: Zip, gender, and date of birth[9].

- **Genetic and biometric data:** The GDPR has included genetic information in the schedule, as well as biometric data such as fingerprints, facial recognition, etc. Anyone who processes large batches of genetic or biometric data will be expected to carry out extensive PIAs to determine linkage issues.

With this in mind, anonymization of data needs to remove all possible ways information, such as genetic or other health data, could be linked back to an individual's directly identifying information, such as their name.

Having Your Cake and Eating it: Risk-based Anonymization

Anonymization of data is a science built around the fine balance between ensuring data is private, while maximizing the data usage for a given process. De-identification of the data is based on classifying:

- **Direct identifiers:** These are unique identifiers, such as social security number, that can be directly linked to an individual.
- **Quasi-identifiers:** These are identifiers such as gender, zip code, etc., that are shared by more than one person, but if used with other identifiers could trace an individual – as shown in Carnegie Mellon University's research[10].

When anonymizing data, you can use two methods:

- **Masking:** This is used to anonymize direct identifiers
- **De-identification:** This is used to anonymize quasi-identifiers

Masking alone will not de-identify a data record. There needs to be a balanced use of both methods to optimize the privacy, while retaining the use of the data. An example would be to remove the name (direct identifier) of the individual associated with the data, while keeping the date of birth (quasi-identifier) but generalizing to the decade or within a five-year span.

Anonymization, performed correctly, balances regulatory requirements and utility. You can "have your cake and eat it" with carefully implemented anonymization methods.

Using a Risk-Based Approach to Anonymization of Data

GDPR regulations are directly associated with the level of risk. Certain actions, such as anonymization, can provide for exemption under the GDPR rules. Having a risk-based approach to anonymization of data and compliance with the GDPR, allows an organization to manage better the workload expected when fully adhering to the compliance requirements. The GDPR itself, to help with compliance, encourages a risk-based approach when applying privacy measures, by having different exemptions and considerations built into the framework. When beginning your GDPR assessment, having an understanding of the level of risk associated with data, within a given process, will allow you to determine the level of de-identification necessary and accordingly help with GDPR compliance.

To achieve a more seamless adherence to GDPR, it is best to create privacy-centric policies and procedures around sensitive data storage and sharing. To do so requires the following:

1. Be privacy aware – understand the risks to individual privacy associated with the storing

and sharing data.

2. Understand the privacy expectations inherent in GDPR.
3. Know how to mitigate those risks using robust anonymization methods.
4. Know how and when to apply anonymization so that you optimize privacy and data use.

An example of a health organization using risk-based anonymization is the American Society of Clinical Oncology (ASCO). The ASCO organization uses risk-based anonymization software to de-identify cancer patients' data[11].This data is then aggregated into a learning healthcare system called CancerLinQ. CancerLinQ is a portal used by oncologists across the US to get better insights into treatments and cures for their patients. This portal could have serious privacy repercussions under the GDPR rules if this data were not anonymized. Without risk-based anonymization, ASCO could not easily accomplish such a wide-reaching project. Risk-based de-identification balances the GDPR privacy requirements while ensuring high-quality data can be shared safely.

Six Basic Steps to Reducing Risk under GDPR

Following some basic steps can help you to manage the requirements of GDPR:

1. Know what personal data you will be using and where you will use it.
2. Understand what aspects of these data fall under the various GDPR categories of 'personal data.'
3. Consider using smart anonymization of health data to simplify GDPR compliance.
4. Use PIAs where necessary to inform your decisions and where best to apply anonymization.
5. Make sure your business associates who may handle health data are also compliant.
6. Keep good documentation on all of your decisions and processes.

A Healthy Approach to GDPR Compliance

The implementation of GDPR compliance is a positive action by enforcing better privacy protection around the use of health data. The GDPR itself uses a risk-based approach to managing modern privacy expectations by having a pragmatic approach to privacy and setting exemptions for best practices such as anonymization. Accessing your risk and taking a risk-based approach to dealing with the requirements of the GDPR will give you leverage in compliance. If you map your data needs to the GDPR, and apply smart anonymization to your processes, you can more easily manage your privacy, while retaining your data quality.

Ultimately, a privacy-first approach will benefit all stakeholders. Your organization benefits from having access to essential health data for research; patients benefit by knowing their data is being put to good use, but is protected; and medicine and research benefits, because individuals feel more confident to share their health data. From this data sharing, better treatments are developed.

Anonymization, done correctly, is a win-win situation for the healthcare industry and patients alike.

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. Cheung, C., et.al., Privacy Attitudes among Early Adopters of Emerging Health Technologies, Nov 10, 2016: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5104519/.
2. Ipsos Mori/Wellcome Trust, The One-Way Mirror: Public attitudes to commercial access to health data, March 2016: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf
3. GDPR Official Portal: http://www.eugdpr.org/
4. The EU Data Protection Directive 95/46/EC: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995L0046
5. EU Factsheet on the Right To Be Forgotten Ruling: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf
6. GDPR Recital 26: https://www.privacy-regulation.eu/en/r26.htm
7. *Ibid.*
8. Efifiom Edem and The Information Commissioner The Financial Services Authority: http://www.bailii.org/ew/cases/EWCA/Civ/2014/92.html
9. Sweeney, L., Carnegie Mellon University, Simple Demographics Often Identify People Uniquely: https://dataprivacylab.org/projects/identifiability/paper1.pdf
10. *Ibid.*
11. For more information, read the ASCO Case Study: https://privacy-analytics.com/files/ASCO-Case-Study.pdf.