WHITEPAPER

Good Governance for Anonymized Data

PRIVACY ANALYTICS
an IQVIA company

Ensuring data is safe before and after anonymization

By Santa Borel, Associate Director, Data Privacy Solutions and Luk Arbuckle, Chief Methodologist

Technology changes fast and can have an immediate impact on data protection and privacy. Contact us to learn how Privacy Analytics can help with in-depth technology analyses or advisory and consulting services.

A lot of focus is given to data governance for personal data, and just as much focus can be given to data governance for anonymized data to manage residual risks. Before anonymization, organizations can establish policies, procedures, and roles for the anonymization process, to communicate effectively with stakeholders, and to know how to identify and manage potential breaches. After anonymization, continuous monitoring and control measures will ensure ongoing safe and responsible use of the data. By following best practice for a governance process before and after anonymization,(1) organizations can maintain the integrity of anonymized data and promote the safe and responsible use for intended purposes.

Before

making data available

How good governance for anonymized data can support the safe and responsible enablement of sensitive data

After

making data available

How to ensure norms for the safe and responsible use of anonymized data are respected and maintained

Before Anonymization

As you prepare to use or share data for new or innovative purposes, you can create a governance plan that supports the safe and responsible enablement of protected data. Assessing the necessity and impact of the use or sharing, as well as the legal premise, will help you determine organizational risk appetite.

We consider the following four aspects to setting up governance for anonymized data:

- **Assign roles and responsibilities** for the anonymization process and safe handling of data
- **Establish principles, policies and procedures** for the anonymization process
- **Communicate with stakeholders** to ensure everything is understood
- **Know how to identify and manage a breach** so that any incidents are handled swiftly

Assign roles and responsibilities

Establishing clear roles and responsibilities provides clarity and efficiency for effective governance of anonymized data. Start by assigning someone in a leadership role to be accountability for the governance process. Make sure the process is clearly defined, shared across the organization, and followed, preferably with documented proof of adherence. Recent regulatory decisions, new privacy legislation, and guidance emphasize the importance of ensuring that employees handling anonymized data have proper training on data governance practices as well as sign confidentiality agreements.(2)

To achieve this, organizations can implement Standard Operating Procedure (SOP) that covers the principles and procedures of the organization's use of anonymized data. Additionally, more detailed expert training can be provided to employees who work closely with data to help them understand how to protect data and what they can and cannot do with it. This training can be tailored to the specific needs of the role and cover topics such as disclosure control, data security, data protection and privacy law, and more.

End users should also receive training on the limitations of the data and anonymization, including the level of identifiability, basics of disclosure control and anonymization, and how the use of the data aligns to agreements. Overall, having well-trained employees will support a strong governance process for anonymized data.

To effectively oversee the anonymization process, one approach is to establish a cross-functional governance committee that aligns on enabling new uses and innovative approaches. At Privacy Analytics, we spearheaded the creation of a stewardship council comprised of senior leaders across different functions. This council's purpose is to enable safe uses of AI and privacy-enhancing technology while ensuring responsible data use. Additionally, a monitoring committee made up of functional leads acts as stewards of responsible data use, working in tandem with the stewardship council to identify areas of improvement and innovation.

Establish principles, policies and procedures

Principles, policies and procedures demonstrate good data governance within your organization and serve as the foundation for your organization's approach to handling anonymized data. The extent to which these are developed, in terms of detail and comprehensiveness, will depend on what is relevant for your organization's data management practices and the applicable regulatory landscape.

- Monitoring for future risks will allow you to anticipate and prepare for potential threats to your organization when preparing for use of data and once data is made available.

- Maintaining a record-keeping system and audit trail will allow you to track the generation of anonymized data and ensure that you are always aware of who has accessed it.

- The privacy impact of using anonymized data can be assessed, for example, through data protection or privacy impact assessments. This process involves a review of the contents of the data being made available and how it will be handled.

- Be prepared to advise on complex cases, such as the addition of genomics data, by being able to identify potential issues and develop strategies for dealing with them. Having clear guidelines in place for such scenarios can help everyone involved understands the impact on identifiability.

- Preparing for unintended disclosures with policies and procedures describing how they might occur and appropriate actions that can be taken to minimize their impact.

For many organizations, policies and procedures related to the use of anonymized data become an extension of their existing data governance framework. While this makes adoption easier, it does require updates to existing practices to account for the intricacies of working with anonymized data. For example, to work with anonymized data, the co-mingling of personal and anonymized data in the same environment or by the same users of data is generally avoided (unless the data are unrelated).

Communicate with stakeholders

Communicating with stakeholders about the use of anonymized data requires careful consideration. Many privacy and data protection authorities expect that the purposes of data use, including the act of anonymization, be shared when collecting and using protected data, including on organizational websites that are easy for data subjects to find and access.

You can start by first identifying relevant stakeholders and engaging with them to determine their needs. The amount of information shared will depend on the stakeholder group and what seems like an appropriate level of detail to share with them, as well as your organization's policies and requirements of applicable regulations.

Emphasizing beneficial uses of the data can be important in helping stakeholders understand the value of the data and promote its responsible use. Consider the data that will be made available, the environments in which that data will be used, and the safe and responsible uses when deciding on the level of information to share. This can help ensure that the communication is transparent and builds trust with stakeholders. Additionally, decide on the appropriate level of confidentiality for the information that will be communicated.

Effective communication can take various forms, including a statement in a consent form, a publicly accessible website that outlines the purposes or through stakeholder discussions and public presentations. By considering these factors, your organization can establish transparent communication with stakeholders about the use of anonymized data, helping to build trust and promote responsible data practices.

Communication plans can become complex when engaging with different stakeholder groups. To tailor the messaging to specific audiences, we sometimes rely on user groups to craft the message. For instance, in a public setting, the emphasize may be placed on the importance of beneficial uses with the appropriate assurances of safe and responsible usage. On the other hand, a business audience may find messaging around the compliant use of data to drive business outcomes more appealing.

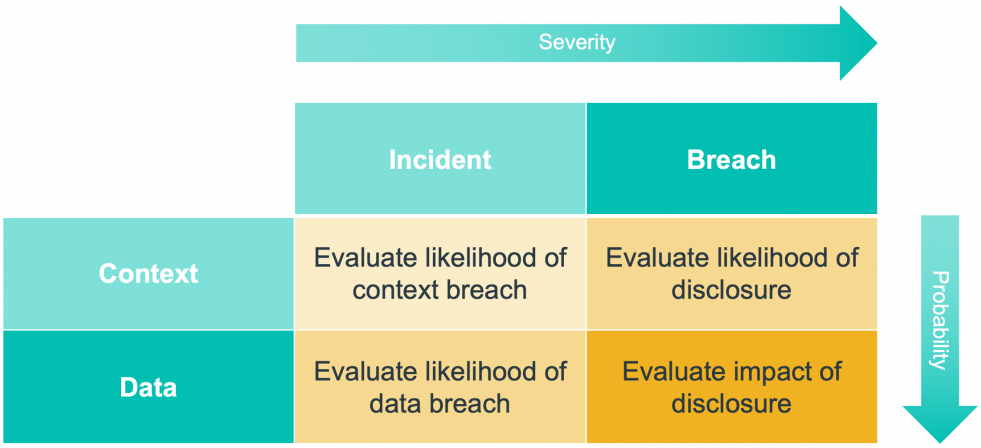
Know how to identify and manage a breach

Identifying and managing a potential breach is part of the overall process of data governance in securing and protecting data from unauthorized access, use, or disclosure. The contextual nature of anonymized data introduces some key differences. In particular, the identifiability of the data can be compromised in multiple ways that need to be assessed to determine the likelihood and impact of a breach.

Start by establishing a clear definition of a disclosure, which is generally broad and can change depending on jurisdiction and other factors. ISO defines disclosure as "revealing confidential or personally identifiable information from a dataset based on a vulnerability that is found or exploited". This definition can be made more specific to the organization's functional area and processing of data.

The distinction between context and data affects whether the an event is considered an incident or a breach, and whether that event needs to be reported to the appropriate authorities. Changes in context can affect the identifiability of the data, so any changes or vulnerabilities that could compromise the security of the data should be reviewed carefully. If the context assessment changes significantly, it may be necessary to treat the event as a *context breach* and evaluate the likelihood of a disclosure. A context breach is a precursor to a more comprehensive analysis to determine whether there was any actual disclosure of information.

Similarly, changes to data can affect its identifiability, so any changes or vulnerabilities that could make the data more identifiable should be reviewed carefully. If the data assessment changes significantly, it may be necessary to treat the event as a *data breach* and evaluate the impact of disclosure. Depending on the jurisdiction, the data breach itself may constitute a reportable breach to the appropriate authorities, or the impact of disclosure will be used to make the final determination on reporting. Either way, the impact of disclosure will help determine next steps.



Organizations take any event that may be classified as a breach very seriously. Although we have helped with crisis management assistance, the circumstances of each situation can vary greatly. When there is a leak of information, a risk assessment is conducted to determine the extent of the leak and whether it warrants a breach response. Similarly, if the circumstances of data use have significantly changed, a risk assessment on the data environment, access, and sharing will determine whether there is a need to evaluate the data for potential disclosures. Evaluating multiple steps before concluding that the data itself is at risk due to contextual changes is important. These cases are often complex, and a quick response is necessary to contain and control the situation before making final determinations.

After Anonymization

Once anonymized data is made available to end users, data governance can support norms for the safe and responsible use of the data are respected and maintained. While the data might have been considered safe at one point in time, monitoring can support ongoing use.

We consider the following four aspects to maintaining governance for anonymized data:

- **Monitor the data environment** for changes in the recipient and their conditions
- **Scan the horizon for changes** to adapt to future data environments
- **Mitigate when there is an incident** to contain a potential disclosure and control the impact
- **Review and continuously improve** your governance process

Monitor the data environment

Anonymization considers who will have access to the data, under what circumstances, and what other data could be available to them. You can monitor these considerations to ensure the identifiability assessment remains valid or whether any changes are needed.

Managing access to data involves registering all data that is made available, tracking recipients of the data, and comparing it to previously released data to highlight any potential overlap that could introduce new risks (such as differencing attacks). If a recipient gains access to a new dataset that can be linked to previously released data, the anonymization could be compromised. Additionally, changes in the environment, such as updates to terms of use, could affect the identifiability of the data by changing its context. Any of these examples could necessitate a re-assessment of the anonymization.

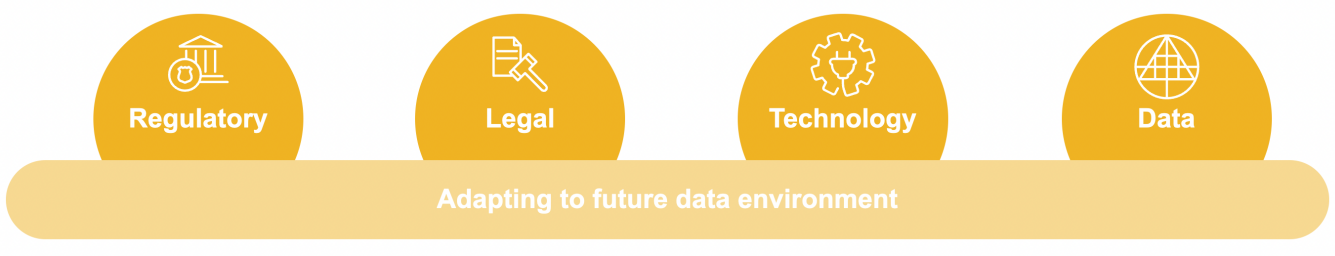
Let's take a simple example of a differencing attack. Assume that aggregated data is represented in a table, with columns representing the identifying attributes of region, sex and age. Say the table was produced for the month of February, in which [region A, sex M, age 30–39] = 15, and for the month of March, an updated table is produced with [region A, sex M, age 30–39] = 16. It's easy to see that one person was added from one month to the next, which would violate the spirit of aggregation rules intended to avoid disclosures about single individuals.(3)

Scan the horizon for changes

Changes in regulations, legal precedents, technology, and data can change the evaluation of identifiability as well as expectations around data anonymization. As an organization, you can scan the horizon for any new developments that will affect your assessments of anonymization or use and sharing of anonymized data to update the governance process and identifiability assessments.

- To stay informed, engage with industry groups, sign up for newsletters or alerts on relevant topics, engage internal stakeholders on knowledge sharing.

- Look for the release of new guidance and decisions by the appropriate authorities to understand concerns and opinions regarding data reuse.
- Keep up with legislative updates, case law, or views impacting operations to assess boundary conditions in the specifics of law and rulings.
- Track technological developments regarding privacy and data reconstruction methods or relevant disclosures, including practical assessments and how they are perceived.
- Watch for any successful attacks on data to discover new or novel disclosure opportunities and assess how realistic they may be in practice.



Data protection and privacy professionals stay informed on industry updates through various means, such as newsletters and conferences. The challenge, however, is often managing the volume and variety of information, which includes legal and policy interpretations as well as technology updates. Making data protection and privacy engineering, with technical know-how, is rapidly becoming a core competency for organizations.

Mitigate when there is an incident

In the event of a potential disclosure, immediate action can contain the incident and minimize damage by following previously defined plans. It's important to determine the extent of the incident and whether it constitutes a breach, taking into account the context and the data involved.

To respond effectively, you probably want to have established procedures and a crisis management plan in place. The initial focus should be on containment. Following containment, the incident can be assessed, and any gaps can be mitigated to prevent further harm.

During the investigation, identify why the incident occurred and what your organization can learn from it. This can be achieved by reviewing the audit trail and crisis management procedures to identify areas of improvement. Documenting lessons learned from the incident can be used to improve future processes.

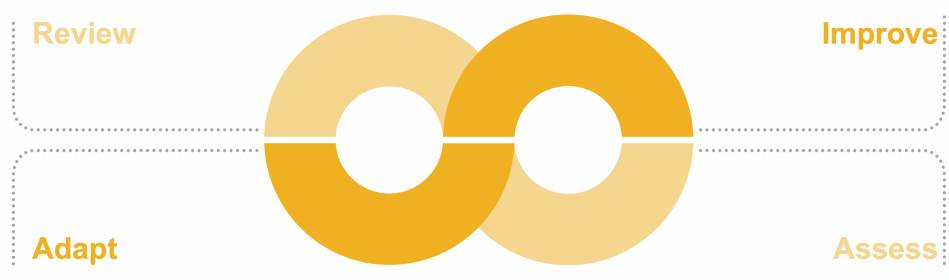
Additionally, motivated intruder (or adversarial) testing can be conducted to improve the mitigating controls in place and demonstrate redress after an incident or breach. This involves simulating an attack on the data and environment to evaluate the effectiveness of the anonymization process and identify any vulnerabilities that need to be addressed.

By demonstrating adherence to established procedures, regularly reviewing audit trails, and conducting motivated intruder testing, organizations can minimize the risk of data breaches and respond effectively when incidents occur.

Organizations we have worked with are often advised by legal to have a third party assess the incident to determine its severity and how to mitigate it in the future. This type of assessment has been proactively shared with the relevant authorities, even when the determination was that the leak was minimal and appropriately managed with mitigation steps. The assessment can also be kept as part of an internal record to explain risks and implement new procedures to prevent future incidents.

Review and continuously improve

To ensure that the governance framework remains effective and up-to-date, periodically review and make updates as needed. This can be done through regular assessments of data, either on a 12-24 month interval or by using a platform or bulk approach to confirm the data environment remains safe for its intended uses.



Adapt your governance to evolving views on the benefits and risks of reusing and sharing data. For example, during the pandemic, many private and public organizations recognized the need to protect patient privacy while advancing accessibility and use of health data to understand the spread of the disease and learn more about successful treatment options for patients.

Data governance can be redefined as a key aspect of data enablement, highlighting the value of data and the supportive role that governance can play in enabling its beneficial use. This reframing allows organizations to view governance as a lever for ensuring access and effective use of valuable data assets. Increasingly, Chief Data Officers are recognizing the significance of this concept and are working within their organizations to shift mindsets towards responsible reuse and sharing of data.

As with any data governance framework, norms and standards will evolve and require us to update our practices accordingly. Even the concept of agility and adaptability is a norm that has been codified in data governance standards. While everything in this document is consistent with best practice, make it part of your approach to continually review and assess.

Conclusion

Anonymized data can be a valuable resource for advancing research and achieving socially and organizationally beneficial outcomes. Effective and responsible use of anonymized data requires the establishment of good governance practices, both before and after anonymization. With these practices in place, organizations can leverage anonymized data to drive innovation and outcomes, while also managing residual risks.

Bibliography

1. International Organization for Standardization. Information Security, Cybersecurity and Privacy Protection – Privacy Enhancing Data De-identification Framework (ISO/IEC Standard No. 27559: 2022) [Internet]. Vernier, Geneva; 2022 p. 22. Available from: https://www.iso.org/standard/71677.html

2. IPC. PHIPA Decision 175, A Group of Medical Clinics and Related Entities [Internet]. Toronto, ON: Information and Privacy Commissioner of Ontario; 2022 Mar p. 25. Report No.: HI19-00007. Available from: https://decisions.ipc.on.ca/ipc-cipvp/phipa/en/item/520967/index.do?q=ipc

3. Arbuckle L. Aggregated Data Provides a False Sense of Security [Internet]. IAPP Privacy Tech. 2020 [cited 2020 Apr 27]. Available from: https://iapp.org/news/a/aggregated-data-provides-a-false-sense-of-security/

4. Information and Privacy Commissioner of Ontario. De-identification Guidelines for Structured Data [Internet]. 2016 Jun. Available from: http://bit.ly/1PkrnMF

5. Elliot M, Mackey E, O'Hara K. The Anonymisation Decision-making Framework (2nd edition): European Practitioners' Guide [Internet]. Manchester, UK: UKAN Publications; 2020 p. 119. Available from: https://eprints.soton.ac.uk/445373/