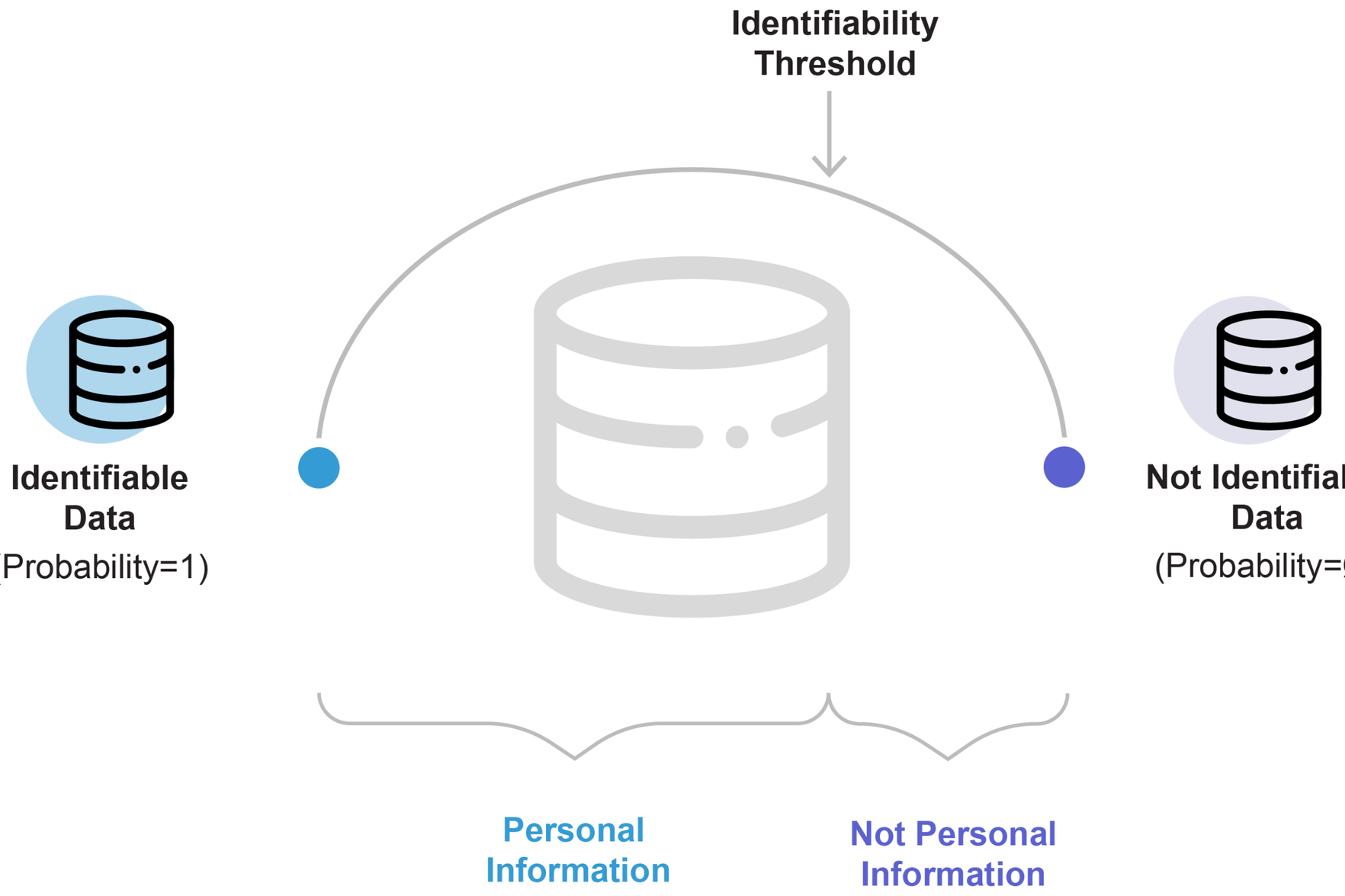
The Definition of Data Anonymity

Khaled El Emam
kelemam@ehealthinformation.ca

Identifiability spectrum and risk thresholds

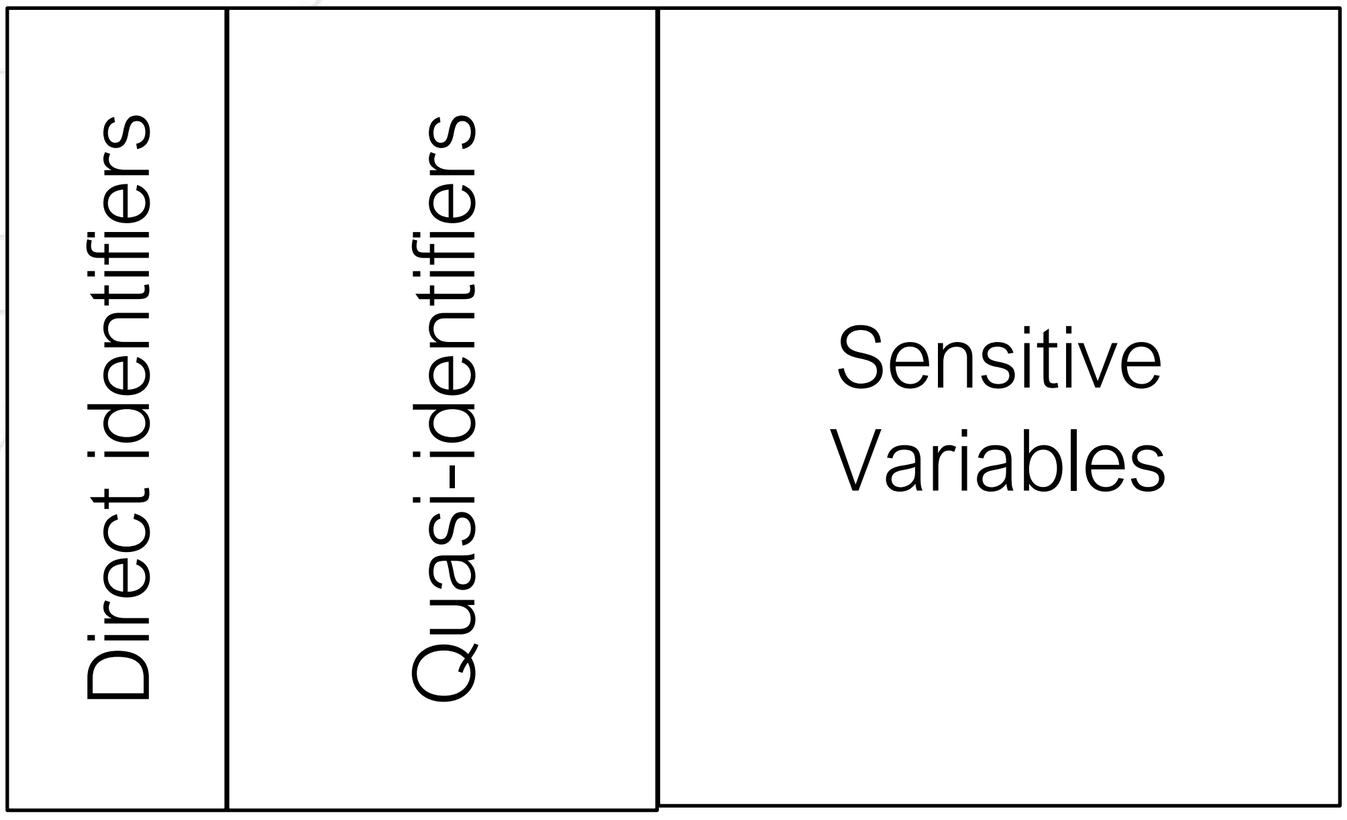


Components of re-identification risk

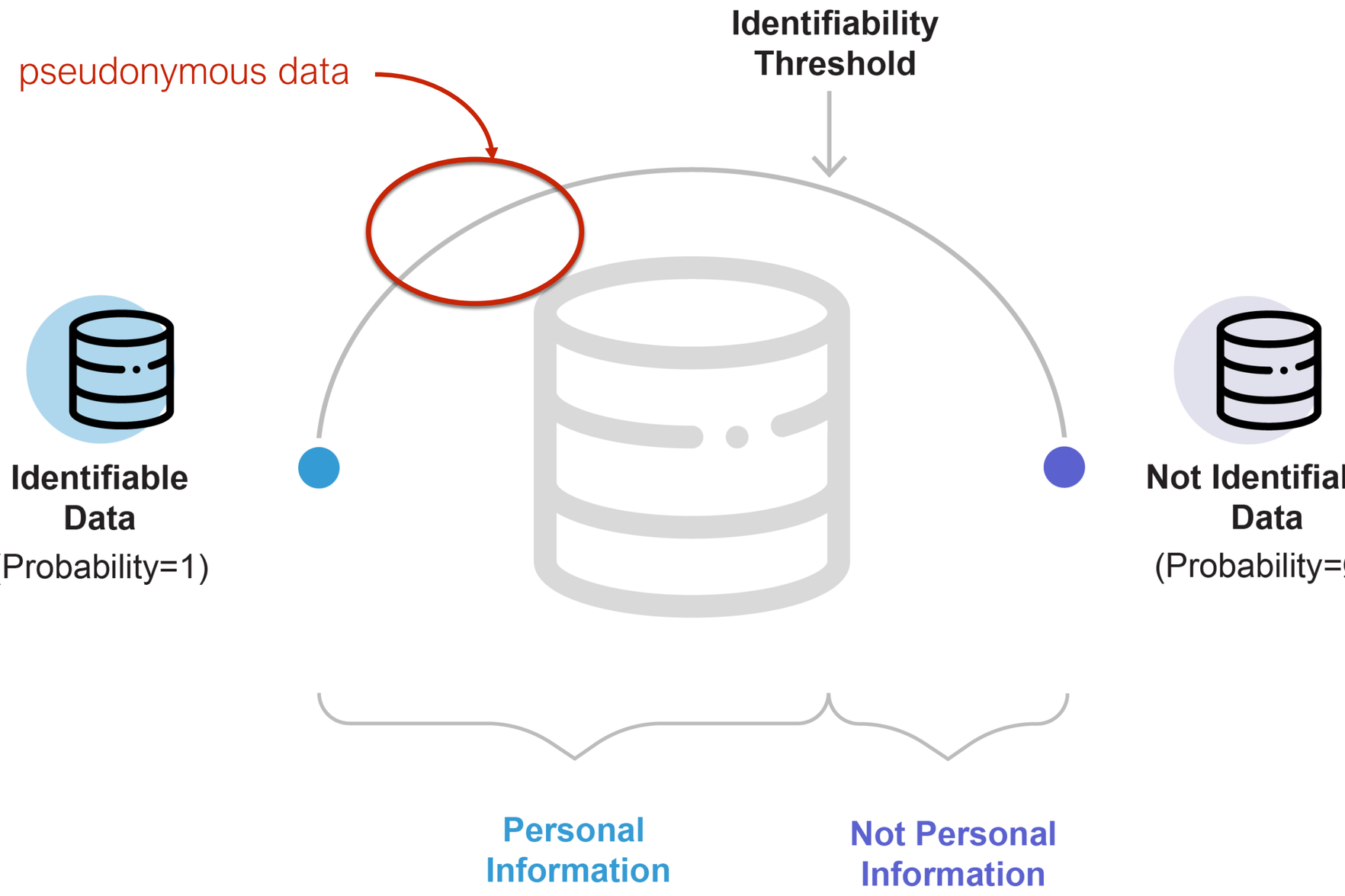
- Security controls
- Privacy controls
- Contractual controls
- Motives & Capacity

A risk-based methodology is consistent with existing standards and guidelines

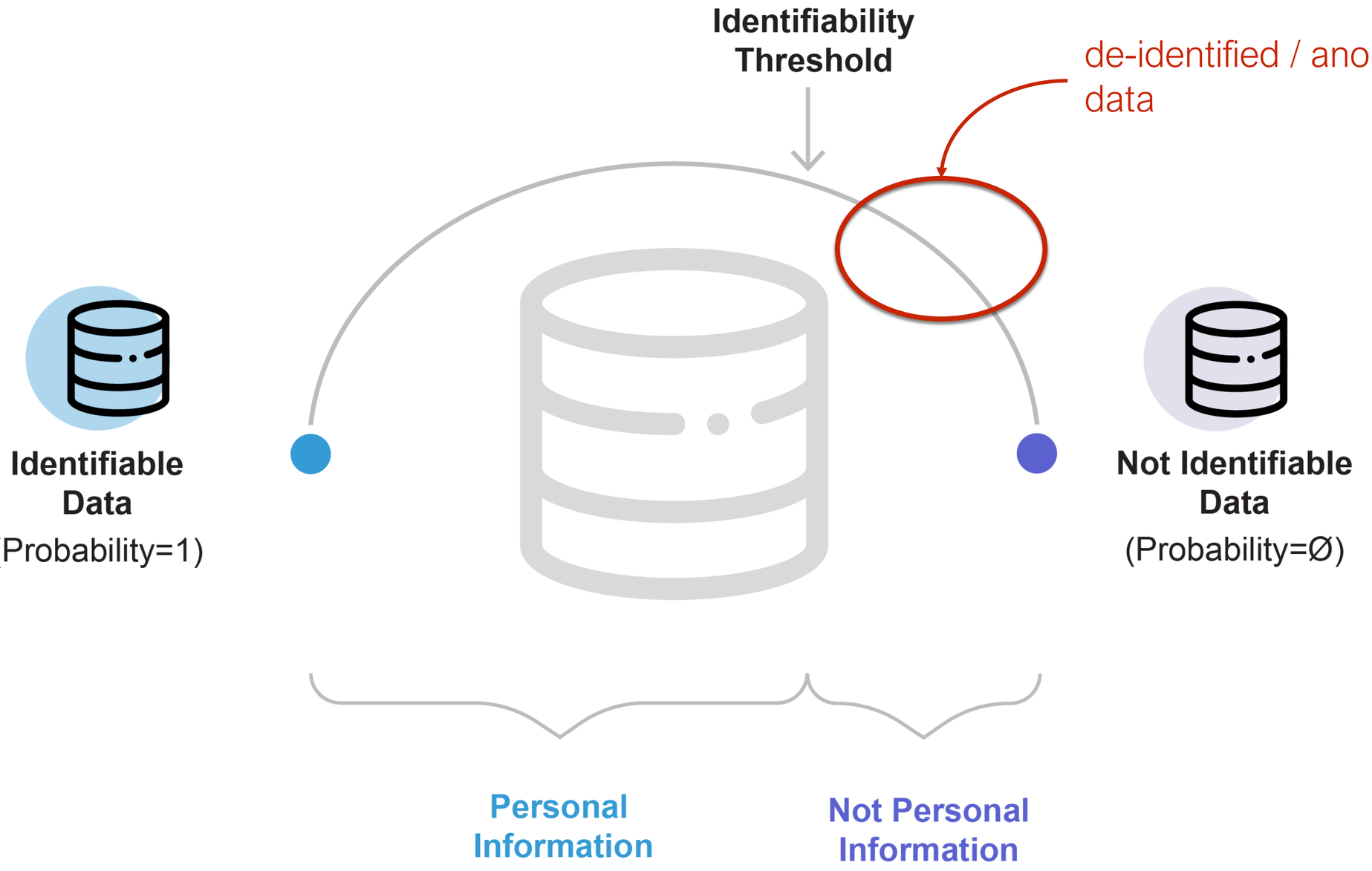
Variables in a dataset can be classified into one of three types



Identifiability spectrum and risk thresholds



Identifiability spectrum and risk thresholds



ISO Standard Thresholds

Scenario	Content (Matrix)	Threshold
Public	High possibility of attack, low impact	Max 0,1
	High possibility of attack, medium impact	Max 0,075
	High possibility of attack, high impact	Max 0,05
Non-public	Low-med possibility of attack, low-medium impact	Avg 0,1
	Medium possibility of attack, medium impact	Avg 0,075
	Medium-high possibility of attack, medium-high impact	Avg 0,05

Ideal definition of identifiability / anonymity

Definition	Source	Explanation
no reasonable basis to believe that the information can be used to identify an individual	45 CFR 164.154 (US HIPAA)	reasonableness standard applied to the overall concept of identifiability
appropriate knowledge and expertise	45 CFR 164.154 (US HIPAA)	the individual(s) performing de-identification must have expertise in de-identification
generally accepted scientific and statistical principles	45 CFR 164.154 (US HIPAA)	use current good practices, which allows the practices to evolve over time based on accumulating evidence
risk is very small	45 CFR 164.154 (US HIPAA)	identifiability is relative and not absolute and there is an implied acceptable risk threshold
alone or in combination with other reasonably available information	45 CFR 164.154 (US HIPAA)	reasonableness standard should be applied in making the assumptions about what an adversary would know (i.e., do not apply the precautionary principle)
anticipated recipient	45 CFR 164.154 (US HIPAA)	the risk is assessed from the perspective of the anticipated recipient and not from the perspective of any third party who may not, in practice, get access to the de-identified information
document the methods and results	45 CFR 164.154 (US HIPAA)	an important documentation requirement

Ideal definition of identifiability / anonymity

Definition	Source	Explanation
reasonably foreseeable in the circumstances	Act to modernize legislative provisions as regards the protection of personal information (Law 25)	include contextual information in the evaluation of identifiability
A health information custodian may use personal health information about an individual […] [for] the purpose of disposing of the information or modifying the information in order to conceal the identity of the individual	PHIPA s. 37(1)(f)	the act of de-identification / anonymisation is a permitted use and does not require additional consent
It is an offence for a person knowingly or recklessly to re-identify information that is de-identified personal data without the consent of the controller responsible for de-identifying the personal data	Data Protection Act 2018 s. 171(1) (UK)	prohibition on re-identification

Ideal definition of identifiability / anonymity

Definition	Source	Explanation
the risk assessment must be repeated every 24 to 36 months	current practices	ensures that evolving technology and risk profiles are taken into account by revisiting assumptions on a regular basis
if a company makes such de-identified data available to other companies […] it should contractually prohibit such entities from attempting to re-identify the data.	FTC report *Protecting Consumer Privacy in an Era of Rapid Change* (2012)	ensuring a chain of obligations if de-identified data is disclosed

QUESTIONS