Privacy-Enhancing Technologies
ovvero: come ho imparato a non preoccuparmi e amare la *data governance*

Avv. Filippo Bianchini

e-privacy XXXII

Roma, 15 giugno 2023

User license

Questo materiale è rilasciato sotto licenza:

Creative Commons Attribuzione - Non commerciale - Condividi allo stesso modo 3.0 Italia (CC BY-NC-SA 3.0 IT)

Alcune immagini della presentazione sono citazioni o “fair use” di opere protette da copyright dei legittimi proprietari.

Tutti i marchi citati appartengono ai legittimi proprietari.

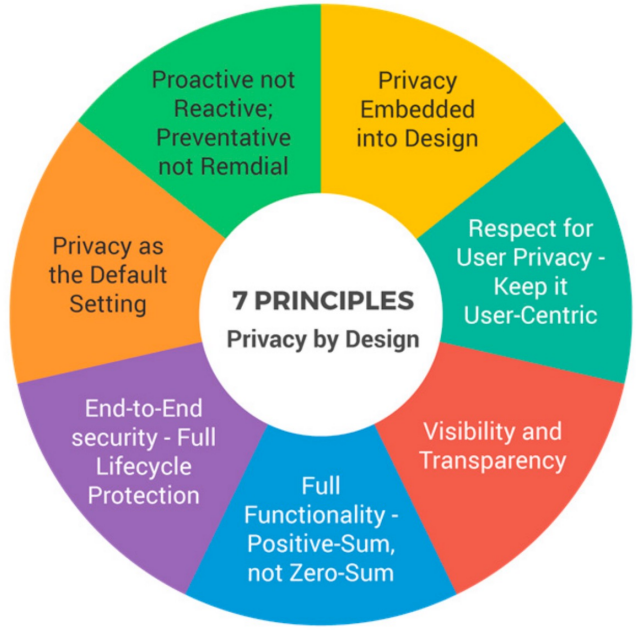
_>whoami

- Avvocato cassazionista, iscritto al Foro di Perugia
- DPO certificato UNI 11697:2017 – Lead Auditor 27001:2012 - CIPP/E
- Membro supplente Autorità Garante per la protezione dei dati personali di San Marino
- Membro del Comitato Direttivo di ASSO DPO
- Membro del *Cybersecurity National Lab* (nodo UniPG)
- Docente al Master universitario in “Data protection, Cybersecurity e Digital forensics” dell'Università per gli Studi di Perugia

In search of a definition

- «Privacy Enhancing Technologies: The Path to Anonymity» (SA Canada e Paesi Bassi,1995)
  - Varietà di tecnologie che salvaguardano la privacy personale, riducendo al minimo o eliminando la raccolta di dati identificabili
- «Inventory of Privacy-Enhancing Technologies» (OCSE, 2002)
  - Ampia gamma di tecnologie che aiutano a proteggere la privacy personale
- «The privacy infrastructure of tomorrow is being built today» (*Lunar Ventures*, 2021)
  - Insieme di metodi crittografici, architetture e flussi di lavoro per la scienza dei dati e sistemi di hardware e software che consentono alla parti avversarie di collaborare su dati sensibili senza dover fare affidamento sulla fiducia reciproca

PETs in the context of privacy regulation



- Principi fondamentali della Privacy by Design (PbD) – Ann Cavuokian, 2010
- Rapporto Federal Trade Commission (FTC) 2012
- Art. 25 Regolamento (UE) 2016/679 (GDPR) → misure tecniche e organizzative adeguate
- Art. 46 Lei Geral de Proteção de Dados Pessoais (LGPD), n° 13.709/2018 – Brasile

- Lo «**stato dell'arte**»: la migliore prestazione di una misura di sicurezza informatica disponibile sul mercato per raggiungere l'obiettivo di sicurezza (Rapporto ENISA-TeleTrust)

What are PETs?

- Le tecnologie di miglioramento della privacy sono tecnologie che garantiscono una maggiore privacy o segretezza per le persone i cui dati vengono elaborati, archiviati e/o raccolti da software e sistemi. Queste tecnologie sono spesso utilizzate come parte di questa elaborazione e modificano le normali modalità di gestione (e spesso di conservazione) dei dati grezzi o in chiaro provenienti direttamente dagli utenti e dai partecipanti interni, come i dipendenti. Aumentando la privacy offerta, si riduce il rischio di proprietà e si forniscono agli utenti scelte migliori su come desiderano che i loro dati siano trattati.

Why now?

- La proliferazione dei sistemi di apprendimento automatico, spesso addestrati utilizzando dati relativi alle persone, ha aumentato la superficie di minaccia per la privacy.
- Questi sistemi a volte producono rischi nuovi e sconosciuti, come dimostra l'espansione della ricerca su come estrarre informazioni private direttamente dai modelli stessi. Esiste anche una ricerca significativa su come l'IA generativa riproduca dati molto simili a quelli di addestramento.
- Fortunatamente, questi problemi stanno ricevendo una maggiore attenzione e c'è una crescente consapevolezza dei rischi a cui si va incontro.
- Le tecnologie per la privacy sono un modo per allineare le esigenze della scienza dei dati con quelle del consenso, della consapevolezza e della privacy degli utenti.

The good and the evil

- Vantaggi
  - Riduzione rischio identificazione delle persone
  - Condivisione, collegamento e analisi dati personali
  - Conformità legislativa a norme su protezione dati
- Svantaggi
  - Maturità non sviluppata
    - T*echnological Readiness Levels* (TRLs): classificano le PET in stadi di sviluppo distinti, da quello concettuale a quello market-ready
    - ENISA: integrazione TRL con scalabilità, fiducia, adattabilità
  - Mancanza di conoscenze sufficienti per la corretta applicazione
  - Errori nell'implementazione pratica rispetto alla teoria

Main kinds of PETs

A. PET che ricavano o generano dati che riducono o eliminano l'identificabilità delle persone
   - *Diffenential privacy, synthetic data*
B. PET per la protezione o la schermatura di dati
   - *Homomorphic encryption, zero-knowledge proofs*
C. PET che suddividono gli insiemi di dati o limitano l'accesso a particolari porzioni di dati
   - *Secure multi-party computation, federated learning*

- PET e **anonimizzazione** sono distinte ma collegate

Brief digression: k-anonymity and differential privacy

k-anonymity

[Samarati and Sweeney, 1998]

A k-anonymous dataset is achieved via suppression to make every combination of potentially identifying attributes appear at least k times

potentially identifying

ZIP	Age	sex	Disease
23456	55	Female	Heart
12345	30	Male	Heart
12346	33	Male	Heart
13144	45	Female	Breast Cancer
13155	42	Male	Hepatitis
23456	42	Male	Viral

2-anonymous

ZIP	Age	sex	Disease
23456	**	*	Heart
1234*	3*	Male	Heart
1234*	3*	Male	Heart
131**	4*	*	Breast Cancer
131**	4*	*	Hepatitis
23456	**	*	Viral

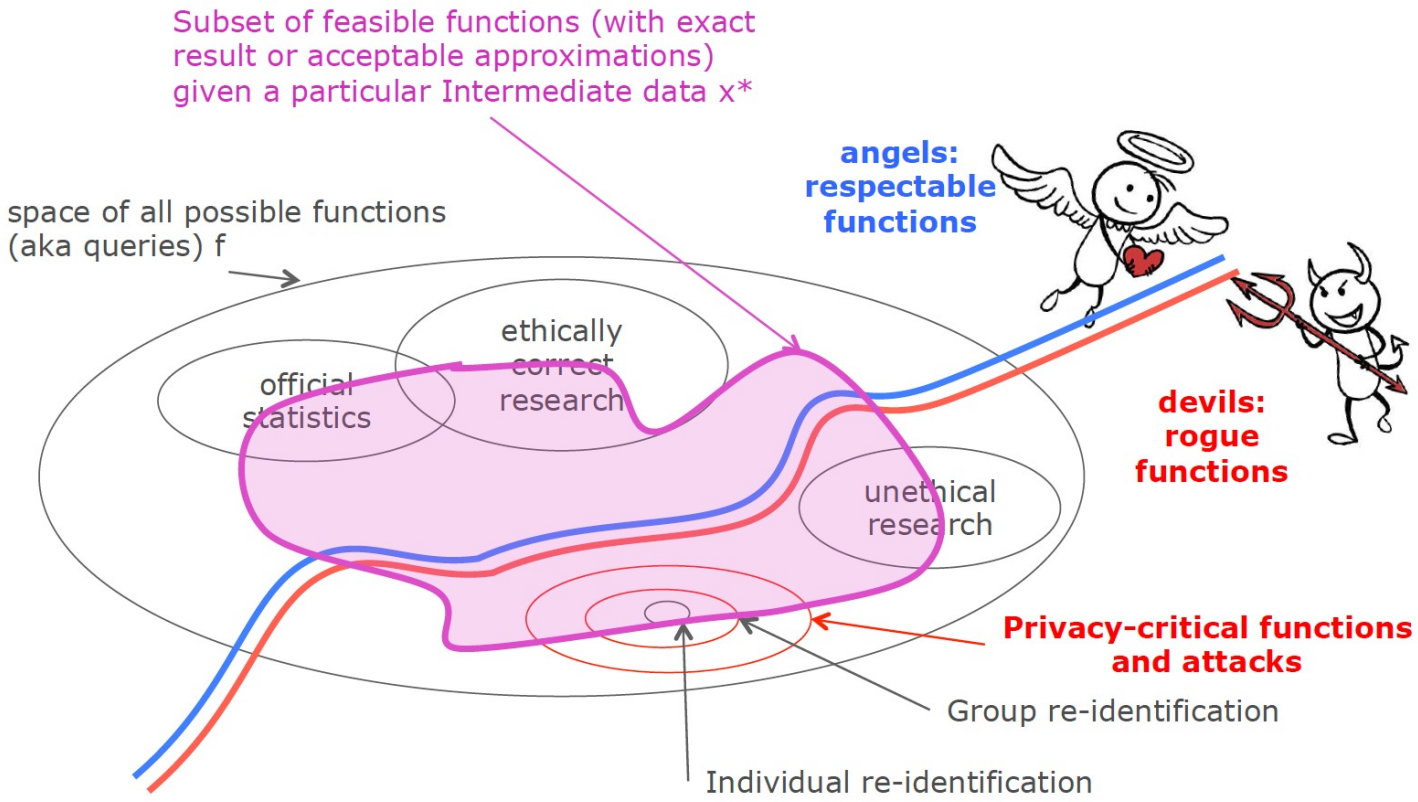
Homogeneity attack

Attacks on privacy

- Re-identification: identifying whose record it is after removal of identifying information [Sweeney 2000]
- Composition attacks: when the combining of two privacy mechanisms results in losing privacy [Ganta Kasiviswanathan smith 2008]
  - Can be applied to k-anonymized data [Cohen 2022]
- Database reconstruction: reconstructing almost the entire underlying dataset [Dinur N 2003]
  - Applied to 2010 Census and Diffix
- Membership inference: determining whether a target individual is in the dataset
  - Applied to genomic data and ML as service

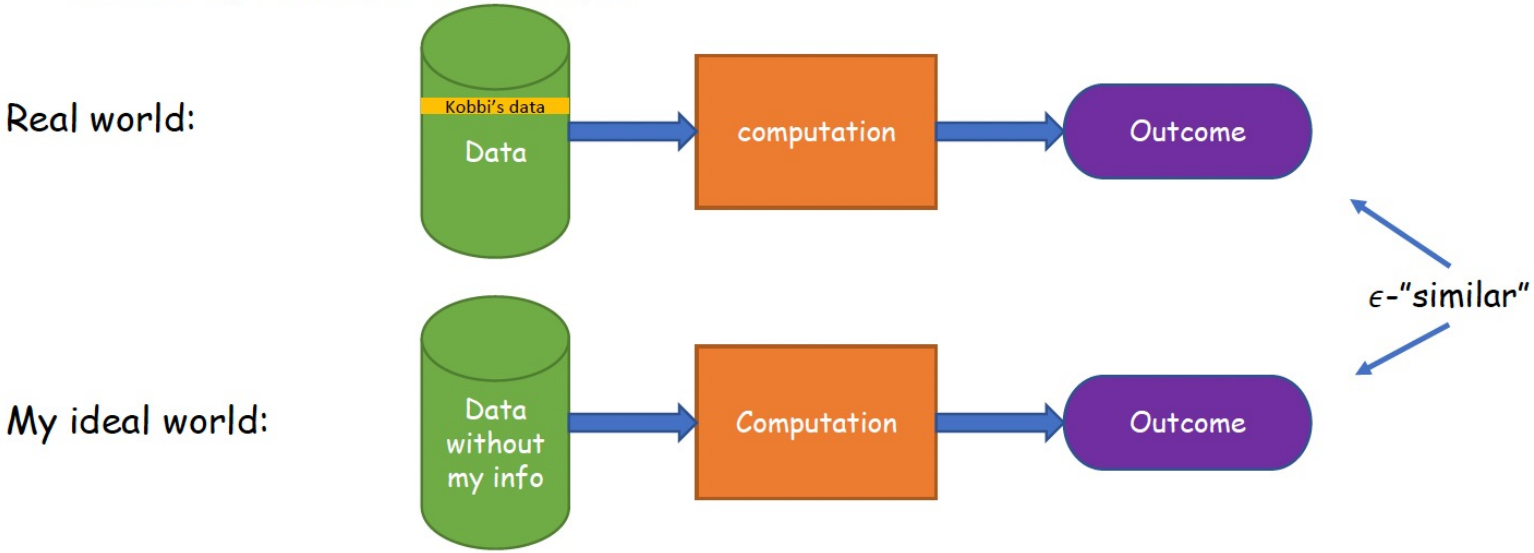
Attacks on "aggregate Statistics"

Setting the scene: angels & devils



The differential privacy *desiderata*
[Dwork and Roth, 2006]

- A computation is differentially private if any information-related risk to a person does not change significantly as a result of that person's information being included, or not, in the analysis



- Provably resilient to attacks from previous slide

Synthetic data

- Insiemi di dati creati «in provetta» per replicare le proprietà statistiche di dati reali senza divulgare le informazioni sensibili del set originario di rispondenti → monetizzazione asset di dati senza divulgazione info sensibili
  - Esempi: UK Office for National Statistics, Statistics Canada
- Vengono generati attraverso *Generative Adversarial Networks* (**GANs**), un tipo di AI in cui gli algoritmi vengono creati a coppie (uno per «imparare» e l'altro per «giudicare»). Utilizzate nell'apprendimento automatico non supervisionato, due reti neurali si sfidano l'una con l'altra per produrre simulazioni sempre migliori di dati reali (es.: creare volti di persone – https://thispersondoesnotexist.com - o grafie - https://www.calligrapher.ai)
- Pro
  - Miglioramento PbD in diverse fasi della progettazione e dello sviluppo di software e sistemi, come il debug, il test, la prototipazione e la convalida del sistema; miglioramento equità tramite attenuazione bias
- Contro
  - La fonte dei dati determina la qualità del modello; difficoltà di mappare gli outliers

Zero-Knowledge proofs

- Le ZKP sono metodi crittografici con i quali una parte (*prover*) può convincere un'altra (*verifier*) della verità di un'affermazione basata su informazioni note solo alla prima, senza rivelare segreti aggiuntivi
- Caratteristiche
  - **Completezza**: se l'affermazione è vera e le due parti seguono il protocollo, il *verifier* accetterà la prova fornita dal *prover*
  - **Solidità**: se l'affermazione è falsa e il *prover* segue il protocollo, il *verifier* non accetterà la prova
  - **Zero-knowledge**: se l'affermazione è vera e il *prover* segue il protocollo, il *verifier* non apprenderà alcuna informazione riservata dall'interazione con il *prover*
- Le ZKP possono essere utilizzati in contesi di verifica dell'identità, ad esempio per dimostrare che qualcuno ha più di una certa età senza rivelare la sua data di nascita esatta, e per le criptovalute, per evitare il *double-spending* (ZeroCash)
- I sistemi di prova *Succint Non-Interactive Argument* (SNArg) generano prove di dimensioni modeste e fisse, ma altri sistemi richiedo elevati costi computazionali

Digital Twin

- Consente di creare una copia digitale del soggetto fisico, costantemente aggiornata attraverso la raccolta continua di dati, per il monitoraggio a distanza dello stato di salute complessivo
- Evoluzione del FSE, che contiene una replica vivente del corpo del paziente aggiornato con i valori degli esami strumenti e l'impiego di dispositivi intracutanei e indossabili (dall'IoT all'IoHB)
- Benefici
  - Virtualizzazione dell'ospedale per creare un ambiente sicuro e a costi minori
  - Personalizzazione del trattamento medico → Percorsi Diagnostici Terapeutici Assistenziali (PDTA)
- Criticità
  - Medical Devices Hijacking (MDH)
  - Rischio di medicina oligarchica e di «digital divide» sanitario

Legal and regulatory issues

- Sfide e rischi
  - La conoscenza e la consapevolezza generale delle PET rimane bassa
  - La standardizzazione delle PET è carente. In futuro, gli standard specifici per le PET potrebbero costituire la base di schemi di garanzia per rafforzare la fiducia degli utenti.
- Opportunità e vantaggi
  - Conformità a PbD
  - Migliore rappresentazione della realtà rispetto all'anonimizzazione
- Ambiente normativo
  - Considerare lo stile della regolamentazione (necessità del consenso informato)
  - Considerare l'influenza su ulteriori diritti dell'interessato (DSAR)

Conclusions

- Le PET non sono un *silver bullet* ai problemi di protezione dei dati; tuttavia, possono essere in grado di fornire nuovi elementi per la costruzione di sistemi di *governance* dei dati responsabili.
- La privacy è solo un aspetto della protezione dei dati. Nella maggior parte dei casi, le PET si occupano di questo aspetto, ma non di come vengono utilizzati i dati o i risultati dell'analisi dei dati.
- Le tecnologie PET stanno già migliorando l'uso responsabile dei dati personali per affrontare importanti sfide contemporanee. Il ruolo emergente delle PET come strumento di partenariato, di rafforzamento della trasparenza e dell'*accountability* può portare a benefici ancora maggiori.

Q&A
Grazie per l'attenzione!

«[…] se non v'è dispiaciuta affatto, vogliatene bene a chi l'ha scritta, e anche un pochino a chi l'ha raccomodata. Ma se in vece fossimo riusciti ad annoiarvi, credete che non s'è fatto apposta.»

Avv. FILIPPO BIANCHINI

Via Bontempi, 1

06122 PERUGIA

(+39) 075 5723243 – (+39) 349 2864103

info@bianchini.legal

studiolegale

@legale

https://www.assodpo.it

Partecipa al Congresso!

https://www.assodpo.it/congresso-annuale-asso-dpo/