The Five Safes of Risk-Based Anonymization

Luk Arbuckle (Privacy Analytics), Felix Ritchie (UWE Bristol)

Introduction

The sharing of data for the purposes of data analysis and research can have many benefits. At the same time, concerns and controversies about data ownership and data privacy elicit significant debate.

So how do we utilize data in a way that protects individual privacy, but still ensures the data are of sufficient granularity that the analytics will be useful and meaningful? Data anonymization (sometimes also called de-identification, depending on the jurisdiction) is the process of removing detail in the data and/or adding other controls to reduce re-identification risk. Good anonymization should mitigate exposure, and allow you to easily demonstrate that you have taken your responsibility toward data subjects seriously.

For example, take the healthcare scientist that requests individual patient data. This scientist may be requesting data be circulated internally within the same organization, from healthcare center to a research wing. Alternatively, this scientist may be requesting data as an external third party, from one healthcare center to another institution.

These different users, settings and objectives complicate decisions about the appropriate level of anonymisation. In this paper we demonstrate how a popular risk-based approach that has emerged from statistical data sharing by government agencies, known as the Five Safes, can be operationalized in a broader setting using concepts from risk-based anonymization.

Risk-Based Anonymization

Relying just on reducing detail to anonymize data risks making the data worthless for analysis, putting the healthcare scientist's goals at risk. On the other hand, not reducing the detail to the minimum risks breaches of confidentiality; for example, a rogue lab analyst/technologist may attempt to re-identify individuals out of curiosity. However, this risk may be better managed by retaining data detail but having strong mitigating controls.

We want the healthcare scientist to achieve the objectives for their lab, to improve healthcare outcomes and/or operational efficiencies, while maintaining data privacy. It is true that data utility is proportional to re-identification risk if data are public. But for non-public data sharing, as in the case of our healthcare scientist, data use becomes safe as a function of both data transformations and technical and administrative controls.

Risk-based anonymization acknowledges that all elements of data management need to be considered jointly: the exact data you're working with, the people you're sharing it with, and the goals of subsequent analysis.[1] With this information in hand, we can think more clearly about safeguards that will result in the responsible sharing and use of data. The process is as follows:

- Identify the need, the likely constraints on usage, and the intended flow of data into and out of the system.
- Evaluate specific implementation and controls with respect to ethical and legal approval or contracts, users, IT systems and outputs.
- Apply appropriate transformations to the data to deal with residual risk.

This user-centered process should ensure that data utility is maintained whilst satisfying the scope of privacy or data protection regulation or legislation. That way data analysis and research will still yield accurate and meaningful results.

Five Safes

One framework that has gained popularity over more than a decade of use is the Five Safes.[2] This framework captures the relevant dimensions to assess the context and results of a data sharing scenario in an effort to make sound decisions. The overall goal is "safe use".

The advantages of the Five Safes as a framework for data sharing is that it can capture a range of options, ensuring data utility is proportional to privacy and confidentiality protections, while considering affordability and feasibility of the data sharing scenario. "Safe" is not a limit but a spectrum (i.e., "how safe is it?"), so that a range of options emerge.

As a risk assessment framework, the Five Safes acknowledges the subjectivity of managing competing risks. Greater emphasis is therefore placed on empirical evidence to drive decision-making. We summarize this framework, operationalized using concepts from risk-based anonymization, in Figure 1.

[1] K. El Emam and L. Arbuckle L. *Anonymizing Health Data: Case Studies and Methods to Get You Started*. O'Reilly Media, 2013 (updated 2014).

[2] F. Ritchie. The 'Five Safes': A framework for Planning, Designing and Evaluating Data Access Solutions. *Data For Policy 2017,* London, UK, Sep 2017.

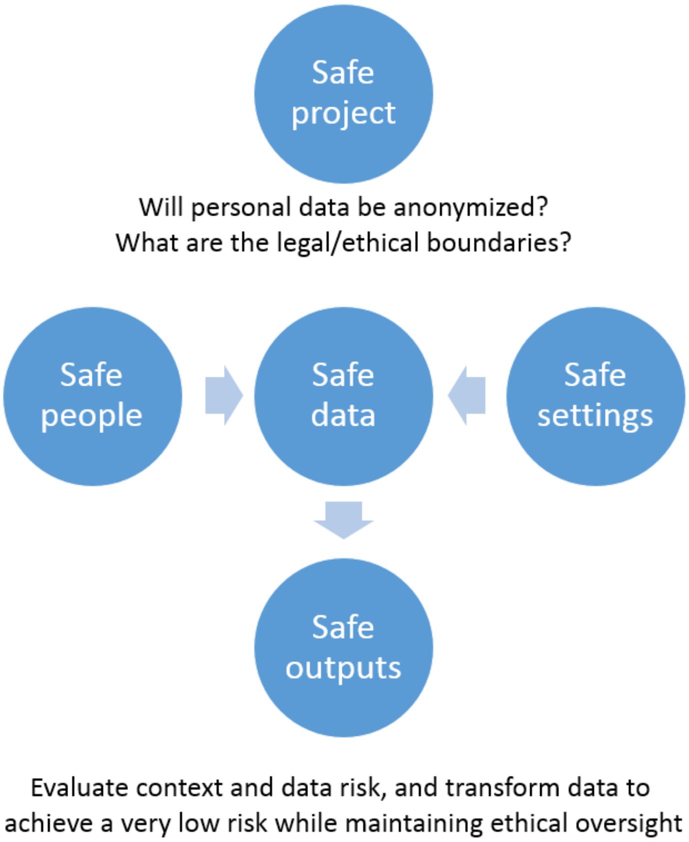
Safe project
Will personal data be anonymized?
What are the legal/ethical boundaries?
Evaluate recipient trust, and manage their motives
Safe people
Safe data
Safe settings
Assess security and privacy controls of the recipient
Safe outputs
Evaluate context and data risk, and transform data to achieve a very low risk while maintaining ethical oversight

Figure 1: Overall risk exposure using the Five Safes, operationalized through risk-based anonymization.

Our healthcare scientist will act a case study in how to apply the Five Safes in practice, in which we are the data custodian deciding how to share data. We demonstrate how overall risk exposure can be kept very low through a combination of both data transformations and technical and administrative controls, with objective support through statistical risk estimation.

Safe Projects

The healthcare scientist's work is for a purpose that is secondary to the original purpose of patient care and claims processing. Personal data will need to be anonymized, and we need to determine the mitigating controls necessary in the scientist's lab (the data environment) to ensure data protection is commensurate with the desired granularity of data. We will also want to know the purposes of the research, to determine whether an ethics review is required.

It's crucial we understand the flow of data at the outset of an anonymization project, and recognize both legal and ethical limits at each stage. Ethical considerations can in fact be as important as legal ones. Ethical breaches are likely to adversely affect the reputations of all parties involved, even if no law is broken.

We use an examination of data flows and data uses to determine whether personal data will be anonymized, otherwise we need not continue with our risk-based assessment of re-identification risk. And if data will be anonymized, for whom and under what circumstances (see the Data Flow and Data Use sidebar). This leads us to consider the people and settings in which data are shared, which will influence overall risk exposure.

Data Flows and Data Use

To initiate a risk-based anonymization project, we first need to understand whether personal data need to be anonymized and for what purpose. Key questions include:

- Where the source data come from and the lawful/ethical basis for collection.
- Where the source data go for processing, and the legal/ethical basis for processing.
- Whether the use of outputs of the sharing raise further legal or ethical issues.

When receiving a request to share data with an internal or external recipient, there are three scenarios to consider:

- Mandatory sharing: No approval is required, and the data do not require anonymization because it is likely that individuals need to be identified (e.g., law enforcement). However, there may be considerable underreporting by data subjects due to privacy concerns (e.g., avoiding treatment or care).
- Permitted sharing: Approval may be optional, under the discretion of the data custodian, for the public good (e.g., public health). There may be reluctance from data custodians to share personal data due to issues of individual and public trust, which anonymization can help remedy.
- Anonymous sharing: When approval is not possible or practical, and there are no exceptions in the legislation, the custodian must anonymize the personal data before sharing with a data recipient.

We should also consider building oversight mechanisms, determine how they can meaningfully engender individual trust for the ethical use of data to avoid causing harm, and ensure these obligations are transferred to downstream organizations. Finally, we should consider what happens at the end of the project, in terms of retention period and deletion.

Safe People

Having determined that patient data will be anonymized, the lab analysts/technologists will be central to an assessment of re-identification risk. Data recipients are human: they make errors, ignoring procedures they dislike or don't understand. More importantly, they are also potential adversaries, re-identifying data for personal gain or purely out of interest, with both motives and capacity.

We are more likely to trust a scientist that is studying healthcare outcomes/operations, with little interest in expending resources seeking external information for identification purposes, or skill/interest in developing novel re-identification attack models. That trust will increase by managing potential motives to re-identify through enforceable contracts or data sharing/use agreements. These need to include very specific clauses (otherwise there are legitimate ways to re-identify data):

- A prohibition on re-identification, contacting data subjects, or unauthorized data linking or sharing.
- Audit requirements allowing for spot checks and/or third-party audits.
- A requirement to pass on the above restrictions to any other party the data is subsequently shared with.

The lab analysts/technologists may also inadvertently re-identify data simply by recognizing individuals, for example an acquaintance or someone that is well-known. Even our trusted healthcare scientist may know someone in the data, especially if they live or work in the same general area. They may also ignore or avoid procedures, thereby allowing others unauthorized access to the data. As the latter is a behavioral problem, it can in part be managed through training and security/privacy controls.

Safe Settings

Having identified the anticipated recipients, we need to evaluate the lab environment and the circumstances in which the data are shared with them (see the Mitigating Controls sidebar). A rogue analyst/technologist is not bound by contractual obligations, and so we need to consider strong mitigating controls. The security and privacy practices of the lab will affect the motives and capability of an analyst/technologist to breach confidentiality. It also determines the likelihood of an outsider gaining access to the shared data, through deliberate hacking or the failure of staff to follow procedures.

Mitigating Controls

An evaluation of mitigating controls needs to be detailed and evidence based, preferably mapped to existing professional, international, and government regulations, standards, and policies, including ISO/IEC 27002, where appropriate. Using a standardized approach also ensures consistency, not only for a single organization that is sharing data, but across organizations, e.g., the HITRUST De-Identification Framework.[1]

Basic protections include

- Controlling access, disclosure, retention, and disposition of data
- Auditable authentication measures
- Remote access that, if allowed, is secure, monitored and logged
- Prevent malicious or mobile code from being run on servers, workstations and mobile devices
- Secure authenticated data transmission
- Physical security to protect access to computers and files.
- Someone in a position of seniority who is accountable for the privacy, confidentiality, and security of data
- Internal or external auditing and monitoring of activity

Depending on the nature of the data, not all protections are necessary. Regardless of what controls are in place, these will be factored into risk estimation.

Reference:

1. HITRUST Alliance. HITRUST De-Identification Framework, 2015.

Security requirements in the healthcare scientist's systems are closely related to the trust we place in their ability to protect data. The healthcare scientist that has poor data protection practices will require more data transformations to eliminate residual risk, ensuring we mitigate against the risk that they misuse or inadvertently lose the data. This is, however, also an incentive for them to improve their security and privacy practices; demonstrable competence in data handling can support access to higher utility/more granular data.

Safe Data

To determine the data transformations necessary to deal with residual risk, we need to understand the risk from the data itself. Data risk comes from both direct and indirect identifiers, not all of which will be required by our healthcare scientist, no matter how safe their lab is:

- Direct identifiers: Attributes that can essentially be used alone to uniquely identify individuals or their households, such as names and known identifiers.

- Indirect identifiers: Attributes that can be used in combination with one another to identify individuals, such as known demographics and events.

Direct identifiers are rarely of analytical value, and so can be removed, or replaced with random values or uninformative pseudonyms (to maintain referential integrity). Indirect identifiers contain useful analytical information that will be needed for data analysis and research (otherwise they can be removed). But specifying the combination of variables that identifies a data subject depends on both data and context.

Risk measurement will combine threat modeling (see the Threat Modeling sidebar) with models of plausible re-identification attacks, making assumptions about the real world. These assumptions should be made explicit, and there is much evidence about plausible risks. However, there will always be an element of subjectivity in these measurements, and this needs to be acknowledged.

Threat Modeling

A structured approach can be used to assess context risk (Safe People and Safe Settings) and evaluate whether an attack will be realized. Consistent with the modelling of threat sources used in information security and risk modelling, there are three plausible attacks that can be made on data:

- Deliberate: A targeted attempt by the data recipient as an entity, or a rogue employee due to a lack of sufficient controls, to re-identify data.
- Accidental: An unintentional re-identification, for example an individual being recognized while a recipient is working with the data.
- Environmental: The data could be lost or stolen in the case where all the controls put in place have failed to prevent a data incident, such as a natural disaster.

To produce Safe Data the overall risk of re-identification needs to be assessed, which is a combination of context risk (the probability of an attack) and data risk (the probability of re-identification when there is an attack).[1] This will drive the data transformations required to reduce risk to be statistically very small.

Reference:

1. C. Marsh et al. The Case for Samples of Anonymized Records from the 1991 Census. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 1991 Mar;154(2):305-40.

With a measure of overall risk, decisions can be taken to reduce detail in the indirect identifiers. To ensure a very small risk, privacy protection will be proportional to analytic utility, and this will depend on the default preference of the healthcare scientist.[3] Information loss needs to be minimal so that the data retain utility for analysis, while ensuring re-identification risk is very small. This way our healthcare scientist gets useful data to derive new insights, while ensuring data are no longer personal.

Safe Outputs

We want to ensure that the use of anonymized data is low risk. However, we also need to consider the result of data use: analytical products, or recommendations for action arising from the analysis. These are outputs, as they are derived from the Safe Data.

The healthcare scientist may, for example, learn from the anonymized data that a specific population group is under vaccinated. A decision to target this group for educational intervention may seem harmless, but what if others in the community who are vaccinated learn of this intervention, or the scientist decides to publish the results? This under-vaccinated population group could be harassed or discriminated against. Or what if the scientist decided to build analytical models that predict under-vaccination ('stigmatizing analytics')?

Outputs from anonymized data can breach confidentiality; good practice requires procedures to check this.[4] An ethics review board can help to manage risks from outputs of anonymized data, by advising and making decisions on whether particular uses of data are appropriate. Again, a clear statement of purpose at the project stage is essential for managing ethical uses of anonymized data.

To determine the Safe Outputs that are produced from our Safe Data, we need to set a risk threshold that will define "very small risk". No matter how much we trust our medical scientist, or the safety of their lab, we must recognize that expectations of privacy will vary based on a variety of factors (see the Risk Threshold sidebar), and our goal here is to capture that variation to make sound decisions around risk tolerance.

Risk Threshold

A subjective criterion is used by the data custodian to determine a risk threshold based on their risk tolerance. That is, how we define statistically very small risk, using precedents involving reputable organizations as a guide, such as national statistical organizations and minimum cell-size rules.[1] Key criteria include:

- The benefit to individuals from analysis of the data.

[3] F. Ritchie. Access to Sensitive Data: Satisfying Objectives Rather Than Constraints. *Journal of Official Statistics*. 2014 Sep 1;30(3):533-45.

[4] M. Brandt et al. *Guidelines for the Checking of Output Based on Microdata Research,* 2010.

- The sensitivity and personal nature of the data (e.g., information about a stigmatized disease or condition).
- The potential injury to individuals from an inappropriate processing or use of the data ('stigmatizing analytics').
- The appropriateness of approval by data subjects for disclosing the data.

Designing data management processes involves judgment of risk. The practical consequence of evaluating the above is that the threshold (the definition of “very small risk”) will be lower under the most invasive scenario, and this affects all decisions made about data reduction and other controls.

References:

1. Federal Committee on Statistical Methodology. *Statistical Policy Working Paper No. 22: Report on Statistical Disclosure Limitation Methodology.* Statistical and Science Policy Office of Information and Regulatory Affairs Office of Management and Budget, 2005.

Conclusion

It is possible to share data with our healthcare scientist that are of sufficient granularity that the analytics will be useful and meaningful, and in a way that protects individual privacy. We can ensure that the overall risk exposure is very low in part by improving their lab’s security and privacy practices.

In many jurisdictions, demonstrating that data have a very small risk of re-identification is a legal or regulatory requirement. The Five Safes, operationalized through risk-based anonymization, provides a basis for meeting these requirements in a defensible, evidence-based way. And the framework allows for the evaluation of scenarios of responsible data sharing, which will be context driven given the impact different scenarios will have on data utility.

Data utility is important for those using anonymized data, because the results of their analyses inform services provided, policy, and investment decisions. The cost of getting access to data is not trivial, making it important to ensure the utility of the data received. We don’t want to spend time and money collecting highly-useful data, only to then watch that usefulness deteriorate through anonymization. Framing design questions using the Five Safes can help to clarify early on where compromises need to be made.