Linee guida 02/2026 sull'anonimizzazione

Versione 1.0

Adottato il 7 luglio 2026

Cronologia delle versioni

Versione	Data	Informazioni sull'adozione
versione 1.0	7 luglio 2026	adozione delle linee guida per la consultazione pubblica

Sintesi

Queste linee guida chiariscono il concetto di dati anonimi e forniscono un quadro pratico per determinare se i dati sono stati anonimizzati con successo. Queste linee guida contribuiranno quindi a garantire che i dati siano anonimizzati in modo sicuro, ove possibile e proporzionato, consentendo un utilizzo libero e proficuo dei dati, nel rispetto dei diritti e delle tutele delle persone fisiche.

Ai sensi del GDPR, i dati sono anonimi se non si riferiscono a una persona fisica identificata o identificabile. La sussistenza di tale condizione può variare da un'entità all'altra. Di conseguenza, l'anonimato deve essere valutato dal punto di vista di ciascuna entità interessata, in genere qualsiasi soggetto per il quale i dati sono destinati a essere anonimi.

Le informazioni possono riferirsi a una persona fisica in virtù del loro contenuto, scopo o effetto. L'esistenza di tale collegamento non deve necessariamente essere immediatamente evidente e potrebbe richiedere un certo trattamento per essere accertata. Una persona fisica è "identificata o identificabile" se può essere distinta da altre persone in un dato contesto mediante mezzi che è ragionevolmente probabile che vengano utilizzati e in un modo che consenta di trattarla in modo diverso. Il termine "mezzi" deve essere interpretato in senso ampio e può includere mezzi accessibili solo tramite terzi. La ragionevole probabilità che vengano utilizzati dipenderà dalla prospettiva dell'entità interessata e dovrà essere valutata alla luce di tutti i fattori oggettivi.

Questa analisi viene quindi integrata in un quadro di riferimento per la valutazione dell'anonimato. Tale quadro può essere applicato in due modi diversi: uno che considera le differenze di capacità tra coloro che potrebbero identificare l'interessato ("approccio contestuale") e uno che non le considera ("approccio semplificato"). L'approccio contestuale riflette tutte le sfumature dello standard legale in materia di anonimizzazione e consente al titolare del trattamento di valutare se i dati sono anonimi per ciascun soggetto interessato in base alle rispettive capacità. L'approccio semplificato, d'altro canto, può andare oltre lo standard legale e può indurre il titolare del trattamento a trattare i dati come se non fossero anonimi, anche se in realtà lo sarebbero per alcuni soggetti interessati. Tuttavia, offre anche un'opzione più comoda per i titolari del trattamento che scelgono di utilizzarla, fornisce una maggiore garanzia che i dati siano effettivamente anonimi e può essere combinato con l'approccio contestuale per affinare i risultati.

Il framework stesso presenta tre criteri che possono essere utilizzati per verificare se i dati sono anonimi: assenza di isolamento dei record, assenza di collegamento e assenza di inferenza. Questi criteri dovrebbero essere utilizzati per valutare l'efficacia di possibili tecniche di (re)identificazione. In generale, la (re)identificazione ha maggiori probabilità di successo su dati a livello di record con elevata dimensionalità e alta risoluzione, ma anche altri fattori sono importanti. Le tecniche dovrebbero essere valutate in base alla loro capacità di generare risultati accurati, in particolare se producono una risposta sufficientemente precisa e affidabile da consentire di distinguere il soggetto dei dati e trattarlo in modo diverso.

Se tutti e tre i criteri vengono soddisfatti, sia con l'approccio contestuale che con quello semplificato, i dati forniti possono essere considerati anonimi in tutta sicurezza. Se un qualsiasi criterio non viene soddisfatto, è necessario effettuare ulteriori analisi per determinare se i dati possano comunque essere considerati anonimi. In particolare, occorre verificare se eventuali record isolati, eventualmente insieme a dati collegati, consentano di identificare singoli individui.

Sommario

Il Comitato europeo per la protezione dei dati

Tenuto conto dell'articolo 70 (1)(e) dellaRegolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (di seguito "GDPR"),

Tenuto conto dell'accordo SEE e in particolare dell'allegato XI e del protocollo 37 dello stesso, come modificati dalla decisione n. 154/2018 del comitato misto SEE del 6 luglio 2018,

Tenuto conto dell'articolo 12 e dell'articolo 22 del suo regolamento

interno, **Ha adottato le seguenti linee guida:**

1 Introduzione

1 I dati vengono utilizzati per molteplici scopi in una vasta gamma di settori e finalità, molti dei quali sono estremamente vantaggiosi per un'altrettanto ampia gamma di soggetti interessati. In alcuni casi, ciò richiede l'utilizzo di dati personali e, in tali casi, il Regolamento generale sulla protezione dei dati (GDPR) garantisce che il trattamento dei dati avvenga in modo da avvantaggiare l'umanità e al contempo tutelare la sicurezza degli interessati. In altri casi, potrebbe essere preferibile utilizzare informazioni anonime,*cioè*informazioni che non si riferiscono a una persona fisica identificata o identificabile[1]I dati anonimi non sono soggetti alla regolamentazione del GDPR.[2]Ciò consente una maggiore libertà di utilizzo, ma significa anche che è estremamente importante garantire che i dati siano adeguatamente anonimizzati in modo che i diritti degli individui non siano messi a rischio.

2 Il parere n. 05/2014 del Gruppo di lavoro Articolo 29 sulle tecniche di anonimizzazione è stato pubblicato nel 2014 e forniva tre criteri per valutare l'anonimato dei dati. Da allora, tuttavia, si sono verificati importanti cambiamenti nel panorama giuridico, della privacy, ingegneristico e tecnologico. Lo sviluppo della giurisprudenza della Corte di giustizia dell'Unione europea (CGUE), l'istituzione di spazi dati a livello europeo e gli sviluppi sia nell'intelligenza artificiale che nella tecnologia in generale, tra le altre cose, hanno reso necessario aggiornare tale parere e perfezionare i criteri ivi presentati.[3] Queste linee guida forniscono tale aggiornamento, unitamente a ulteriori indicazioni sull'argomento.

3 Le linee guida iniziano con un'analisi giuridica dell'anonimizzazione e dei suoi requisiti ai sensi del GDPR. Questa analisi viene poi incorporata in un quadro di valutazione dell'anonimizzazione, affrontando i tre criteri oltre la loro formulazione originale nel precedente parere. Le linee guida definiscono anche due approcci per questa valutazione: uno che tiene conto delle differenze di capacità tra coloro che potrebbero identificare l'interessato ("l'approccio contestuale") e uno che, per semplicità, non lo fa ("l'approccio semplificato"). Quando si valuta se un dataset è anonimo, la domanda diventa: tale dataset soddisfa i tre criteri di assenza di isolamento dei record, assenza di collegamento e assenza di inferenza? Se tali criteri sono

[1]Considerando 26 del GDPR. Questa definizione rispecchia l'articolo 4, paragrafo 1, del GDPR, che definisce i dati personali come informazioni che si riferiscono a una persona fisica identificata o identificabile (ossia l'interessato). Lo stesso considerando afferma inoltre che le informazioni che inizialmente costituiscono dati personali possono successivamente diventare anonime se vengono trattate in modo tale che "l'interessato non sia o non sia più identificabile".
[2]Articolo 2(1) GDPR.
[3]Per maggiore chiarezza, un titolare del trattamento che abbia valutato un insieme di dati come anonimo in conformità al parere 05/2014 prima della pubblicazione delle presenti nuove linee guida non è tenuto a effettuare una nuova valutazione. Tuttavia, è buona prassi per qualsiasi soggetto rivalutare periodicamente la probabilità di reidentificazione dei dati anonimizzati che elabora.

Se uno di questi criteri viene soddisfatto, sia con l'approccio contestuale che con quello semplificato, i dati forniti possono essere considerati anonimi in tutta sicurezza. Qualora uno di questi criteri non venga soddisfatto, è necessario effettuare un'analisi più approfondita: i dati forniti potrebbero essere personali oppure potrebbero comunque essere considerati anonimi. Le presenti linee guida si concludono fornendo indicazioni su tale analisi aggiuntiva.

4 Gran parte del contenuto di queste linee guida si colloca nello spazio intermedio tra dati anonimi e dati personali: poiché le linee guida definiscono i criteri per stabilire se un set di dati rientra nell'una o nell'altra categoria, adottano una terminologia specifica per questo contesto. Non è possibile, ad esempio, stabilire se la persona che detiene i dati debba essere definita titolare del trattamento o meno, poiché tale risposta dipende dalla natura personale o meno dei dati in esame. Pertanto, anziché fare riferimento costantemente a "la persona che può o non può essere titolare del trattamento", queste linee guida utilizzano una serie di termini chiave per facilitarne la lettura, tutti descritti in dettaglio in un glossario alla fine del presente documento.

5 In particolare, le linee guida usano il termine "individuo" per riferirsi a una persona fisica che può essere o meno un soggetto interessato e "dati forniti" per riferirsi ai dati che vengono testati e che possono essere o meno dati personali. Le linee guida usano poi la parola "record" per riferirsi a tutti i valori (incluse le immagini)*ecc.*.) all'interno dei dati forniti che si sa essere relativi allo stesso individuo. Usano anche il termine "entità" per riferirsi a una parte il cui ruolo dipende dalla situazione specifica. Ad esempio, un'entità potrebbe essere un potenziale titolare del trattamento, un potenziale responsabile del trattamento o una potenziale terza parte in possesso di informazioni aggiuntive necessarie per identificare l'interessato. Ogni entità ha quindi una "prospettiva", che dipende dalle circostanze del caso.

2 Analisi giuridica

2.1 Introduzione

6 Il test fondamentale per verificare l'anonimato prevede due domande:

a. Le informazioni “riguardano” una persona fisica? In caso affermativo,

b. Si tratta di una persona fisica “identificata o identificabile”?

Se la risposta a una di queste domande è "no", i dati devono essere considerati anonimi. È importante sottolineare che queste risposte possono variare da un'entità all'altra. Ciò significa che le informazioni possono essere anonime per alcune entità, ma non per altre.[4].

7 Ad esempio, le informazioni possono riferirsi in modo inequivocabile a un individuo, ma solo un'organizzazione può essere in grado di identificare effettivamente tale individuo. In questo caso, la risposta a entrambe le domande sarebbe "sì" dal punto di vista di quell'organizzazione, e quindi le informazioni sarebbero considerate dati personali per quell'organizzazione. Allo stesso tempo, la risposta alla seconda domanda sarebbe – al di fuori di particolari circostanze che verranno spiegate in seguito[5]– sarebbe “no” dal punto di vista di tutti gli altri, e quindi le informazioni sarebbero considerate anonime per tutti gli altri. È quindi necessario considerare le diverse prospettive dei soggetti interessati quando si esegue questo test. Ciò significa che l'anonimizzazione può essere effettuata in modo che i dati siano anonimi per tutti, oppure in modo che siano anonimi solo per determinati soggetti (che possono includere o meno il titolare del trattamento che effettua l'anonimizzazione).

[4] C-413/23 P *EPA contro SRB.*

[5] Vedere, *per esempio*, paragrafi 20, 21 e 29.

8 In molti casi, la seconda domanda riceverà una risposta qualificata ("L'individuo potrebbe essere identificabile, ma solo se..."). In particolare, potrebbero esserci diversi modi in cui un'entità potrebbe identificare un individuo in teoria, ma in cui la probabilità che ciò accada è insignificante nella pratica. Questa domanda deve quindi essere considerata in termini di probabilità: anziché chiedere se l'individuo è identificato o identificabile in senso assoluto, dovrebbe chiedere quanto è probabile che l'individuo venga identificato o sia identificabile da qualche entità.[6].

9 Occorre inoltre riconoscere che queste domande presuppongono che l'entità in questione abbia, o possa avere, un qualche tipo di accesso o controllo sui dati. Se, tenendo conto dei mezzi che l'entità è ragionevolmente probabile che utilizzi, un'entità non può effettivamente accedere o elaborare i dati, non è significativamente collegata ad altre entità (o catene di entità) che abbiano tale accesso e non è probabile che riceva i dati da tali altre entità[7], quindi non dovrebbe essere necessario valutare se le informazioni siano anonime per tale entità.

10 Infine, le presenti linee guida si riferiscono al concetto di "anonimizzazione" in senso ampio, intendendolo come qualsiasi trattamento che produca con successo dati anonimi ai sensi del GDPR, indipendentemente dal fatto che i dati originali vengano cancellati o meno. Ciò include l'anonimizzazione che non comporta la cancellazione dei dati originali e che quindi non contribuisce alla limitazione della conservazione. Le presenti linee guida non forniscono pertanto indicazioni per l'interpretazione di specifiche disposizioni di legge esistenti che potrebbero imporre la cancellazione dei dati originali.

2.2 Individuazione della/e prospettiva/e applicabile/i

11 Come già accennato, la natura personale di un'informazione può variare da un'entità all'altra. È pertanto importante individuare le entità rilevanti per la valutazione dei dati in questione. In genere, la natura personale di un'informazione per una determinata entità viene valutata in base alla sua prospettiva, sebbene, come spiegato di seguito, in alcuni casi sia opportuno adottare la prospettiva di un'altra entità. L'obiettivo dell'anonimizzazione è garantire che i dati risultanti siano anonimi per tutte le entità coinvolte nella valutazione.

12 A volte può essere immediatamente evidente se una determinata entità è rilevante o meno. La domanda fondamentale è: per chi si intende rendere anonimi i dati? Ad esempio, se l'intenzione è di rendere anonimi i dati per uso proprio di un'entità, la prospettiva di tale entità sarà rilevante per la valutazione. Allo stesso modo, se l'intenzione è di rendere anonimi i dati per un uso indipendente da parte di chiunque li riceva, l'anonimato dovrebbe generalmente essere valutato dalla prospettiva di coloro che avranno accesso ai dati dopo il loro trasferimento. In alcuni casi, potrebbe esserci solo una prospettiva rilevante (*per esempio*, se i dati sono conservati in modo sicuro da un'unica entità e nessun altro potrà mai accedervi) mentre, in altri, possono esserci molte prospettive rilevanti diverse (*per esempio*(se i dati fossero accessibili da più entità).

Esempio 1: In*OC contro Commissione*[8], la CGUE ha esaminato se un comunicato stampa contenesse dati personali. La Corte ha osservato che il comunicato stampa era, per sua stessa natura, destinato a

[6]V., ad esempio, C-413/23 P EDPS/SRB e C-582/14 Breyer/Bundesrepublik Deutschland.

[7]Ciò può verificarsi, ad esempio, in caso di collusione o cooperazione tra tali entità, oppure quando è probabile che le informazioni possano essere trasferite (direttamente o indirettamente) da un'entità all'altra. Si veda, ad esempio, C-582/14.*Breyer contro Repubblica federale di Germania*e C-319/22*Gesamtverband Autoteile-Handel eV contro Scania CV AB*. Vedere anche C-210/16 *Centro regionale per la protezione dei dati dello Schleswig-Holstein contro Wirtschaftsakademie Schleswig-Holstein GmbH*, dove la Corte di giustizia dell'UE ha stabilito che una parte che determina le finalità e i mezzi del trattamento dei dati personali può comunque essere considerata contitolare del trattamento ed è comunque soggetta alla normativa europea in materia di protezione dei dati, anche se non ha accesso a tali dati.

[8]C-479/22 P OC contro Commissione.

La pubblicazione è stata distribuita ai giornalisti, compresi i giornalisti investigativi con competenze di ricerca superiori a quelle di una persona media. Pertanto, nel valutare se la pubblicazione contenesse dati personali, è stato necessario effettuare la valutazione almeno dal punto di vista di questi giornalisti investigativi.

13 In altri casi, diversi fattori possono essere importanti per identificare le prospettive rilevanti. In particolare, è importante esaminare la natura e il contesto specifici del trattamento, come ad esempio chi ha accesso ai dati o ne ha il controllo, se i dati vengono trasmessi da una parte all'altra, la relazione tra tali parti e se un destinatario dei dati potrebbe elaborarli per conto di terzi.

14 La prospettiva rilevante[9] può essere influenzato da specifici obblighi del GDPR e può influenzare il modo in cui un determinato obbligo viene applicato nella pratica, come si è visto in *EPA contro SRB*.

Esempio 2: In *EPA contro SRB* La Corte di giustizia dell'Unione europea ha esaminato l'obbligo di fornire informazioni quando i dati personali vengono raccolti direttamente dagli interessati. Tale obbligo richiede, tra l'altro, che gli interessati siano informati sui destinatari di eventuali trasferimenti di dati personali. L'Autorità garante per i diritti degli Stati Uniti (SRB) ha sostenuto che, nel decidere se tale obbligo fosse applicabile, fosse importante considerare la prospettiva dell'entità ricevente. Seguendo il ragionamento dell'SRB, se le informazioni fossero anonime dal punto di vista dell'entità ricevente, la condivisione dei dati con tale entità non si configurerebbe come un trasferimento di dati personali. *per sé* e tale obbligo non si applicherebbe.

La Corte di giustizia dell'Unione europea, tuttavia, ha respinto tale argomentazione. La Corte ha invece osservato che tale obbligo deve essere adempiuto al momento della raccolta dei dati e che esso costituisce parte integrante del rapporto giuridico tra l'interessato e il titolare del trattamento. La Corte ha stabilito che, nel valutare l'applicabilità di tale obbligo, la natura personale dei dati deve essere considerata unicamente dal punto di vista del titolare del trattamento al momento della raccolta. Poiché, in quel momento, i dati erano personali per il titolare del trattamento, qualsiasi trasferimento di tali informazioni si configura come trasferimento di dati personali, indipendentemente dal punto di vista del destinatario. Di conseguenza, il titolare del trattamento è tenuto a informare gli interessati in merito ai destinatari dei dati.[10].

15 È importante notare che la prospettiva applicabile utilizzata per una particolare entità dipende dal fatto che essa essi stessi determinano i mezzi e le finalità del trattamento[11]. Se un'entità elabora informazioni per conto di un'altra, la valutazione della natura personale dei dati forniti per tale entità deve essere effettuata in riferimento alla prospettiva dell'entità che controlla. In pratica, ciò significa che se un'entità elabora informazioni per conto di un titolare del trattamento (*cioè* Se una persona per la quale le informazioni sono dati personali e che decide le finalità e i mezzi del trattamento è responsabile del trattamento, tali informazioni devono essere considerate dati personali anche per il soggetto che effettua il trattamento. Tale soggetto deve pertanto essere considerato responsabile del trattamento ai sensi del GDPR ed essere soggetto agli obblighi applicabili ai responsabili del trattamento.[12] Ciò deriva da un'applicazione teleologica del GDPR, tenendo presente che:

- La normativa UE in materia di protezione dei dati si basa sul concetto di responsabile del trattamento per impedire ai titolari del trattamento di eludere le tutele del GDPR esternalizzando le proprie attività di trattamento a terzi,

[9] Consulta la definizione del termine nel Glossario.
[10] C-413/23 P *EPA contro SRB*, punti 102-103 e 108-111.
[11] Per ulteriori chiarimenti sul concetto di trattamento dei dati per conto di terzi, si vedano le Linee guida 07/2020 dell'EDPB sui concetti di titolare e responsabile del trattamento nel GDPR, paragrafi 79-84.
[12] Ciò include, in particolare, il rispetto degli articoli 27, 28, 29, 30(2), 31, 32, 33(2), 37, 38, 44, 46, 48 e 49(6) del GDPR e di qualsiasi ordine emesso dalle autorità di controllo ai sensi dell'articolo 58 del GDPR.

- il GDPR rafforza ulteriormente questo principio imponendo determinati obblighi direttamente ai responsabili del trattamento e
- Sia gli accordi contrattuali (o altri atti giuridici applicabili) sia gli obblighi previsti dal GDPR devono operare in sinergia per garantire che il responsabile del trattamento gestisca le informazioni in modo appropriato.

Esempio 3: Un rivenditore di e-commerce trasmette estratti dei propri dati clienti a un'agenzia terza per l'analisi. In questo caso, il rivenditore determina le finalità e le modalità del trattamento, mentre l'agenzia è un'entità separata che elabora i dati per suo conto. È quindi necessario adottare la prospettiva del rivenditore per entrambe le parti quando si determina se le informazioni sono da considerarsi personali. In questo caso, il rivenditore è in grado di identificare i clienti da tali dati e le informazioni devono pertanto essere trattate come dati personali sia per il rivenditore (in qualità di titolare del trattamento) sia per l'agenzia (in qualità di responsabile del trattamento).

Esempio 4: Un ospedale trasmette estratti delle cartelle cliniche dei pazienti, che di per sé non consentono l'identificazione degli individui, a un istituto indipendente a fini di ricerca. I dati non saranno mai restituiti all'ospedale e l'istituto è libero di determinare le modalità e le finalità del trattamento, senza agire su istruzioni dell'ospedale. Poiché agisce in modo indipendente, la questione se le informazioni siano da considerarsi personali per l'istituto di ricerca deve essere valutata dal suo punto di vista.

2.3 Se le informazioni “riguardano” una persona fisica

16 Un'informazione si riferisce a una persona fisica quando, per il suo contenuto, scopo o effetto, è collegata a tale persona. Il collegamento deve essere valutato alla luce di tutte le circostanze del caso specifico:[13]

UN.**Contenuto.**Le informazioni riguardano quella persona*per sé*,*per esempio*Descrive una o più delle loro caratteristiche o azioni. Ad esempio, potrebbe trattarsi del profilo di un dipendente, della cartella clinica di un paziente o di una foto o un video di una persona.

B.**Scopo.**Le informazioni riguardano o quella persona, o riguardano principalmente qualcosa o qualcun altro che è direttamente o indirettamente correlato a quella persona a causa di una qualche relazione o interazione, e tali informazioni sono utilizzate o potrebbero essere utilizzate allo scopo di valutare, trattare o influenzare lo status o il comportamento dell'individuo in un modo particolare. Ciò può includere informazioni su un oggetto (come un telefono, un'auto o una casa di proprietà della persona), un'entità organizzativa (come un'azienda, un dipartimento, un'associazione o un club di cui la persona è membro) o un associato o un parente della persona fisica (come un amico o un familiare). Ad esempio, potrebbero essere dati sul valore di una casa utilizzati allo scopo di determinare le tasse del proprietario, o il registro di manutenzione di un'auto che potrebbe essere utilizzato allo scopo di accertare la produttività del meccanico responsabile.

C.**Effetto.**Le informazioni riguardano principalmente un oggetto (come un telefono, un'auto o una casa), un'organizzazione (come un'azienda, un dipartimento, un'associazione o un club) o un'altra persona, ma

[13]C-434/16*Nowak contro il Commissario per la protezione dei dati*, paragrafo 35 e C-413/23 P*EPA contro SRB*, paragrafi 55 e 56. Si veda anche il parere n. 4/2007 del Gruppo di lavoro sulla protezione dei dati dell'articolo 29, adottato il 20 giugno 2007, pag. 10, sebbene tale parere utilizzi il termine “risultato” anziché “effetto”.

Il trattamento di tali dati potrebbe incidere sui diritti e sugli interessi dell'interessato. Affinché sussista tale nesso causale, deve esserci una concreta possibilità che, a seguito del trattamento, l'individuo possa essere trattato in modo diverso da altre persone. Ad esempio, si potrebbero raccogliere i dati di geolocalizzazione dei taxi al fine di ottimizzare gli itinerari, ma che consentono anche di monitorare le prestazioni dei tassisti e di individuare i momenti di pausa.

17 Le informazioni possono riferirsi a una persona anche se non è immediatamente evidente, o se è necessaria un'ulteriore elaborazione per rivelare tale collegamento. Ciò è particolarmente vero per i dati aggregati che, a prima vista, contengono informazioni solo su gruppi di persone, non su singoli individui. Ad esempio, esistono molte tecniche in grado di estrarre informazioni relative a uno o più singoli individui da dati aggregati. Quando una di queste tecniche può essere utilizzata, i dati aggregati possono riferirsi a quell'individuo (e a qualsiasi altro individuo per il quale tale tecnica potrebbe essere utilizzata).

2.4 Se la persona fisica è “identificata o identificabile”

18 Non è sufficiente che le informazioni si riferiscano semplicemente a una persona fisica. Affinché queste informazioni siano dati personali, tale persona fisica deve anche essere identificata o identificabile utilizzando mezzi ragionevolmente probabili da utilizzare. Ciò è particolarmente importante quando le informazioni vengono anonimizzate per scopi di ricerca o di analisi, dove l'obiettivo è spesso quello di avere un insieme di dati anonimi che consentano di apprendere qualcosa sulle persone in questione.*cioè*informazioni che sono ancora correlate a tali individui) ma lo fa in modo da mantenere private le loro identità (*cioè*non sono identificati né identificabili).

19 Identificare una persona fisica significa distinguerla dalle altre all'interno di un dato contesto, rendendo di conseguenza possibile trattarla in modo diverso rispetto a queste ultime. Ciò può avvenire tramite identificatori diretti (come nome, numero di identificazione o codice di riferimento) o tramite uno o più attributi relativi a tale persona.

20 Un'entità può identificare un individuo direttamente o indirettamente. Un individuo è identificabile "direttamente" da una particolare entità se quest'ultima può, attraverso i propri mezzi, identificare l'individuo sulla base delle informazioni stesse. Un individuo è identificabile "indirettamente" dal punto di vista di una particolare entità se, attraverso mezzi ragionevolmente utilizzabili, tale entità potrebbe identificare l'individuo tramite:

- ottenere informazioni aggiuntive[14],
- applicando metodi (come la decrittazione di parti crittografate dei dati) controllati da terzi e/o
- mettere i dati a disposizione di altre entità che a loro volta dispongono di mezzi che ragionevolmente possono essere utilizzati per identificare l'individuo[15].

Di conseguenza, un individuo può ancora essere identificato anche se alcune delle informazioni necessarie – o alcuni dei mezzi di identificazione – sono distribuiti tra diverse entità e richiedono uno sforzo per essere ottenuti, a condizione che questi elementi possano essere combinati attraverso mezzi che è ragionevolmente probabile che vengano utilizzati dalle rispettive entità[16].

[14] C-604/22 IAB Europe contro Gegevensbeschermingsautoriteit, punto 39.
[15] C-319/22 Gesamtverband Autoteile-Handel eV contro Scania CV AB, punti 45 – 49.
[16] Considerando 26 GDPR. Cfr. anche C-582/14 Breyer contro Bundesrepublik Deutschland, punto 43 e C-319/22 Gesamtverband Autoteile-Handel eV contro Scania CV AB, punti 45 – 49.

21 A tale riguardo, l'EDPB sottolinea che se (a) è ragionevolmente probabile che un'entità possa trasferire i dati a terzi e (b) non si può escludere che tale entità ricevente abbia i mezzi per identificare l'individuo attraverso mezzi che è ragionevolmente probabile che vengano utilizzati, i dati devono essere considerati personali sia per tale trasferimento sia per qualsiasi trattamento successivo dei dati da parte di tale destinatario.[17]Inoltre, questo trasferimento implica che le informazioni trasferite diventerebbero indirettamente personali anche per l'entità che le trasferisce, anche se tale entità non disporrebbe altrimenti di mezzi che potrebbero ragionevolmente essere utilizzati per identificare le persone in questione.[18].

22 È importante sottolineare che, affinché le informazioni siano anonime, la probabilità che l'individuo venga distinto con successo dagli altri all'interno del contesto dato non deve essere pari a zero; al contrario, tale probabilità dovrebbe essere insignificante nella realtà.[19]Questa probabilità può variare notevolmente da una prospettiva all'altra e il fatto che un individuo sia stato, o potrebbe essere, identificato da una prospettiva non significa necessariamente che lo stesso valga per un'altra.

2.4.1 Identificatori e altri attributi

23 In molti casi, l'identificazione avviene tramite un identificativo univoco all'interno di un dato contesto. Nei casi più semplici, questo potrebbe consistere, ad esempio, in un singolo dato univoco, come un nome, un indirizzo IP, un codice fiscale o un indirizzo e-mail.

24 In altri casi, l'identificazione può basarsi su uno o più attributi che, considerati insieme, risultano unici per un determinato individuo, anche se ciascuno dei singoli valori non sarebbe di per sé unico. La tipologia di attributi coinvolti può essere estremamente ampia e può includere, tra le altre cose, attributi fisici, genetici, psicologici, comportamentali, economici o culturali. Ad esempio, una persona potrebbe essere identificata come "quella con l'accento del nord" o "quella che vive sottoterra".

Esempio 5: Osservando le immagini in diretta delle telecamere di sorveglianza che riprendono le persone in strada, si nota che molte indossano abiti neri e molte portano a spasso il cane. Nel contesto di quella strada, "la persona in abito nero" non è quindi un elemento univoco e tale descrizione non permette di identificare un individuo specifico. Lo stesso vale per "la persona con il cane". Tuttavia, in quella strada c'è solo un individuo che indossa un abito nero e porta a spasso il cane. Nel contesto di quella strada, la combinazione di questi attributi è quindi unica per quell'individuo e la descrizione "la persona in abito nero che porta a spasso il cane" permette di identificarlo.

25 È importante sottolineare che gli identificatori o gli attributi non devono necessariamente riguardare l'individuo*per sé*ma può anche riguardare altri individui, gruppi o oggetti che sono altrimenti collegati ad essi. Ad esempio, il dispositivo di un utente può avere un identificatore hardware, come un indirizzo MAC. – che può identificare il dispositivo e, a sua volta, l'individuo che utilizza il dispositivo. Allo stesso modo, un sito web può "improntare" i suoi visitatori in base alla loro combinazione unica di browser web, sistema operativo, risoluzione dello schermo, fuso orario*ecc.*e utilizzare questa impronta digitale per identificarli mentre si spostano da una pagina all'altra e nel corso di diverse visite.

[17]C-413/23 P*EPA contro SRB*, paragrafo 85.
[18]C-413/23 P*EPA contro SRB*, punto 84, in riferimento a C-319/22*Gesamtverband Autoteile-Handel eV contro Scania CV AB*, paragrafi 45 e 49.

[19]C-413/23 P*EPA contro SRB*, paragrafo 82.

Esempio 6: Diversi siti web utilizzano lo stesso fornitore terzo per la pubblicità mirata. I siti web non condividono nomi, indirizzi IP o email.*ecc.*...con terze parti; invece, forniscono informazioni sulle pagine visitate insieme a uno pseudonimo basato sulle impronte digitali dei dispositivi degli utenti. La terza parte utilizza quindi questi pseudonimi per collegare i dati degli utenti tra i vari siti, profilare tali utenti e decidere quali annunci pubblicitari mostrare loro.

In questa situazione, sia i dati grezzi delle impronte digitali (come combinazione di attributi) sia lo pseudonimo derivato da tali dati (come identificativo) consentono l'identificazione dell'individuo. Ciò vale sia per i siti web che per terze parti: anche se non interagiscono direttamente con essi, le terze parti possono (e di fatto lo fanno) utilizzare gli pseudonimi per distinguere l'individuo e trattarlo in modo diverso dagli altri, rendendo quindi tali individui identificati o identificabili dal loro punto di vista.

2.4.2 Mezzi che è ragionevolmente probabile che vengano utilizzati

26 Piuttosto che descrivere quali mezzi è ragionevolmente probabile che vengano utilizzati*per sé*La Corte di giustizia dell'Unione europea ha definito questo concetto facendo riferimento a elementi che impedirebbero che un determinato soggetto utilizzi ragionevolmente i mezzi. In particolare, non è ragionevolmente probabile che un mezzo venga utilizzato se la probabilità di identificazione appare, in realtà, insignificante perché sarebbe impossibile da realizzare; perché comporterebbe uno sforzo sproporzionato in termini di tempo, costi e manodopera; o a causa di un divieto legale.[20]Questa sottosezione fornisce ulteriori indicazioni su questo concetto e su come valutare se un mezzo possa essere considerato ragionevolmente probabile che venga utilizzato oppure no.

27 Come punto di partenza, il test “mezzi ragionevolmente probabili da utilizzare” si riferisce solo all’identificazione possibile e futura (*cioè*L'individuo è "identificabile"? Se l'individuo a cui si riferiscono le informazioni è già stato identificato, allora le informazioni non sono anonime.

28 Il concetto di “mezzi” dovrebbe essere inteso in senso ampio, comprendendo qualsiasi cosa, dalla semplice lettura di un documento all'effettuazione di una ricerca sul web, fino all'esecuzione di analisi e inferenze incredibilmente complesse. Inoltre, i mezzi di un'entità possono includere anche lo sfruttamento di mezzi disponibili a un'altra entità; laddove ciò sia possibile, è importante valutare la probabilità che questa “catena” complessiva di mezzi[21]si unisce.

Esempio 7: Un titolare del trattamento desidera condividere determinate informazioni con il proprio partner commerciale. Cerca di anonimizzare i dati, ma, durante la verifica dell'efficacia dell'anonimizzazione, si rende subito conto che gli interessati potrebbero essere reidentificati tramite una specifica tecnica di reidentificazione. Il titolare del trattamento sa che il partner commerciale non sarebbe in grado di eseguire direttamente tale tecnica. Tuttavia, potrebbe facilmente incaricare una terza parte di farlo. Pertanto, tale tecnica si qualifica comunque come un mezzo ragionevolmente probabile che venga utilizzato dal partner commerciale.

29 La probabilità che un'entità utilizzi determinati mezzi per identificare un individuo deve essere valutata alla luce di tutti i fattori oggettivi che potrebbero influenzare la disponibilità di tali mezzi e la probabilità del loro effettivo utilizzo, tra cui:

[20]C-413/23 P*EPA contro SRB*, paragrafi 82-83.

[21]La catena dei mezzi comprende qualsiasi trasferimento dei dati in questione a (o accesso da parte di) un soggetto coinvolto nel rispettivo processo di identificazione. La probabilità di tali trasferimenti (o accessi) deve essere valutata a partire dalla ricezione dei dati da parte di un soggetto per il quale si presume che essi rimangano anonimi.

a. le proprietà dei dati stessi – tra cui se i dati sono aggregati, l'unicità dei record, nonché l'accuratezza e la precisione delle informazioni – e come queste influenzano la vulnerabilità dei dati ai diversi possibili mezzi di (re)identificazione;

b. il contesto in cui i dati vengono rilasciati e/o elaborati[22], compresa la verifica dell'esistenza di eventuali restrizioni effettive e permanenti all'accesso ai dati;
c. qualsiasi informazione aggiuntiva che permetta l'identificazione degli individui e la probabilità che tali informazioni possano essere ottenute;
d. i costi e il tempo necessario all'entità per ottenere tali informazioni aggiuntive (nel caso in cui non siano già a sua disposizione); e
e. la tecnologia disponibile al momento del trattamento, nonché gli sviluppi tecnologici ragionevolmente prevedibili.

In particolare, se la legge prevede mezzi per accedere a informazioni aggiuntive, è oggettivamente chiaro che tale accesso costituisce un mezzo che è ragionevolmente probabile venga utilizzato[23].

30 Il GDPR non stabilisce alcuna condizione in merito alle entità che sono in grado di identificare l'individuo[24]Le entità rilevanti possono quindi includere chiunque riceva i dati direttamente o indirettamente, compresi soggetti non autorizzati e malintenzionati. A seconda delle circostanze, queste possono includere (tra gli altri):[25]:

- Il titolare del trattamento dei dati che si intendono anonimizzare;
- Qualsiasi ente ricevente;
- Dipendenti infedeli;
- I partner, gli amici, i colleghi, i vicini dell'individuo,*ecc.*Questi soggetti avranno spesso un accesso limitato ai dati forniti (soprattutto se vengono implementate misure di sicurezza adeguate), ma potrebbero avere accesso a informazioni aggiuntive che possono quindi essere utilizzate per (re)identificare l'individuo o gli individui presenti nei dati in questione;
- Giornalisti investigativi. Ciò è particolarmente rilevante per i dati relativi a individui che sono, o si prevede che saranno presto, sotto i riflettori;
- Forze dell'ordine o agenzie di intelligence nazionali. Questi enti dispongono spesso di ampi poteri legali per indagare e raccogliere dati, nonché di un ampio accesso a risorse finanziarie, elevata potenza di calcolo ed esperti tecnici;
- Agenzie di intelligence straniere. Le stesse considerazioni che si applicano alle forze dell'ordine nazionali si applicano anche alle agenzie di intelligence straniere. Inoltre, le agenzie di intelligence straniere potrebbero non essere vincolate dalle stesse leggi che limitano le azioni delle forze dell'ordine nazionali;
- Aziende non etiche. Queste entità possono avere accesso a grandi insiemi di dati e possono sfruttare lacune legali o zone grigie per ottenere l'accesso a dati che altrimenti sarebbero inaccessibili; e
- Criminali informatici. Queste entità possono semplicemente violare la legge per ottenere l'accesso a dati non disponibili da parte di soggetti che rispettano la legge.

Diverse entità possono essere rilevanti in modi diversi. Ad esempio, gli amici e i familiari di un individuo potrebbero essere in grado di fornire informazioni aggiuntive che potrebbero aiutare la reidentificazione, ma

[22]Si veda il paragrafo 148 della decisione vincolante 1/2021: "[L]'EDPB sottolinea che è necessario considerare l'intero contesto del trattamento, poiché 'tutti i fattori oggettivi' incidono sulla 'ragionevole probabilità che vengano utilizzati mezzi per identificare la persona fisica'".
[23]Ciò include,*per esempio*, la capacità di un ente privato di ottenere informazioni aggiuntive attraverso canali legali, come indicato nel documento C-582/14 *Breyer contro Repubblica federale di Germania*, paragrafi 47 e 49.
[24]C-479/22 P*OC contro Commissione europea*, paragrafo 56.
[25]Non tutte queste prospettive devono essere considerate sistematicamente in ogni caso, poiché ciò dipende dalle circostanze specifiche di ciascun caso e dagli enti che ricevono i dati, direttamente o indirettamente.

incapaci di accedere ai dati forniti o di eseguire effettivamente la necessaria tecnica di reidentificazione. Queste diverse capacità possono, pertanto, influenzare i mezzi che ciascuna entità potrebbe avere a disposizione, il modo in cui ciascuna entità potrebbe inserirsi in una possibile catena di mezzi e se tali mezzi (o catene di mezzi) siano effettivamente ragionevolmente probabili da utilizzare.

31 Il Comitato europeo per la protezione dei dati (EDPB) mette in guardia dall'utilizzare la "mancanza di motivazione" come fattore in questa analisi. Innanzitutto, la motivazione degli individui può essere difficile da valutare o dimostrare oggettivamente e può cambiare nel tempo. Inoltre, le azioni di un'entità potrebbero non essere coerenti con le sue motivazioni. Ad esempio, l'identificazione potrebbe avvenire a causa di un incidente o di negligenza. Ciò potrebbe accadere anche se un'entità è costretta da circostanze esterne a identificare un individuo o a trasferire informazioni che consentano tale identificazione.[26]. Nel valutare i mezzi che è ragionevolmente probabile che vengano utilizzati, il potenziale valore dei dati reidentificati per l'entità che effettua la reidentificazione (che non si limita al valore monetario) [27]Tale aspetto dovrebbe essere preso in considerazione nella misura in cui compensa i costi e gli sforzi necessari per la reidentificazione. Allo stesso modo, i rischi che la reidentificazione comporta per l'entità che la effettua dovrebbero essere valutati nell'ambito di tale analisi.

32 La Corte di giustizia dell'Unione europea ha affermato che non è necessario considerare i mezzi se, in realtà, la probabilità del loro utilizzo è insignificante perché sono vietati dalla legge.[28]Occorre precisare che questi casi riguardavano soggetti che erano legittimamente limitati dallo stato di diritto, cosicché la legge impediva di fatto l'uso dei mezzi in questione. In effetti, si può generalmente presumere che le persone rispettino la legge. Tuttavia, potrebbero esserci anche casi in cui l'illegalità non costituisce un ostacolo significativo all'identificazione, poiché i soggetti coinvolti non saranno effettivamente vincolati da tale divieto.

33 L'assunto generale secondo cui le persone rispettano la legge può, pertanto, essere confutato qualora vi siano prove sufficienti di un rischio concreto che, ciononostante, sia ragionevolmente probabile che vengano utilizzati mezzi illeciti. La valutazione di tale rischio dovrebbe considerare, in particolare, le prove del mancato rispetto del divieto in questione da parte dei soggetti coinvolti. I fattori rilevanti possono includere, a titolo esemplificativo ma non esaustivo:

- prove che il divieto legale non viene monitorato, applicato e dissuaso in modo efficace;
- prove che i vantaggi derivanti dalla violazione del divieto legale potrebbero superare i costi, gli sforzi e i potenziali rischi connessi a tale azione;
- prove di una situazione in cui i dati rilevanti per la reidentificazione (*cioè* sia i dati forniti sia, potenzialmente, i dati aggiuntivi utili per la reidentificazione) sono particolarmente vulnerabili all'accesso illegale tramite mezzi proibiti, a causa della mancanza di protezione; e
- prove che il divieto è stato precedentemente violato o eluso in situazioni analoghe.

Esempio 8: Un'azienda mediatica desidera condividere determinate informazioni con un'autorità pubblica locale, a condizione che possano essere rese anonime prima di farlo. L'azienda mediatica è l'unica entità ad avere accesso a queste informazioni ed è certa che non siano disponibili tramite altre fonti. L'azienda mediatica, tuttavia, non vuole eliminare le fonti originali utilizzate per generare tali informazioni e desidera utilizzare una tecnica di anonimizzazione che

[26] Vedere, *per esempio*, la capacità di un ente privato di ottenere informazioni aggiuntive attraverso canali legali, come indicato nel documento C-582/14 *Breyer contro Repubblica federale di Germania*, paragrafo 47.

[27] Il potenziale valore dei dati personali risiede nel beneficio che un'entità può trarre dalla reidentificazione e deve essere distinto dalla sensibilità dei dati stessi. Ad esempio, le informazioni sugli hobby di una persona possono non essere sensibili, ma risultare economicamente preziose ai fini del marketing mirato.

[28] Vedi, ad esempio, C-582/14 *Breyer contro Repubblica federale di Germania*, paragrafo 46; C-479/22 P *OC contro Commissione europea*, paragrafo 51; e C-413/23 P *EPA contro SRB*, paragrafo 82.

In pratica, l'identificazione non è possibile a meno che i dati (altrimenti anonimi) non vengano combinati con informazioni riservate in possesso solo dell'azienda mediatica. Nella sua valutazione, l'azienda mediatica osserva che tali informazioni riservate sono archiviate sui propri server e che, sebbene la sicurezza della rete stia diventando obsoleta, l'accesso di terzi a tali server è vietato dalla legge. Poiché l'autorità pubblica è legittimamente limitata dallo stato di diritto, l'azienda mediatica conclude che il rischio di identificazione è insignificante dal punto di vista dell'autorità pubblica e che le informazioni sono anonime per tale autorità.

Tuttavia, la società di media riconosce anche il rischio che i dati condivisi vengano intercettati illegalmente da terze parti non autorizzate. Tra queste terze parti rientrano entità situate al di fuori dell'ambito di applicazione del diritto dell'Unione e degli Stati membri, non vincolate di fatto da leggi che vietano l'accesso non autorizzato. Inoltre, si sono verificati diversi episodi di accesso illegale che hanno interessato i sistemi informatici di questa e di altre società di media che memorizzano dati personali.

Poiché le misure di sicurezza obsolete dell'azienda mediatica implicano che anche terze parti potrebbero probabilmente violare i server dell'azienda stessa, questi destinatari non autorizzati dei dati sarebbero in grado di accedere alle informazioni riservate e quindi reidentificare le persone attraverso mezzi ragionevolmente utilizzabili. Le informazioni condivise non sarebbero quindi considerate anonime dal loro punto di vista. L'azienda mediatica decide pertanto di aggiornare la sicurezza della propria rete prima di testare nuovamente l'anonimato dei dati.

34 Un divieto previsto in un contratto (anche se legalmente vincolante per le parti) non dovrebbe essere considerato un divieto di legge. Sebbene le clausole contrattuali possano influenzare i mezzi che ragionevolmente si prevede di utilizzare, esse dovrebbero essere impiegate solo a complemento di misure tecniche ed è importante valutare attentamente l'effettiva efficacia del loro effetto. Tra le altre cose, le misure dovrebbero essere affidabili, verificabili e applicabili.[29]tenendo presente che le misure contrattuali possono tipicamente essere soggette a revisione, o addirittura essere completamente ignorate dalle parti.

35 Infine, la probabilità di reidentificazione in genere aumenta nel tempo a causa dei progressi tecnologici e delle tecniche utilizzate per la reidentificazione, nonché della maggiore disponibilità di informazioni aggiuntive. Se la probabilità di identificazione aumenta fino a un livello non più trascurabile, i dati precedentemente anonimi devono essere nuovamente considerati personali e il titolare del trattamento sarà responsabile di qualsiasi trattamento di tali dati personali.[30]Pertanto, è buona prassi per gli enti che elaborano dati anonimizzati, ove possibile e opportuno, rivalutare periodicamente la probabilità di reidentificazione.

2.5 Considerazioni aggiuntive

2.5.1 Insiemi di dati contenenti un mix di dati anonimi e personali

36 Alcune tecniche di anonimizzazione potrebbero rendere insignificante la probabilità di identificazione solo per alcuni individui all'interno di un dataset. In questi casi, è importante distinguere tra l'anonimato dei singoli record e l'anonimato dell'intero dataset.*per sé*In particolare, il dataset nel suo complesso dovrebbe essere considerato anonimo solo se l'anonimizzazione è efficace per tutti gli individui inclusi: ciò non richiede necessariamente un livello di protezione uguale per tutti gli individui, ma richiede che la probabilità di reidentificazione sia insignificante per tutti.

[29]Commissione europea e EDPB, Linee guida comuni sull'interazione tra la legge sui mercati digitali e il regolamento generale sulla protezione dei dati (versione per la consultazione pubblica), paragrafo 181.
[30]Per le condizioni generali sulla responsabilità dei titolari del trattamento per comportamenti illeciti, si veda C-807/21.*Deutsche Wohnen SE contro Staatsanwaltschaft Berlino*, paragrafi 76 e 78.

questi individui. Laddove ciò non avvenga, e quindi un dataset contenga un mix di dati personali e anonimi, l'intero dataset deve essere trattato come contenente dati personali (e quindi rientrante nell'ambito di applicazione del GDPR) in qualsiasi situazione in cui le singole parti del dataset non siano trattate separatamente.[31].

2.5.2 Conformità al GDPR e processo di anonimizzazione

37 L'EDPB sottolinea che, quando si rendono anonime le informazioni da alcune prospettive ma non da altre, è importante considerare le possibili implicazioni per la conformità al GDPR. Ad esempio, quando le informazioni vengono anonimizzate per l'entità che riceve i dati, ma non per il titolare del trattamento che le anonimizza, quest'ultimo dovrà comunque trattare i dati come dati personali e rispettare tutti i propri obblighi GDPR in merito.[32].

38 I principi di protezione dei dati si applicano ogniqualvolta vengono trattati dati personali, comprese le operazioni di trattamento effettuate per ottenere informazioni anonime. Ciò significa, tra l'altro, che l'anonimizzazione deve avere una base giuridica ai sensi dell'articolo 6 del GDPR e che, se il set di dati contiene dati personali di categoria particolare, deve applicarsi una delle esenzioni previste dall'articolo 9, paragrafo 2, del GDPR. La base giuridica e, se del caso, la relativa esenzione di cui all'articolo 9, paragrafo 2, si presumono coincidenti con la base giuridica e l'esenzione del trattamento precedente all'anonimizzazione, qualora l'anonimizzazione sia parte della stessa attività di trattamento e il titolare del trattamento persegua con essa le stesse finalità.[33].

39 L'anonimizzazione (almeno se combinata con la cancellazione dei dati originali) può essere utile anche per i titolari del trattamento che desiderano basarsi sul legittimo interesse per le attività di trattamento precedenti. In particolare, l'anonimizzazione immediata dei dati e la cancellazione immediata dei dati originali possono limitare l'impatto complessivo delle attività di trattamento sugli interessati, fungendo quindi da fattore positivo nel bilanciamento degli interessi legittimi.[34].

40 I titolari del trattamento devono inoltre rispettare gli obblighi di trasparenza previsti dagli articoli 5, paragrafo 1, lettera a), e 12-15 del GDPR in merito al trattamento di anonimizzazione. Tra le altre cose, i titolari del trattamento devono dichiarare chiaramente che i dati personali saranno trattati al fine di produrre dati anonimi, che quindi non rientrano nell'ambito di applicazione del GDPR, ed evitare dichiarazioni ambigue o fuorvianti. In particolare, i titolari del trattamento non devono utilizzare termini quali "anonimo", "deidentificato" o "depersonalizzato" se gli individui sono ancora identificabili.

41 Allo stesso modo, i titolari del trattamento devono garantire un'adeguata documentazione del processo di anonimizzazione. Tale documentazione (compresa la verifica dei set di dati presumibilmente anonimi) consente di dimostrare sia la conformità del processo di anonimizzazione al GDPR, sia l'efficacia dell'anonimizzazione stessa. Questa documentazione deve essere conservata anche dopo il completamento del processo di anonimizzazione.

42 In alcuni casi, un incidente di sicurezza può portare a una rivalutazione dell'anonimato. Ciò si verifica in particolare se la valutazione dell'anonimato si basava sul mantenimento della riservatezza di determinate informazioni, ma l'incidente fa sì che gli individui possano ora essere identificati con mezzi

---

[31] Si veda, a titolo di confronto, il fatto che un insieme di dati contenente un mix di dati personali di categoria speciale e non speciale deve essere trattato collettivamente come dati personali di categoria speciale ogniqualvolta le diverse parti non possono essere trattate separatamente: C-252/21 *Meta v Bundeskartellamt*, paragrafo 89.

[32] Vedere, ad esempio, C-413/23 P *EPA contro SRB* e gli effetti dell'anonimizzazione (o della sua mancanza) su alcuni aspetti dell'obbligo di informazione, come discusso nell'Esempio 2.

[33] Si veda anche l'EDPB Linee guida 1/2026 sul trattamento dei dati personali a fini di ricerca scientifica, adottate il 15 aprile 2026, versione per consultazione pubblica, sezione 3.1.

[34] Per ulteriori informazioni sul test di bilanciamento dell'interesse legittimo, consultare le Linee guida 1/2024 dell'EDPB sul trattamento dei dati personali ai sensi dell'articolo 6, paragrafo 1, lettera f) del GDPR.

ragionevolmente probabile che venga utilizzato. In tal caso, l'entità che ha subito l'incidente dovrebbe quindi determinare il proprio ruolo rispetto al trattamento e se deve adempiere a eventuali obblighi ai sensi degli articoli 33 e 34 del GDPR.[35].

3. Analisi tecnica dell'anonimizzazione

3.1 Introduzione

43 Non sarà sempre immediatamente evidente se le informazioni si riferiscono a una persona fisica identificata o identificabile. Ciò si verifica soprattutto quando il titolare del trattamento che effettua l'anonimizzazione ha deliberatamente cercato di rendere difficile il riconoscimento del fatto che le informazioni si riferiscono a una persona fisica identificata o identificabile. È pertanto necessario applicare un'analisi tecnica che tenga conto dello stato dell'arte delle tecniche di reidentificazione e che possa essere utilizzata per stabilire se i dati sono anonimi o meno.

44 Considerata l'ampia gamma di possibili set di dati e tecniche di anonimizzazione, nonché i futuri sviluppi nelle tecniche di (re)identificazione, le valutazioni richiederanno spesso un'analisi caso per caso e potrebbero necessitare di considerazioni non completamente trattate in questo documento. Ciononostante, questa sezione fornirà indicazioni su un quadro di riferimento utilizzabile per testare l'anonimato dei dati.

45 Il quadro di riferimento descritto in questa sezione può essere utilizzato in due modi. Il primo è un approccio contestualizzato, basato sull'analisi giuridica sopra descritta, che utilizza informazioni sui mezzi a disposizione di tutti i soggetti interessati e sulla probabilità del loro utilizzo. Il secondo è un approccio semplificato, che non tiene conto delle differenze tra i soggetti in merito ai mezzi che è ragionevolmente probabile che utilizzino. Questo approccio semplificato potrebbe indurre il titolare del trattamento ad anonimizzare i dati come se non fossero anonimi, anche se in realtà lo sarebbero per i soggetti interessati. Tuttavia, può fornire maggiore affidabilità e può essere integrato con l'approccio contestualizzato per affinare i risultati.

46 Dopo aver illustrato in dettaglio questi due approcci, questa sezione esamina brevemente i casi più semplici, in cui i dati forniti sono chiaramente personali e l'utilizzo dell'intero framework non è necessario. Si passa quindi alla parte principale del framework stesso, che si compone di tre criteri generali: Nessun isolamento dei record, Nessun collegamento e Nessuna inferenza. Se tutti e tre i criteri sono soddisfatti, le informazioni possono essere considerate anonime; se uno o più di essi vengono violati, potrebbe essere necessaria un'ulteriore analisi. La sezione si conclude fornendo indicazioni su come applicare e valutare il framework e su cosa comporta tale ulteriore analisi.

3.2 Approcci alla valutazione

47 Lo standard legale per l'anonimizzazione può produrre risultati diversi a seconda delle prospettive applicabili. Questo approccio contestuale porta a una valutazione differenziata, in cui le rispettive capacità delle entità interessate dovrebbero essere valutate individualmente. In pratica, ciò può essere piuttosto complesso, soprattutto quando è difficile tenere conto (o addirittura conoscere) tutti i possibili elementi contestuali coinvolti. Questo approccio contestuale corre quindi il rischio di falsi positivi, in cui il valutatore può concludere erroneamente che i dati sono anonimi perché non è a conoscenza dei mezzi di una particolare entità che è ragionevolmente probabile che vengano utilizzati per

[35]Per ulteriori indicazioni su questo argomento, si veda la Guida EDPB 9/2022 sulla notifica delle violazioni dei dati personali ai sensi del GDPR.

identificazione. Tuttavia, se eseguito correttamente, questo approccio consente un'analisi completa del set di dati e permette ai titolari del trattamento di essere certi che il set di dati rientri nell'ambito di applicazione del GDPR.

48 Per semplicità, quindi, alcuni titolari del trattamento dei dati anonimizzati potrebbero decidere di adottare un approccio semplificato per la valutazione. In base a tale approccio, il titolare del trattamento può semplicemente scegliere di ignorare le differenze tra le entità in merito all'accesso ai dati forniti, alle informazioni aggiuntive o ad altre risorse che potrebbero contribuire alla reidentificazione. Occorre sottolineare che questo approccio semplificato non costituisce una norma giuridica alternativa.[36]Questo approccio, invece, rappresenta un modo per spostare volontariamente il rischio dai falsi positivi ai falsi negativi, in cui il titolare del trattamento che effettua l'anonimizzazione tratta i dati anonimi come dati personali perché ha (di fatto) sovrastimato la probabilità che vengano utilizzati determinati mezzi. In alcuni casi, questo approccio fornirebbe un livello di protezione più prudente di quanto strettamente necessario, ma così facendo garantisce che gli standard legali stabiliti dal GDPR siano sempre rispettati in caso di dubbio.

49 In pratica, una combinazione dei due approcci può spesso rivelarsi utile. Ad esempio, un responsabile dell'anonimizzazione potrebbe voler iniziare con l'approccio semplificato, chiedendosi se la reidentificazione sia teoricamente possibile. Se ciò non è possibile, i dati possono essere considerati anonimi in tutta sicurezza. Se, tuttavia, l'approccio semplificato individua l'esistenza di determinati metodi di reidentificazione, il responsabile dell'anonimizzazione può passare all'approccio contestuale e valutare se tali metodi siano ragionevolmente probabili dal punto di vista delle entità interessate. In alternativa, il responsabile dell'anonimizzazione potrebbe effettuare una mappatura iniziale dei dati e rendersi immediatamente conto che esiste un'ampia gamma di metodi di reidentificazione teoricamente possibili, ma che possono essere rapidamente esclusi in quanto impossibili da applicare ai dati in questione nella pratica. Tale responsabile dell'anonimizzazione potrebbe quindi decidere di passare direttamente all'approccio contestuale, che consente di scartare immediatamente questi metodi perché, in realtà, non sono mezzi ragionevolmente probabili da utilizzare da parte di alcuna entità.

Esempio 9: Il titolare del trattamento desidera anonimizzare un set di dati contenente una combinazione di informazioni demografiche e mediche. Inizia con un'analisi semplificata e scopre che i soli dati demografici non consentirebbero la reidentificazione degli individui in questo caso. Tuttavia, scopre anche che la reidentificazione potrebbe essere possibile se le informazioni demografiche venissero combinate con alcune informazioni mediche aggiuntive. In base all'analisi semplificata, concluderebbe quindi che le informazioni non sono anonime. Tuttavia, il titolare del trattamento decide di estendere l'analisi e passa a un'analisi contestuale. Inizia quindi a identificare le prospettive applicabili e a verificare se le rispettive entità potrebbero accedere a tali informazioni mediche aggiuntive con mezzi ragionevolmente probabili da utilizzare.

3.3 Casi semplici e mappatura dei dati

50 In alcuni casi, può essere immediatamente evidente che le informazioni non sono anonime. Ciò include, ad esempio, un set di dati che include il nome, l'indirizzo e-mail, il numero di telefono o il numero di previdenza sociale di un individuo; un set di dati che contiene pseudonimi che possono essere facilmente decodificati per scoprire un identificatore, o che può essere utilizzato per identificare gli individui; un set di dati privo di identificatori diretti ma che contiene chiaramente attributi che consentono l'identificazione degli individui; o un set di dati che viene chiaramente utilizzato per l'identificazione.

[36]Ciò significa, in particolare, che mentre l'approccio semplificato può essere piuttosto utile per i titolari del trattamento che desiderano anonimizzare i dati al fine di dimostrare che le informazioni sono anonime e non rientrano nell'ambito di applicazione del GDPR, non può essere utilizzato da terzi per dimostrare in modo definitivo che le informazioni sono dati personali e che, pertanto, il trattamento di tali dati da parte del titolare del trattamento viola il GDPR.

scopi di valutazione, trattamento o influenza dello status o del comportamento di un individuo specifico in un modo particolare.

51 Per individuare questi e altri casi semplici, si raccomanda di mappare i dati, nonché qualsiasi informazione aggiuntiva che possa facilitare l'identificazione (o la riidentificazione), sia che sia in possesso di altri soggetti interessati o accessibile tramite mezzi che è ragionevolmente probabile vengano utilizzati. Se i dati non sono chiaramente anonimi in questa fase, non è necessario applicare i criteri indicati di seguito. Questa operazione dovrebbe essere eseguita come fase iniziale della valutazione e, anche se non emerge alcuna semplice identificazione, può comunque contribuire a semplificare le fasi successive.

3.4 I tre criteri

52 Questa sottosezione definisce i tre criteri: Assenza di isolamento dei dati, Assenza di collegamento e Assenza di inferenza, con ulteriori indicazioni sulla loro applicazione contenute nella sottosezione 3.5. Se tutti e tre i criteri sono soddisfatti, le informazioni possono essere considerate anonime. D'altra parte, la violazione di un criterio non implica necessariamente che le informazioni debbano essere considerate dati personali; piuttosto, se uno o più criteri vengono violati, è necessario proseguire l'analisi, come indicato nella sottosezione 3.5.3, per valutare l'impatto di tale violazione e se i dati possano ancora essere considerati anonimi.

53 Gli stessi criteri vengono utilizzati sia per l'approccio contestuale che per quello semplificato. A scopo esplicativo, la descrizione di questi criteri nelle sottosezioni 3.4.1–3.4.3 si basa sull'approccio semplificato, il che consente a questa sottosezione di concentrarsi sui metodi stessi. Le questioni contestuali verranno poi reintrodotte nella sottosezione 3.5.

3.4.1 Nessun isolamento dei record

54 Un metodo comune utilizzato per distinguere gli individui dagli altri in un dato gruppo consiste nell'individuarli facendo riferimento a uno o più attributi[37]. Per sua stessa natura, è molto più facile farlo quando si registra –*cioè*Tutti i valori all'interno dei dati forniti che si riferiscono a un singolo individuo sono unici. Pertanto, l'isolamento dei record può portare alla violazione dell'anonimato per i dati in questione. Questo porta al criterio di non isolamento dei record.

55 Il criterio di assenza di isolamento dei record è soddisfatto se i dati non contengono una combinazione univoca di valori di attributi che si riferiscono a un singolo individuo.

56 Quanto più un record è grande e quanto più attributi contiene, tanto maggiore è la probabilità che il record sia univoco all'interno dei dati forniti.

57 Il criterio di assenza di isolamento dei record può essere generalmente verificato esaminando i dati forniti singolarmente. Poiché un record comprende tutte le informazioni note relative a un singolo individuo, il criterio di assenza di isolamento dei record può comportare l'analisi, ad esempio, di più righe in una tabella o in una tabella di database, a seconda di come sono strutturati i dati.

[37] Il concetto di individuazione verrà spiegato in dettaglio nella sezione 3.5.3.

Example 10: A research institute holds a dataset containing individual records of patients. Each record contains information about the individual's sex, date of birth, postcode, and the autoimmune disease affecting the patient. The data was verified before inclusion in the dataset and has not been modified.

Sex	Date of birth	Postcode	Disease
Female	28-09-1955	43221	celiac disease
Female	06-01-1955	43210	Crohn's disease
Male	08-07-1959	89127	Vitiligo
Male	10-06-1959	89127	multiple sclerosis
Male	27-09-1959	89127	Diabetes mellitus type 1

In this case, all of the records are unique. The No Record Isolation criterion is violated because of the unique combinations of the attributes.

58 Notably, aggregate data will usually satisfy the No Record Isolation criterion – if the aggregation was done correctly, it should result in information that relates to groups of individuals and, therefore, should not actually contain any individual records which could be isolated.

3.4.2 No Linkage

59 Another common method of identifying an individual is by linking information from one dataset to another, or multiple other, dataset(s). By establishing that a record in Dataset A relates to the same individual as a record in Dataset B (and, if relevant, Datasets C, D and E *etc.*), the linkage increases the amount of information about that individual, which in turn makes it more likely that the individual can then actually be identified. Linkages can then lead to the violation of anonymity for the given data. This leads to the No Linkage criterion.

60 The No Linkage criterion is met if the data does not contain an individual's record which could be linked to another record which (a) also relates (with certainty or high likelihood) to that same individual, and (b) comes from a different dataset.

61 Unlike the No Record Isolation criterion, which was assessed by reference to the given data alone, this criterion requires knowledge and understanding of other datasets that could be linked to the given data with means reasonably likely to be used[38].

62 Linkage may often be possible by reference to a common identifier (e.g., an internal customer number) which is used in both datasets, or by matching records based on certain combinations of attributes.

Example 11: A board game shop keeps a record of every purchase made by every customer. This is clearly personal data, since each purchase is linked to a specific user. The shop wishes

[38] As noted in para. 53 above, this section will explain the criteria following the simplified approach, to focus the explanation on the core parts of the criterion. It will not, therefore, consider whether the relevant entities could access those other datasets with means reasonably likely to be used. Under the contextual approach, however, it would also be necessary to ask whether such links are actually possible with means reasonably likely to be used from the relevant perspective, as explained in sub-sections 2.2 and 2.4 and discussed in sub-sections 3.5.2 and 3.5.3 below. This is true for all three of the criteria, but can be especially helpful to emphasise for the sake of the examples set out in this sub-section.

to anonymise the data and so decides to test if simply erasing the direct identifiers would be enough to do so.

The shop immediately notices that many of the records in the given data are unique, and that the given data therefore fails the No Record Isolation criterion. However, since failing a criterion does not necessarily mean that the given data is personal, the shop continues to test the other criteria before starting its assessment of the results.

To test the No Linkage criterion, the shop checks for other datasets which might contain matching records based on certain combinations of attributes (in this case, the specific combinations of purchased board games). This leads to a very popular website where users create lists of the games which they own and mark how often they play each of those games. The shop finds that several records on the website match those in the given data, where customers have purchased games from the shop and then entered that information onto the website. These matching records mean that the datasets can be linked, and the shop therefore concludes that the given data also violates the No Linkage criterion.

63 In practice, data which satisfies the No Record Isolation criterion will often also satisfy the No Linkage criterion. However, the two criteria should still be assessed separately, as there are ways to link non-unique data.

Example 12: A research institute processes a dataset containing individual records of some patients. In the dataset, the date of birth and postcode attributes have been generalised as follows:

Sex	Date of birth	Postcode	Disease
Male	**-**-1946	54***	sclerosis
Male	**-**-1946	54***	sclerosis
Female	**-**-1951	32***	lung cancer
Female	**-**-1951	32***	lung cancer
Female	**-**-1951	32***	lung cancer

The data satisfies the No Record Isolation criterion since no record is unique in the dataset. However, looking at the first two records, this dataset reveals that all individuals born in 1946 and living in any postcode starting with 54 (and whose record was included in the dataset) have sclerosis.

Therefore, the data could be linked to information in other datasets which contain the same individual's sex, year of birth and generalised postcode. The fact that the datasets contain information about the same individual could, for example, be established through prior knowledge that two datasets contained information about the same individual, from information contained in either dataset which makes this link clear, or though information in a third dataset which acts as a bridging link. If this link can be established, the No Linkage criterion is violated.

64 It may still be possible to achieve linkage on records that are not fully accurate (including where records are intentionally modified, *e.g.*, by adding noise to the data). This can often be done to a high degree of accuracy and with a likelihood of success by, for example, using

attacchi robusti di collegamento dei record[39]Pertanto, è generalmente difficile stabilire se le imprecisioni siano sufficientemente ampie da impedire il collegamento, poiché i dati rumorosi possono divulgare informazioni in modi inaspettati. Inoltre, l'efficacia dell'attacco dipende fortemente dalla disponibilità di informazioni aggiuntive.

65 Il criterio di non collegamento è probabilmente soddisfatto se le informazioni non sono state registrate altrove e non sono correlate a informazioni simili sullo stesso individuo registrate in altri contesti. Questo criterio può essere soddisfatto, ad esempio, nel caso delle risposte fornite a un sondaggio d'opinione, a condizione che non contengano, tra l'altro, dati demografici che potrebbero essere ricondotti a un individuo. Un altro esempio potrebbe essere rappresentato dalle registrazioni di particolari stati fisiologici dei pazienti, a condizione che, tra l'altro, vengano raccolti valori di attributi simili per molti pazienti e che tale stato sia suscettibile di variare nel tempo.

3.4.3 Nessuna inferenza

66 In generale, l'"inferenza" è il processo di giungere a una conclusione mediante deduzione da prove e ragionamenti.[40]Attraverso questo processo, un'entità competente può ottenere efficacemente informazioni dai dati forniti. Le inferenze possono essere tratte sia da dati a livello di record che da dati aggregati. Le inferenze possono anche "arricchire" i dati in modo da rivelarne la pertinenza personale, anche se le informazioni inferite non erano chiaramente presenti nel set di dati prima che l'inferenza venisse effettuata.[41]Tali inferenze possono violare l'anonimato dei dati forniti. Ciò porta al criterio di non inferenza.

67 Il criterio "Nessuna inferenza" è soddisfatto se non è possibile trarre alcuna inferenza specifica e significativa dai dati forniti. Le inferenze possono essere basate sui dati forniti e/o su informazioni aggiuntive, comprese le informazioni relative al processo di anonimizzazione. In particolare:

- Un'inferenza è specifica se le informazioni inferite si riferiscono a un singolo individuo identificato o identificabile (*cioè*si tratta di dati personali); e
- Un'inferenza è specifica e significativa se il trattamento delle informazioni inferite è suscettibile di incidere sui diritti e sugli interessi dell'interessato, si basa sui dati forniti e non potrebbe essere ricavata da conoscenze generali o da dati riguardanti la popolazione nel suo complesso.[42].

68 Questo criterio deve essere valutato facendo riferimento ai dati forniti, unitamente a qualsiasi informazione esterna che possa essere utilizzata in combinazione con i dati forniti per ricavare informazioni, compreso il contesto in cui i dati forniti sono stati prodotti e qualsiasi insieme di dati esterni che possa essere integrato con essi.

69 Spesso è possibile dedurre alcuni dati personali da dati forniti in qualche modo, anche se i dati forniti stessi non sono personali. In tali casi, i dati personali dedotti sarebbero certamente

---

[39]Vedere,*per esempio*, Arvind Narayanan e Vitaly Shmatikov, "De-anonimizzazione robusta di grandi dataset sparsi" (2008) 2008 IEEE Symposium on Security and Privacy 111, https://doi.org/10.1109/SP.2008.33; e Julien Freudiger, Reza Shokri e Jean-Pierre Hubaux, "Valutazione del rischio per la privacy dei servizi basati sulla posizione" in George Danezis (a cura di)*Crittografia finanziaria e sicurezza dei dati: 15ilConferenza internazionale, FC 2011, Gros Islet, Santa Lucia, 28 febbraio - 4 marzo 2011, Atti selezionati e rivisti.* (Springer 2012), https://doi.org/10.1007/978-3-642-27576-0_3.

[40]Vedi le voci "infer" e "inference" in Catherine Soanes & Angus Stevenson (a cura di).*Dizionario inglese Oxford conciso*((Undicesima edizione riveduta, OUP 2009).

[41]A tale proposito, è importante ricordare che, come discusso nei paragrafi 2.3 e 2.4, le informazioni possono essere considerate dati personali anche se il loro collegamento con un individuo non è immediatamente evidente o se le informazioni stesse non contengono tutto il necessario per identificare l'individuo a cui si riferiscono.

[42]In questo contesto, "informazioni sulla popolazione nel suo complesso" si riferisce a qualsiasi gruppo numeroso di persone per le quali le informazioni non vengono raccolte individualmente, ma estrapolate da un campione, senza alcun collegamento con i singoli individui. In termini tecnici, tali inferenze possono essere considerate parte delle "credenze a priori".

Le informazioni si riferiscono a un individuo identificato o identificabile, ma ciò non significa che lo stesso valga per i dati in questione. Ad esempio, quando si lavora con dati a livello di record, è possibile dedurre informazioni su un individuo completamente nuovo che non era incluso nel dataset originale, ma che presenta attributi simili a quelli di persone che vi erano incluse. In tal caso, le informazioni dedotte si riferiscono certamente a questo nuovo individuo, ma ciò non significa di per sé che i dati in questione si riferiscano a esso.

70 È quindi importante distinguere tra due situazioni. Nella prima situazione, i dati personali vengono dedotti dai dati forniti, ma non sono effettivamente collegati al dataset originale (*cioè* (le informazioni utilizzate per generare i dati forniti). Questo tipo di inferenza non porterebbe a una violazione del criterio di non inferenza perché l'inferenza non è significativa: è ottenuta da informazioni contenute nei dati forniti che sono di conoscenza generale o riguardano la popolazione in generale – risultato di una generalizzazione dal dataset originale rappresentato dai dati forniti. Nella seconda situazione, i dati personali inferiti sono effettivamente collegati al dataset originale ed è ragionevolmente probabile che abbiano un effetto sui diritti e sugli interessi dell'interessato. L'inferenza, quindi, è significativa perché il collegamento al dataset originale dimostra che non è ottenuta da conoscenza generale o da dati generalizzati sulla popolazione in generale, ma da informazioni sull'interessato contenute nelle informazioni inferite rappresentate dai dati forniti e provenienti dal dataset originale. Questo tipo di inferenza porterebbe a una violazione del criterio di non inferenza.

71 Affinché i dati forniti possano essere considerati effettivamente relativi a un individuo, l'inferenza deve essere specifica e significativa. Ciò significa che l'inferenza produrrà dati personali il cui trattamento è suscettibile di incidere sui diritti e sugli interessi dell'interessato, si basa sui dati forniti e non potrebbe essere ottenuta da conoscenze generali o da dati relativi alla popolazione nel suo complesso. Questo può essere generalmente dimostrato in uno dei tre modi seguenti:

a. Idealmente, ciò può essere fatto tracciando il modo in cui si giunge all'inferenza.

b. Se ciò non è possibile,*per esempio*A causa dell'opacità dell'elaborazione automatizzata che ha portato alle informazioni inferite, una mancanza di affidabilità potrebbe essere dimostrata anche confrontando il risultato dell'inferenza con due versioni dei dati forniti: una prodotta dal dataset originale e una prodotta da una versione modificata del dataset originale che esclude i record relativi all'individuo. Se l'inferenza può essere tratta solo dalla prima versione dei dati forniti, è chiaro che l'inferenza si basa sui record relativi all'individuo e, di conseguenza, che i dati forniti si riferiscono all'individuo.[43].

c. Se nessuna di queste opzioni è fattibile,*per esempio*poiché le informazioni sono state aggregate e il dataset originale è stato cancellato, è anche possibile fare affidamento su una presunzione di confutazione secondo cui un'inferenza non si basava sui dati forniti dimostrando che

[43]In termini più tecnici, l'argomentazione presentata in questo paragrafo è la seguente: inferire significa fare una previsione su una proprietà S avendo osservato i dati forniti A(D) e informazioni aggiuntive, dove A(D) è stato prodotto eseguendo il processo di anonimizzazione A sui dati originali D e le informazioni aggiuntive fungono da fattore condizionante. Supponiamo che S sia una proprietà relativa a una persona di nome John Smith e che i dati relativi a tale persona siano inclusi in D. Indichiamo con D' una copia del dataset D, con l'unica differenza che tutti i dati relativi a John Smith sono stati rimossi. Pertanto, D' non contiene alcun dato relativo a John Smith. Potrebbe comunque essere possibile fare delle inferenze su John Smith basandosi su A(D'), anche sulla proprietà S. Ad esempio, ciò potrebbe essere possibile se gli altri record in D' si riferiscono a persone "simili" a John Smith (ad esempio perché vivono nella stessa città). Tuttavia, questa inferenza non può dipendere dai dati di John Smith, poiché non ci sono dati relativi a John Smith in D'. Pertanto, se l'inferenza su John Smith funziona allo stesso modo per A(D) e A(D'), non si tratta di un'inferenza tratta dai suoi dati, bensì da dati relativi alla popolazione nel suo complesso, di cui D' è rappresentativo.

avrebbero potuto essere ricavate da conoscenze generali o da informazioni sulla popolazione nel suo complesso[44].

Esempio 13: L'affermazione non qualificata "Il verde è un colore popolare" è un dato anonimo poiché non si riferisce ad alcuna persona fisica identificata o identificabile. Da questa informazione potremmo dedurre che a Connor piace il colore verde. L'affermazione dedotta "A Connor piace il verde" è chiaramente un'informazione relativa a una persona fisica identificata e dovrebbe quindi essere trattata come dato personale. Tuttavia, questa deduzione non significa che anche "Il verde è un colore popolare" sia un dato personale.

Sebbene l'informazione inferita "A Connor piace il verde" sia specifica (poiché si riferisce a un singolo individuo identificato), non è significativa perché può essere (e, di fatto, è stata) ottenuta da informazioni riguardanti l'intera popolazione. Questa inferenza non comporta, pertanto, una violazione del criterio di non inferenza per i dati forniti "Il verde è un colore popolare".

Esempio 14: Un dataset anonimizzato a livello di record, contenente informazioni storiche sui prestiti di una banca, mostra che gli individui con una particolare combinazione di attributi hanno una probabilità di non riuscire a rimborsare il prestito. La banca utilizza questa informazione per valutare un nuovo richiedente di prestito, che non era mai stato incluso nel dataset, e deduce che tale richiedente presenta un alto rischio di inadempienza. Tuttavia, questa non è un'inferenza significativa nel senso definito in precedenza, e pertanto il dataset potrebbe comunque soddisfare il criterio di "Nessuna inferenza".

72 È importante sottolineare che le inferenze che, a una prima analisi, non sembrano specifiche e significative, possono comunque rivelarsi utili per la (re)identificazione e dovrebbero essere ulteriormente analizzate in combinazione con i dati disponibili. Ad esempio, i dati disponibili, unitamente alle informazioni inferite, potrebbero condurre a ulteriori inferenze che risulterebbero effettivamente specifiche e significative, violando quindi il criterio di "Nessuna inferenza".

73 Esiste un ampio spettro di processi che possono condurre a inferenze rilevanti. Da un lato ci sono le conclusioni raggiunte semplicemente ragionando logicamente sui dati, dall'altro quelle ottenute tramite analisi elaborate o l'utilizzo di processi automatizzati che seguono algoritmi sofisticati. Alcune forme di inferenza possono essere più semplici, o più pertinenti, per i dati a livello di record, mentre altre possono essere particolarmente efficaci nell'estrarre informazioni da dati aggregati. Le seguenti sottosezioni discuteranno e forniranno esempi di alcuni di questi processi. Tuttavia, una data inferenza può contribuire alla reidentificazione in più di un modo e l'elenco seguente non deve essere considerato esaustivo.

Aggiunta di un attributo inferito a un determinato record

74 Potrebbe essere possibile aggiungere informazioni inferite ai dati a livello di record in un modo che porti poi a una violazione del criterio di non inferenza. Queste informazioni inferite non devono necessariamente essere aggiunte al record: è sufficiente che sia possibile allegarle al record.

Esempio 15: Un controller riceve dati di posizione precisi sugli utenti di un'app mobile, rimuove tutti gli identificatori diretti e li raggruppa in tracce di 24 ore per ciascun utente. Il controller desidera valutare se tali tracce costituiscono dati anonimi. Le tracce mostrano che,

[44] In tal caso, la "credenza a priori" (vedi nota 40) potrebbe includere la conoscenza derivante da una versione alternativa dei dati forniti, prodotta senza la registrazione delle persone che vengono identificate – A(D') nella nota 41 – ma che si applica comunque a loro e ad altre persone per estrapolazione. Le inferenze derivanti dai dati forniti dovrebbero produrre una "credenza a posteriori" (sulla quale potrebbe basarsi un'azione) significativamente diversa per essere considerate correlate all'individuo in un modo che classifichi i dati forniti come personali.

Ogni giorno feriale, un singolo utente è presente in un determinato edificio residenziale dalle 0:00 alle 08:00 e dalle 20:00 alle 24:00, e in un determinato edificio commerciale dalle 09:00 alle 17:00. Queste informazioni consentono di dedurre in modo significativo e specifico gli indirizzi di residenza e di lavoro dell'utente, che potrebbero poi essere aggiunti al suo profilo.

75 Il semplice fatto che sia possibile aggiungere informazioni dedotte a un record non significa necessariamente che il criterio di non inferenza sia violato. Sarà quasi sempre possibile trovare informazioni dedotte che possono essere aggiunte, sia analizzando i dati sia semplicemente applicando ad essi fatti generali. È quindi importante considerare la natura delle informazioni dedotte e se siano significative e specifiche, come indicato in precedenza.

Generazione di dati a livello di record a partire da dati aggregati

76 Potrebbe essere possibile dedurre dati a livello di record analizzando e/o combinando dati aggregati. Ciò è particolarmente vero per i dati anomali, dove questo tipo di inferenza è spesso più probabile.

Esempio 16: Un'azienda decide di pubblicare informazioni sugli stipendi totali corrisposti a ciascuna categoria di dipendenti. Le informazioni indicano che gli ingegneri dell'azienda ricevono complessivamente 550.000 euro all'anno, ma non specificano l'importo corrisposto a ciascun ingegnere. Tuttavia, i report interni dell'azienda indicano anche che l'azienda impiega un totale di sei ingegneri, che il team di Ingegneria di Prodotto è composto da cinque ingegneri (e non comprende altri dipendenti) e che il team di Ingegneria di Prodotto ha una spesa salariale annua di 480.000 euro. Sebbene entrambi i report si limitino a dati aggregati, un'entità in possesso di tutte queste informazioni può dedurre che il sesto ingegnere percepisce uno stipendio annuo di 70.000 euro.

77 Questa forma di inferenza può anche essere significativamente più sofisticata. Può, ad esempio, partire da grandi insiemi di statistiche aggregate accessibili in blocco o da sistemi di interrogazione dei dati. Spesso, questo tipo di inferenza utilizzerà più set di dati aggregati, soprattutto se tali set sono stati tutti prodotti dagli stessi dati originali. Ciò è particolarmente vero per le tecniche di deaggregazione avanzate, inclusi gli attacchi in grado di estrarre dati da modelli di intelligenza artificiale apparentemente anonimi.[45]Tuttavia, un singolo set di dati contenente una quantità sufficientemente ampia di informazioni può consentire di ricavare informazioni relative a un singolo individuo, senza bisogno di altre fonti.

78 È possibile trarre inferenze anche dall'assenza di un valore.

Esempio 17: Un'azienda organizza un sondaggio in cui ogni team può esprimere "in forma anonima" la propria opinione sul responsabile di linea. Il sondaggio include domande come: "Da 1 a 5, in che misura la cultura aziendale ti incoraggia a commettere errori e a imparare da essi?". Le risposte individuali non vengono condivise con i responsabili di linea, ai quali viene comunicato solo il numero totale di membri del loro team che hanno selezionato ciascuna risposta. Tuttavia, se un responsabile di linea viene informato che tutti i membri del suo team hanno partecipato al sondaggio e nessuna delle risposte ha dato un punteggio superiore a 3, può dedurre che tutti i membri del team hanno una visione critica della cultura aziendale. Sebbene aggregate, queste informazioni si riferiscono comunque a ciascun membro del team individualmente.

[45]Vedi anche EDPB: Parere 28/2024 su taluni aspetti di protezione dei dati relativi al trattamento dei dati personali nel contesto dei modelli di intelligenza artificiale.

Deduzione di appartenenza

79 Si può dedurre che i dati di un individuo fossero inclusi nel dataset originale utilizzato per produrre i dati presentati. Questo tipo di inferenza di appartenenza può rappresentare il primo passo per altri tipi di inferenze, oppure può di per sé rivelare informazioni sull'individuo.

Esempio 18: Un set di dati apparentemente anonimo viene generato a partire da informazioni sui membri di un club di canottaggio. Tuttavia, analizzando alcuni attributi contenuti nei dati forniti, è possibile dedurre che Alex era incluso nel set di dati originale e che quindi è un membro di quel club.

Collegamento dedotto a informazioni aggiuntive

80 Il collegamento con altri dati, come quello discusso nella sezione 3.4.2, può anche essere visto come una forma di inferenza. Collegando i dati forniti a informazioni aggiuntive, è possibile dedurre che le informazioni contenute nei dati forniti possono essere applicate alle informazioni aggiuntive e *viceversa* Questo tipo di inferenza può anche andare oltre il tipo di collegamento descritto nel criterio "Nessun collegamento". Ad esempio, se le informazioni non possono essere abbinate in modo univoco tra i dataset, ma le informazioni aggiuntive potrebbero essere abbinate a diverse possibili voci nei dati forniti, potrebbe comunque essere possibile trarre inferenze rilevanti. Ciò si verifica in particolare quando un valore di attributo compare in un numero preponderante di record corrispondenti.

81 Inoltre, mentre il criterio "Nessun collegamento" si riferiva in particolare ai dati a livello di record, questo tipo di inferenza si applica anche ai dati aggregati.

Esempio 19: I dati dell'Esempio 12 vengono aggregati per creare un istogramma delle diagnosi raggruppate per combinazione di sesso, anno di nascita e codice postale abbreviato. Nonostante l'aggregazione, è comunque possibile dedurre la diagnosi di un individuo collegando questo istogramma a (a) dati relativi al sesso, all'anno di nascita e al codice postale generalizzato dell'individuo; e (b) dati che indicano che la registrazione dell'individuo è stata inclusa nel dataset.

82 Questo tipo di inferenza potrebbe essere tratto anche da dati aggregati che rappresentano correlazioni tra elementi delle informazioni originali, piuttosto che dalle informazioni stesse. Ciò si verificherebbe, in particolare, per i modelli di intelligenza artificiale o i dati sintetici. È possibile trarre inferenze specifiche interrogando (o, nel caso dei modelli di intelligenza artificiale, sollecitando) i dati forniti con informazioni aggiuntive per ottenere nuove informazioni su un particolare individuo. Qualora tale inferenza sia anche significativa, ciò costituirebbe una violazione del criterio di non inferenza.

3.5 Applicazione dei criteri

3.5.1 Valutazione dell'efficacia delle tecniche di (re)identificazione

83 I tre criteri sopra descritti dovrebbero essere utilizzati per valutare l'efficacia delle tecniche di (re)identificazione applicabili ai dati in esame. Queste tecniche spaziano da quelle molto semplici, come una ricerca di base in altri dataset per la corrispondenza degli attributi, a quelle molto complesse, come l'utilizzo di agenti di intelligenza artificiale specifici per implementare una sofisticata combinazione di tecniche probabilistiche. A seconda della tecnica, queste possono essere utilizzate intenzionalmente o involontariamente e possono essere adattate per sfruttare potenziali vulnerabilità derivanti da contesti specifici o da difetti specifici nei metodi di anonimizzazione applicati.

84 Lo stato dell'arte di tali tecniche è in continua evoluzione e deve essere preso in considerazione durante questa valutazione. Il Comitato europeo per la protezione delle informazioni (EDPB) incoraggia pertanto lo sviluppo, la diffusione e l'applicazione di competenze scientificamente fondate su questo argomento. Tali competenze acquisiranno sempre maggiore importanza con l'aumentare della complessità dei metodi di (re)identificazione e/o del rischio legato alla loro corretta applicazione.

85 Sebbene non sia possibile stilare un elenco esaustivo, diversi fattori chiave saranno generalmente rilevanti nella valutazione dell'efficacia di queste tecniche:

- Sia che i dati siano aggregati o a livello di singolo record;
- La dimensionalità dei dati. Si riferisce alla quantità di informazioni disponibili su ciascun individuo (*per esempio*il numero di colonne in una tabella per ogni singolo record);
- La risoluzione dei dati. Ciò si riferisce al livello di dettaglio degli attributi registrati (*per esempio*una data di nascita completa ha una risoluzione maggiore rispetto al solo anno di nascita);
- La diversità dei dati. Ciò si riferisce alla quantità di differenze tra i dati (*per esempio*se molti individui condividono gli stessi valori degli attributi, i dati presentano una bassa diversità);
- Il numero di individui i cui dati sono inclusi nel record; e
- La quantità di informazioni aggiuntive disponibili per la combinazione con i dati.

86 Come regola generale, le tecniche di (re)identificazione hanno maggiori probabilità di essere efficaci contro dati a livello di record con elevata dimensionalità e alta risoluzione. Al contrario, molte tecniche di (re)identificazione hanno meno probabilità di essere efficaci contro dati che vengono aggregati tramite funzioni matematiche in indicatori statistici (*per esempio*(medie), sebbene ciò possa dipendere dalla quantità di dati aggregati. Questi fattori chiave sono in genere interdipendenti. Ad esempio, la quantità di informazioni aggiuntive necessarie affinché una tecnica di (re)identificazione sia efficace può essere molto inferiore per i dati a livello di record rispetto ai dati aggregati.

87 L'efficacia di una tecnica di (re)identificazione dovrebbe essere valutata in base alla sua capacità di generare risultati accurati. Ciò non richiede necessariamente una precisione perfetta o una certezza assoluta; piuttosto, il risultato dovrebbe essere almeno sufficientemente vicino e fornire almeno un livello di confidenza sufficientemente elevato, date le circostanze, da consentire di distinguere i soggetti dei dati e di trattarli in modo diverso. Inoltre, una tecnica di (re)identificazione può avere successo anche se risulta efficace solo nei confronti di un singolo individuo all'interno di un set di dati più ampio.[46].

Esempio 20: Un'entità utilizza un dataset contenente informazioni su tutti i dipendenti di una grande organizzazione per dedurre, con ragionevole precisione, lo stipendio dei dipendenti. Tuttavia, l'inferenza ha successo solo per un dipendente con un livello di confidenza significativo e l'entità non ha modo, se non per tentativi, di stabilire di chi sia stato correttamente dedotto lo stipendio. Poiché questa tecnica non può, con un livello di confidenza sufficientemente elevato, portare all'identificazione dell'individuo, non è sufficientemente efficace e il suo utilizzo non viola il criterio di non inferenza.

3.5.2 Valutazione dei criteri secondo l'approccio contestuale

88 Quando si utilizza l'approccio contestuale, è importante riconoscere che le capacità delle diverse entità possono variare nel tempo. Ad esempio, una fuga di dati potrebbe fornire a determinate entità l'accesso a nuovi set di dati che in precedenza non erano a loro disposizione. Ciò consentirebbe a sua volta a tali entità di collegare le informazioni con mezzi che, almeno dal punto di vista di tale entità, in precedenza non erano disponibili ma che ora è ragionevolmente probabile che vengano utilizzati, portando a un

[46]Vedi paragrafo 36.

violazione del criterio di non collegamento. Per garantire la corretta protezione degli interessati, il Comitato europeo per la protezione dei dati (EDPB) raccomanda pertanto che le analisi effettuate secondo l'approccio contestuale includano, per quanto possibile, margini di sicurezza adeguati, assicurando che la valutazione rimanga solida man mano che le capacità degli enti si evolvono.

89 Un altro aspetto importante da considerare nell'ambito dell'approccio contestuale è l'accessibilità dei dati. Se i dati sono pubblicamente disponibili, chiunque può accedervi con mezzi che è ragionevolmente probabile vengano utilizzati. Se i dati sono soggetti a controlli di sicurezza adeguati (in particolare per quanto riguarda la riservatezza dei dati), l'accesso ai dati costituirà un mezzo che è ragionevolmente probabile venga utilizzato solo dagli utenti autorizzati (e non da soggetti non autorizzati). Ciò è rilevante sia per i dati stessi sia per qualsiasi informazione aggiuntiva necessaria che potrebbe essere utilizzata nell'ambito del collegamento o dell'inferenza.

Esempio 21: Tornando all'Esempio 16, il rapporto interno dell'azienda è una parte necessaria per dedurre lo stipendio del sesto ingegnere. Quando si applica l'approccio contestuale, è quindi importante verificare se ogni entità rilevante ha accesso a tale rapporto con mezzi che è ragionevolmente probabile che vengano utilizzati. Se tale rapporto interno è pubblico, si dovrebbe presumere che tutti possano accedervi e che il criterio di non inferenza sia violato dal punto di vista di tutti. Se, tuttavia, l'accesso a tale rapporto interno è limitato, e supponendo che non ci sia altro modo per ottenere queste informazioni aggiuntive, il criterio di non inferenza sarà violato solo dalle prospettive di coloro che effettivamente vi hanno accesso.

---

90 Le limitazioni all'accesso devono essere valutate in modo olistico, tenendo conto delle capacità dei diversi soggetti e di eventuali elementi contestuali che possano avere un impatto maggiore o minore sulle modalità di accesso che questi probabilmente utilizzeranno. Le misure tecniche, organizzative e contrattuali possono contribuire a limitare l'accesso, ma non sono di per sé sufficienti a rendere i dati anonimi; sebbene i titolari del trattamento siano incoraggiati ad adottare tali misure ove opportuno, dovrebbero comunque esaminare le capacità di (re)identificazione di chiunque possa accedere ai dati con modalità ragionevolmente probabili.

91 Una crittografia opportunamente implementata può essere utilizzata per limitare l'accesso ai dati (o, quantomeno, limitare l'accesso a una forma dei dati intelligibile) e, in questo modo, può contribuire a un'anonimizzazione efficace. Tuttavia, la crittografia, per sua stessa natura, è reversibile e si differenzia dall'anonimizzazione. È quindi importante valutare se i soggetti che hanno accesso ai dati possano comunque decifrarli con mezzi ragionevolmente probabili, anche se al momento non possiedono la chiave di decrittazione (ad esempio, tramite crittanalisi).

92 La maggior parte delle tecniche di (re)identificazione attualmente documentate stanno diventando sempre più accessibili, richiedono risorse limitate e possono essere eseguite entro un lasso di tempo ragionevole, anche su hardware commerciale. Ciò vale soprattutto per la re-identificazione dei dati a livello di record. Alcune tecniche che operano su dati aggregati possono richiedere maggiori risorse computazionali, ma l'accesso a tali risorse è comunque possibile come servizio, a costi relativamente accessibili. In genere, tempo e costi non sarebbero proibitivi per l'esecuzione di queste tecniche esistenti, sebbene possano rappresentare una sfida per (*per esempio*) ottenendo le informazioni aggiuntive necessarie per utilizzare con successo tale tecnica. Il Comitato europeo per la protezione dei dati (EDPB) osserva inoltre che gli sviluppi nell'IA, e in particolare nell'IA agentiva, probabilmente ridurranno ulteriormente i tempi e i costi necessari per accedere e implementare queste tecniche.

93 L'effetto del tempo e del costo sui mezzi ragionevolmente utilizzabili può quindi essere particolarmente evidente laddove le informazioni aggiuntive necessarie per la (re)identificazione siano distribuite su più luoghi o entità, oppure siano difficili da individuare o da consultare. Tuttavia, il Comitato europeo per la protezione dei dati (EDPB) osserva nuovamente che gli sviluppi nell'IA possono anche ridurre il tempo associato alla raccolta e

combinando informazioni provenienti da più fonti (almeno nei casi in cui tali fonti siano pubblicamente disponibili), il che può portare ad aumenti significativi dei mezzi che è ragionevolmente probabile che vengano utilizzati.

94 L'esempio seguente illustra a grandi linee come si potrebbe valutare il criterio di non collegamento utilizzando l'approccio contestuale.

Esempio 22: L'istituto di ricerca descritto nell'Esempio 12 decide di utilizzare l'approccio contestuale per valutare l'anonimato dei dati contenuti nella tabella dopo aver eliminato i dati personali originali.

L'istituto di ricerca inizia identificando chiunque potrebbe, attraverso mezzi ragionevolmente utilizzabili, essere in grado di accedere alla tabella stessa. La tabella non è stata resa pubblica e l'istituto di ricerca attualmente impiega solide misure di sicurezza informatica all'avanguardia, nonché altri mezzi tecnici per limitare l'accesso non autorizzato a questo set di dati e impedire ai suoi membri di crearne copie. Per precauzione, l'istituto di ricerca ritiene che, soprattutto nel lungo periodo, sia possibile che i dati vengano comunque divulgati. Ciò potrebbe includere, tra l'altro, fughe di dati dovute a errori umani o tentativi di hacking da parte di soggetti esterni. L'istituto di ricerca conclude pertanto che la tabella è accessibile, attraverso mezzi ragionevolmente utilizzabili, da almeno:

1. membri dell'istituto di ricerca a cui è stato concesso l'accesso al tavolo,
2. individui che potrebbero ricevere accidentalmente i dati a causa di un errore umano,
3. individui che potrebbero violare il sistema e rubare e distribuire i dati,
4. autorità governative o giudiziarie che potrebbero accedere ai dati nel corso dell'adempimento delle loro funzioni pubbliche, e
5. Altre entità che potrebbero ottenere i dati forniti da qualcuno appartenente a uno dei quattro gruppi sopracitati.

Tuttavia, l'istituto di ricerca conclude che è estremamente improbabile che persone a conoscenza della vita privata dei soggetti interessati ricevano i dati in seguito a una fuga di notizie, evento peraltro già di per sé poco probabile. La stessa conclusione si applica a tutte le altre persone che potrebbero essere a conoscenza della partecipazione di un individuo ad attività che portano all'inclusione di dati che lo riguardano nei dati in questione. Pertanto, l'istituto di ricerca conclude che queste persone non possono essere considerate soggetti rilevanti.

L'istituto di ricerca valuta quindi se sia ragionevolmente probabile che i mezzi a disposizione di ciascun ente interessato vengano utilizzati, tenendo conto dei fattori oggettivi indicati al paragrafo 29.

È già stato concluso che i dati possono essere collegati ad altri dati relativi agli stessi individui da qualsiasi soggetto che conosca (a) il sesso, l'anno di nascita e il codice postale generalizzato dell'individuo e (b) che l'individuo è stato incluso nei dati in questione. Sebbene queste informazioni possano essere note, ad esempio, agli amici intimi e ai familiari degli individui, nessuno dei soggetti rilevanti che sia a conoscenza dell'inclusione di un individuo nei dati in questione rientra tra i soggetti interessati. Ciò include i membri dell'istituto di ricerca, poiché i suoi membri non hanno avuto interazioni dirette con i pazienti e non sono stati coinvolti nel loro trattamento, e non esiste alcuna registrazione di quali pazienti siano stati inclusi nel dataset.

In conclusione, nessuna entità rilevante saprebbe che i dati di un determinato individuo sono stati inclusi nel dataset. Inoltre, non è possibile dedurre dal dataset stesso che i dati di un determinato individuo sono stati inclusi nel dataset. Queste considerazioni dipendono dal contesto specifico del caso dato e devono essere supportate da un ragionamento accurato. Tuttavia, in questo

instance, the research institute is able to conclude that no relevant entity is likely to have access to the additional information that can be linked to the given data and that the No Linkage criterion has been met.

3.5.3 Compiling the results

95 If the given data fulfil all three criteria under either the simplified or contextual approaches, they can be presumed anonymous. If, however, one or more criteria is violated, further analysis may be needed to see if the information should be considered as personal data.

96 If the data is found to be personal under the simplified approach, it is also possible to perform a further assessment through the contextual approach. If none of the three criteria are violated from the relevant perspective, then the information can still be considered anonymous from that perspective. In some cases, this finding of anonymity may be limited to the perspective of only one entity, while in others it may extend to all of the relevant entities. As previously stated, the goal of anonymisation should be to ensure that the resulting data is anonymous for all of the relevant entities.

97 This is shown in the following table:

	Criterion is satisfied under either approach	**Criterion is violated under simplified approach**	**Criterion is violated under contextual approach**
No Record Isolation	The data can be considered anonymous if the other two criteria are also passed.	The data may not be anonymous. See paras. 98 to 00 below. To reach a conclusive result, consider continuing to the contextual approach.	The data may not be anonymous from at least one of the relevant perspectives. See paras. 98 to 100 below.
No Linkage	The data can be considered anonymous if the other two criteria are also passed.	The data may not be anonymous. See paras. 101 to 103 below. To reach a conclusive result, consider continuing to the contextual approach.	The data may not be anonymous from at least one of the relevant perspectives. See paras. 101 to 103 below.
No Inference	The data can be considered anonymous if the other two criteria are also passed.	The data should not be considered anonymous under the simplified approach, consider continuing to the contextual approach.	The data should not be considered anonymous from at least one of the relevant perspectives.

98 If the data violates the No Record Isolation criterion, it should be checked if the isolated records allow for singling out, so that individuals are distinguished from others by matching those records with the individual's corresponding attributes. These corresponding attributes could,

for example, become apparent to the relevant entity in the course of their interactions with the individual, or may be available through additional information that allows for the individual to be identified[47]. The larger the isolated record and the more attributes it contains, the easier the singling out of the individuals becomes. If the individual can be singled out in this way, then they should be considered identifiable.

Example 23: Returning to the dataset from Example 10 although the dataset violated the No Record Isolation criterion in several places, this does not necessarily mean that the individuals can be identified by isolating their records. However, in this case it is very likely: The records are unique, and their values are likely to be unique for most individuals in this population.

Further, the information needed to identify the individuals by reference to the demographic attributes may be available through means reasonably likely to be used (*e.g.* through public registers). If this is possible, the given data would be personal.

99 If this singling out cannot be done – and the other two criteria have been successfully met – then the given data can still be considered anonymous, despite failing the No Record Isolation criterion. This could happen if, among other things, the record is unique (and so the No Record Isolation criterion is failed) but the attributes cannot be matched to any other data available to the relevant entity or if the unique record is not intelligible.

Example 24: A cinema decides to run a study to better understand its clients' preferences. The survey is answered by a very large number of customers, and only contains questions related to the movie itself, that were written specifically for this survey and that do not reveal anything that can be reasonably linked back to the respondent. The responses are then collected in a secret ballot, and later turned into a dataset:

Survey number	Question 1	Question 2	...	Question 30
1	D	C	…	B
2	C	A	...	A
3	…	…	…	...

Because of the large number of questions, it is very likely that at least some (if not all) records are unique in the dataset, violating the No Record Isolation criterion. However, this does not mean that participants can be identified. In fact, given the circumstances, there is no way to link the answers to their identity or to other information relating to them using means reasonably likely to be used, nor to treat any respondent differently based on their individual answers. If the data passes the other two criteria, it is, therefore, anonymous, despite failing the No Records Isolation criterion.

100 If this singling out can be done, the information should not be considered anonymous. However, if the analysis has progressed under the simplified approach (*i.e.* testing whether an entity can single out an individual on the basis of the given data, while assuming that if means reasonably likely to be used exist then at least someone will be able to use them), this

[47] See, for example, C-479/22 P *OC v European Commission*, paras. 60 & 63, where the CJEU looked at a unique combination of attribute values which could then be compared against the corresponding attributes available online.

L'analisi potrebbe essere ulteriormente ampliata utilizzando l'approccio contestuale, come descritto nei paragrafi 48 e 49, e seguendo le linee guida stabilite nella sottosezione 3.5.2 di cui sopra.

101 Se i dati violano il criterio di non collegamento, occorre verificare se i dati collegati possono comunque condurre direttamente all'identificazione dell'individuo.

Esempio 25: Tornando all'Esempio 12, il criterio di non collegamento verrebbe violato se le informazioni contenute nei dati forniti potessero essere collegate a un altro set di dati che includesse il sesso, l'anno di nascita e il codice postale generalizzato dello stesso individuo.

Poiché il sesso, l'anno di nascita e il codice postale generico di un individuo non sono univoci, non è possibile distinguere gli individui e identificarli. Tuttavia, se il secondo set di dati contiene attributi aggiuntivi relativi allo stesso individuo che ne consentirebbero l'identificazione (o se contiene un identificativo diretto per tale individuo), allora il collegamento porterebbe direttamente all'identificazione dell'individuo. In tali casi, i dati forniti sarebbero considerati dati personali.

102 Se i dati forniti possono essere collegati a un altro dataset, le nuove informazioni (collegate) devono essere nuovamente testate rispetto ai tre criteri. Se le informazioni collegate superano i criteri di assenza di isolamento dei record, assenza di (ulteriori) collegamenti e assenza di inferenze, possono essere considerate anonime.

103 Se le nuove informazioni (collegate) possono portare direttamente all'identificabilità dell'individuo, le informazioni non devono essere considerate anonime. Tuttavia, se l'analisi è progredita secondo l'approccio semplificato (*cioè*verificando se le informazioni collegate possono portare all'identificabilità di un individuo, presupponendo che se esistono mezzi ragionevolmente probabili da utilizzare, almeno qualcuno sarà in grado di utilizzarli), questa analisi potrebbe essere ulteriormente estesa utilizzando l'approccio contestuale, come descritto nei paragrafi 48 e 49, e seguendo le linee guida stabilite nella sottosezione 3.5.2 di cui sopra.

4 Glossario

Per rendere queste linee guida più facili da leggere, vengono utilizzati diversi termini che non hanno una definizione legale. Questi termini sono inclusi nel presente glossario.

Dati aggregati Dati relativi a gruppi di persone e che non contengono informazioni individuali.

Controller anonimizzato: Un responsabile che esegue l'anonimizzazione e ne valuta l'esito.

Prospettiva applicabile: La prospettiva da cui valutare la natura personale di alcuni dati in possesso di una specifica entità, in base alle circostanze concrete del caso. Può essere la prospettiva dell'entità stessa o quella di un'altra entità, ad esempio, nel caso di trattamento per conto di tale altra entità.

Entità: Una persona fisica o giuridica, un'autorità pubblica, un'agenzia o altro ente. A seconda che le informazioni siano personali o meno, potrebbero essere, tra gli altri, un potenziale titolare del trattamento, un potenziale responsabile del trattamento o un potenziale terzo in possesso di informazioni aggiuntive necessarie per identificare l'interessato.

Dati forniti Dati di cui si deve valutare l'anonimato.

IndividualePersona fisica, identificabile o meno, alla quale le informazioni contenute nei dati forniti possono o meno riferirsi.

Prospettiva: Ogni entità ha un**prospettiva**, che dipenderà dalle loro circostanze e informa sui mezzi che è ragionevolmente probabile che vengano utilizzati da tale entità. Se un'entità ha o può ottenere mezzi che è ragionevolmente probabile che vengano utilizzati per identificare la persona fisica a cui si riferiscono le informazioni, si può dire che le informazioni sono dati personali dal suo punto di vista.

Documentazione: Tutti i valori dei dati (incluse le immagini)*ecc.*.) all'interno dei dati forniti che si sa essere relativi allo stesso individuo, indipendentemente dal fatto che tale individuo sia identificabile o meno. Ad esempio, in un dataset contenente foto di volti, un record potrebbe essere costituito da una, diverse o tutte le foto che mostrano la stessa persona. Di solito, i record relativi a persone diverse sono chiaramente distinguibili. Tuttavia, in alcuni casi la stessa informazione può essere relativa a più di un individuo e quindi può anche appartenere a record diversi. Per chiarezza, questo termine non viene utilizzato nel senso di "record di database". Ai fini di queste linee guida, un record può quindi essere costituito da dati che (ad esempio) coprono diverse righe in una tabella, a condizione che ciascuna di queste righe si riferisca allo stesso individuo.

Entità ricevente: Un'entità che riceve i dati forniti.

Dati a livello di record Dati strutturati in modo da distinguere facilmente i singoli record.

ente: Le entità sono**pertinente**per valutare l'anonimato dal punto di vista di una particolare entità, qualora potessero contribuire all'identificazione diretta o indiretta da parte di tale entità.

Reidentificazione: Un processo in cui si scopre che i dati, ritenuti correttamente anonimizzati, sono in realtà personali, in quanto permette di identificare l'individuo a cui si riferiscono.

Allegato 1: Diagramma di flusso per l'analisi tecnica di anonimato

Questo diagramma di flusso è stato sviluppato come strumento di supporto per le organizzazioni che intendono anonimizzare i dati. In modo semplificato, presenta una possibile modalità di valutazione in conformità con le linee guida.