Standard bezpieczeństwa

STANDARDY OTWARTOŚCI DANYCH

Fundusze Europejskie Polska Cyfrowa
Ministerstwo Cyfryzacji
OTWARTE DANE
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

Spis treści

Wstęp

Otwieranie danych publicznych zwiększa transparentność działań administracji oraz zakres kontroli nad działalnością państwa przez obywateli, a także umożliwia ich dalsze wykorzystywanie w produktach, aplikacjach czy usługach, w tym dla celów naukowych czy biznesowych. Rekomendowaną praktyką dla dysponentów danych jest organizowanie konsultacji z środowiskami interesariuszy (sektor badań i rozwoju, naukowcy, firmy komercyjne, organizacje pozarządowe) w celu określenia oczekiwanej zawartości informacyjnej zbiorów. Na podstawie takiego zapotrzebowania oraz źródłowego zakresu danych osobowych można przeprowadzać procedurę otwierania danych publicznych.

Otwieranie danych publicznych do ponownego wykorzystywania może jednocześnie rodzić ryzyko wkroczenia w sferę autonomii informacyjnej jednostek, dlatego istotne jest zapewnienie ochrony danych osobowych tak przez podmioty udostępniające lub przekazujące dane publiczne do ponownego wykorzystywania, jak i użytkowników tych danych. Przez dane osobowe rozumie się informacje o zidentyfikowanej lub możliwej do zidentyfikowania osobie fizycznej („osobie, której dane dotyczą"). Możliwa do zidentyfikowania osoba fizyczna to osoba, którą można bezpośrednio lub pośrednio zidentyfikować, w szczególności na podstawie identyfikatora takiego jak: imię i nazwisko, numer identyfikacyjny, dane o lokalizacji, identyfikator internetowy lub jeden bądź kilka szczególnych czynników określających fizyczną, fizjologiczną, genetyczną, psychiczną, ekonomiczną, kulturową lub społeczną tożsamość osoby fizycznej.

Dobór odpowiednich technik – w kontekście udostępniania danych z zasobów publicznych – pozwalających na zapewnienie prywatności, ale umożliwiających zachowanie wartości informacyjnych danego zbioru i potencjału dla ponownego wykorzystywania, jest w tym wypadku kwestią kluczową.

[1] W rozumieniu art. 2 ust. 2 ustawy z dnia 25 lutego 2016 r. o ponownym wykorzystywaniu informacji sektora publicznego, dalej: upw.

[2] W rozumieniu Rozporządzenie Parlamentu Europejskiego i Rady (UE) 2016/679 z dnia 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych), dalej: RODO.

1.

Minimalne czynności dotyczące udostępnienia danych do ponownego wykorzystywania

Opisane czynności dotyczą podstawowych zagadnień, jakie powinien wziąć pod uwagę dysponent danych (na gruncie przepisów o ochronie danych osobowych jest to administrator) udostępniający je do ponownego wykorzystywania. Istotną częścią jest ocena skutków dla ochrony danych osobowych, której elementem jest analiza ryzyka. Przeprowadzenie poniższych czynności jest kluczowe z punktu widzenia spełnienia wymogu rozliczalności. Pomoże on dysponentom danych w podjęciu decyzji o udostępnieniu danych osobowych do ponownego wykorzystywania.

Nr	Czynność	Załącznik, którego dotyczy czynność	Działanie	Czy czynność realizowana jest zawsze?
1.	Analiza zawartości rejestru/zasobu/ bazy danych	-	• Dostarczenie informacji czy rejestr/ zasób/baza danych, która ma potencjał dla ponownego wykorzystywania zawiera dane osobowe; • Jeśli rejestr/zasób/baza nie zawiera danych osobowych, dysponent może udostępnić dane publiczne do ponownego wykorzystywania; • Jeśli rejestr/zasób/zbiór danych publicznych zawiera dane osobowe konieczne jest przejście do kolejnych czynności.	Tak
2.	Określenie przypadku w jakim będzie dochodzić do ujawniania danych publicznych	Załącznik nr 1	• Przesądzenie podstawy ujawniania danych osobowych, która jest związana z kategorią danych i kategorią osób, których dotyczy ujawnienie danych publicznych; • Przesądzenie czy ujawnianie danych publicznych wymagać będzie anonimizacji; • Przesądzenie czy potrzebne będzie przeprowadzenie testu zgodności celów (art. 6 ust. 4 RODO[3]);	Tak

[3] Art. 6 ust. 4 Jeżeli przetwarzanie w celu innym niż cel, w którym dane osobowe zostały zebrane, nie odbywa się na podstawie zgody osoby, której dane dotyczą, ani prawa Unii lub prawa państwa członkowskiego stanowiących w demokratycznym społeczeństwie niezbędny i proporcjonalny środek służący zagwarantowaniu celów, o których mowa w art. 23 ust. 1, administrator - aby ustalić, czy przetwarzanie w innym celu jest zgodne z celem, w którym dane osobowe zostały pierwotnie zebrane - bierze pod uwagę między innymi:
a) wszelkie związki między celami, w których zebrano dane osobowe, a celami zamierzonego dalszego przetwarzania;
b) kontekst, w którym zebrano dane osobowe, w szczególności relację między osobami, których dane dotyczą, a administratorem;
c) charakter danych osobowych, w szczególności czy przetwarzane są szczególne kategorie danych osobowych zgodnie z art. 9 lub dane osobowe dotyczące wyroków skazujących i czynów zabronionych zgodnie z art. 10;
d) ewentualne konsekwencje zamierzonego dalszego przetwarzania dla osób, których dane dotyczą;
e) istnienie odpowiednich zabezpieczeń, w tym ewentualnie szyfrowania lub pseudonimizacji.

			• Przesądzenie jakie operacje będą realizowane w ramach ujawniania danych publicznych. Uwaga! Jeżeli informacja, która ma zostać ujawniona obejmuje dane osobowe osób innych niż wskazanych w art. 6 ust. 2[4] upw (tj., gdy wymagane będzie przeprowadzenie testu zgodności celów) dopiero ocena zgodności celów (**załącznik nr 4**) i ocena ryzyka (**załącznik nr 5**) ostatecznie potwierdzi czy przyjęte założenie o możliwości ujawnienia danych osobowych jest prawidłow	Tak
3.	Weryfikacja aktualności i kompletności wykazu typowych (źródeł) ryzyk dotyczących operacji na danych, które będą miały zastosowanie do przypadku zakwalifikowanego w oparciu o załącznik nr 1	Załącznik nr 2	• Zapewnienie rozliczalności w zakresie analizy ryzyka naruszenia praw lub wolności	Tak

[4] Art. 6 ust. 2. Prawo do ponownego wykorzystywania podlega ograniczeniu ze względu na prywatność osoby fizycznej lub tajemnicę przedsiębiorcy. Ograniczenie to nie dotyczy informacji o osobach pełniących funkcje publiczne, mających związek z pełnieniem tych funkcji, w tym o warunkach powierzenia i wykonywania funkcji, oraz przypadku gdy osoba fizyczna lub przedsiębiorca rezygnują z przysługującego im prawa.

4.	Określenie tła analizy ryzyka	Załącznik nr 3	• Spełnienie wymogu, o którym mowa w art. 25 ust. 1 RODO[5] oraz art. 35 ust. 1 RODO[6] w zakresie określenia celu, zakresu, charakteru i kontekstu przetwarzania danych; • Sporządzenie systematycznego opisu planowanych operacji przetwarzania i celów przetwarzania (art. 35 ust. 7 lit. a) RODO)[7]; • Weryfikacja wstępna czy wymagane będzie przeprowadzenie oceny skutków w oparciu o wykaz Prezesa UODO operacji przetwarzania wymagających oceny skutków dla ochrony danych. Uwaga! Dany zasób może obejmować różne kategorie danych np. osób pełniących oraz niepełniących funkcji publicznych.	Tak

[5]Art. 25 ust. 1. Uwzględniając stan wiedzy technicznej, koszt wdrażania oraz charakter, zakres, kontekst i cele przetwarzania oraz ryzyko naruszenia praw lub wolności osób fizycznych o różnym prawdopodobieństwie wystąpienia i wadze wynikające z przetwarzania, administrator – zarówno przy określaniu sposobów przetwarzania, jak i w czasie samego przetwarzania –wdraża odpowiednie środki techniczne i organizacyjne, takie jak pseudonimizacja, zaprojektowane w celu skutecznej realizacji zasad ochrony danych, takich jak minimalizacja danych, oraz w celu nadania przetwarzaniu niezbędnych zabezpieczeń, tak by spełnić wymogi niniejszego rozporządzenia oraz chronić prawa osób, których dane dotyczą.

[6]Art. 35 ust. 1. Jeżeli dany rodzaj przetwarzania – w szczególności z użyciem nowych technologii – ze względu na swój charakter, zakres, kontekst i cele z dużym prawdopodobieństwem może powodować wysokie ryzyko naruszenia praw lub wolności osób fizycznych, administrator przed rozpoczęciem przetwarzania dokonuje oceny skutków planowanych operacji przetwarzania dla ochrony danych osobowych. Dla podobnych operacji przetwarzania danych wiążących się z podobnym wysokim ryzykiem można przeprowadzić pojedynczą ocenę.

[7]Art. 35 ust. 7 lit. a) Ocena zawiera co najmniej systematyczny opis planowanych operacji przetwarzania i celów przetwarzania, w tym, gdy ma to zastosowanie – prawnie uzasadnionych interesów realizowanych przez administratora.

5.	Przeprowadzenie testu zgodności celów	Załącznik nr 4	• Spełnienie wymogu testu zgodności celów, o którym mowa w art. 6 ust. 4 RODO[8];	Nie. Dotyczy wyłącznie przypadku, gdy prowadzony będzie test zgodności celów. Nie dotyczy to: - ujawniania danych anonimowych; - osób pełniących funkcje publiczne w związku z ich pełnieniem; - osób, które wyraziły zgodę na przetwarzanie danych w celu ponownego wykorzystania/ zrezygnowały z prawa do prywatności na gruncie art. 6 ust. 2upw
6.	Przeprowadzenie oceny ryzyka naruszenia praw lub wolnośc	Załącznik nr 5	• Spełnienie wymogu uwzględnienia ochrony danych osobowych w fazie projektowania (art. 25 ust. 1 RODO[9]) oraz oceny skutków dla ochrony danych (art. 35 ust. 1 RODO[10]). Uwaga!	Tak – także w przypadku ujawniania danych anonimowych. W przypadku, gdy przeprowadzana jest ocena zgodności celów, przeprowadzona ocena ryzyka może potwierdzić możliwość ujawnienia danych osobowych lub to wykluczyć.

[8] Art. 6 ust. 4. Jeżeli przetwarzanie w celu innym niż cel, w którym dane osobowe zostały zebrane, nie odbywa się na podstawie zgody osoby, której dane dotyczą, ani prawa Unii lub prawa państwa członkowskiego stanowiących w demokratycznym społeczeństwie niezbędny i proporcjonalny środek służący zagwarantowaniu celów, o których mowa w art. 23 ust. 1, administrator - aby ustalić, czy przetwarzanie w innym celu jest zgodne z celem, w którym dane osobowe zostały pierwotnie zebrane - bierze pod uwagę między innymi:
a) wszelkie związki między celami, w których zebrano dane osobowe, a celami zamierzonego dalszego przetwarzania;
b) kontekst, w którym zebrano dane osobowe, w szczególności relację między osobami, których dane dotyczą, a administratorem;
c) charakter danych osobowych, w szczególności czy przetwarzane są szczególne kategorie danych osobowych zgodnie z art. 9 lub dane osobowe dotyczące wyroków skazujących i czynów zabronionych zgodnie z art. 10;
d) ewentualne konsekwencje zamierzonego dalszego przetwarzania dla osób, których dane dotyczą;
e) istnienie odpowiednich zabezpieczeń, w tym ewentualnie szyfrowania lub pseudonimizacji.

[9] Art. 25 ust. 1. Uwzględniając stan wiedzy technicznej, koszt wdrażania oraz charakter, zakres, kontekst i cele przetwarzania oraz ryzyko naruszenia praw lub wolności osób fizycznych o różnym prawdopodobieństwie wystąpienia i wadze wynikające z przetwarzania, administrator – zarówno przy określaniu sposobów przetwarzania, jak i w czasie samego przetwarzania –wdraża odpowiednie środki techniczne i organizacyjne, takie jak pseudonimizacja, zaprojektowane w celu skutecznej realizacji zasad ochrony danych, takich jak minimalizacja danych, oraz w celu nadania przetwarzaniu niezbędnych zabezpieczeń, tak by spełnić wymogi niniejszego rozporządzenia oraz chronić prawa osób, których dane dotyczą.

[10] Art. 35 ust. 1. Jeżeli dany rodzaj przetwarzania – w szczególności z użyciem nowych technologii – ze względu na swój charakter, zakres, kontekst i cele z dużym prawdopodobieństwem może powodować wysokie ryzyko naruszenia praw lub wolności osób fizycznych, administrator przed rozpoczęciem przetwarzania dokonuje oceny skutków planowanych operacji przetwarzania dla ochrony danych osobowych. Dla podobnych operacji przetwarzania danych wiążących się z podobnym wysokim ryzykiem można przeprowadzić pojedynczą ocenę.

			Uwaga! W ramach oceny ryzyka należy rozważyć możliwość wystąpienia każdego typowego źródła ryzyka (załącznik nr 2) a także źródeł, które są specyficzne w danym stanie faktycznym.	
7.	Przeprowadzenie testu proporcjonalności i niezbędności	Załącznik nr 6	Spełnienie wymogu oceny czy operacje przetwarzania są niezbędne oraz proporcjonalne w stosunku do celów jako elementu oceny skutków dla ochrony danych, o którym mowa w 35 ust. 7 lit. b) RODO[11];	Nie. Dotyczy sytuacji, gdy konieczne jest przeprowadzenie oceny skutków. Potrzeba oceny skutków może zostać ujawniona już na początku ustalenia tła oceny ryzyka (na poziomie **załącznika nr 3**), gdy operacje na danych spełniają kryteria wskazane w wykazie Prezesa UODO operacji wymagających oceny skutków albo po przeprowadzeniu oceny ryzyka (**załącznik nr 5**) – gdy ocena ta wykaże, że istnieje wysokie ryzyko naruszenia praw lub wolności.
8.	Rejestracja przeprowadzanych czynności	Załącznik nr 7	Spełnienie wymogu rozliczalności (art. 5 ust. 2 RODO[12])	Tak

Niezależnie od realizacji czynności opisanych w tabeli, zadaniem dysponenta jest śledzenie osiągnięć w dziedzinie zabezpieczania systemów informatycznych i wdrażanie takich narzędzi oraz sposobów zarządzania danymi, które zapewnią bezpieczeństwo danych osobowych.

[11] Art. 35 ust. 7 lit. b) Ocena zawiera co najmniej ocenę, czy operacje przetwarzania są niezbędne oraz proporcjonalne w stosunku do celów.

[12] Art. 5 ust. 2. Administrator jest odpowiedzialny za przestrzeganie przepisów ust. 1 i musi być w stanie wykazać ich przestrzeganie („rozliczalność").

2.

Grupy danych, których ponowne wykorzystywanie jest ułatwione

a) Ułatwienia dla dysponenta danych

Dysponent, który udostępnia dane publiczne zawierające dane osobowe do ponownego wykorzystywania jest zwolniony ze spełnienia obowiązku informacyjnego, o którym mowa w art. 13 ust. 3 RODO[13].

b) Ułatwienia dla użytkownika

Użytkownik zwolniony jest z konieczności spełnienia obowiązku informacyjnego, o którym mowa w art. 14 ust. 1-4 RODO[14] w przypadku, w którym ponownie wykorzystuje następujące kategorie danych:

[13] Art. 13 ust. 3. Jeżeli administrator planuje dalej przetwarzać dane osobowe w celu innym niż cel, w którym dane osobowe zostały zebrane, przed takim dalszym przetwarzaniem informuje on osobę, której dane dotyczą, o tym innym celu oraz udziela jej wszelkich innych stosownych informacji, o których mowa w ust. 2.

[14] Art. 14. 1. Jeżeli danych osobowych nie pozyskano od osoby, której dane dotyczą, administrator podaje osobie, której dane dotyczą, następujące informacje:
a) swoją tożsamość i dane kontaktowe oraz, gdy ma to zastosowanie, tożsamość i dane kontaktowe swojego przedstawiciela;
b) gdy ma to zastosowanie – dane kontaktowe inspektora ochrony danych;
c) cele przetwarzania, do których mają posłużyć dane osobowe, oraz podstawę prawną przetwarzania;
d) kategorie odnośnych danych osobowych;
e) informacje o odbiorcach danych osobowych lub o kategoriach odbiorców, jeżeli istnieją;
f) gdy ma to zastosowanie – informacje o zamiarze przekazania danych osobowych odbiorcy w państwie trzecim lub organizacji międzynarodowej oraz o stwierdzeniu lub braku stwierdzenia przez Komisję odpowiedniego stopnia ochrony lub w przypadku przekazania, o którym mowa w art. 46, art. 47 lub art. 49 ust. 1 akapit drugi, wzmiankę o odpowiednich lub właściwych zabezpieczeniach oraz informację o sposobach uzyskania kopii tych zabezpieczeń lub o miejscu ich udostępnienia.
2. Poza informacjami, o których mowa w ust. 1, administrator podaje osobie, której dane dotyczą, następujące informacje niezbędne do zapewnienia rzetelności i przejrzystości przetwarzania wobec osoby, której dane dotyczą:
a) okres, przez który dane osobowe będą przechowywane, a gdy nie jest to możliwe, kryteria ustalania tego okresu;
b) jeżeli przetwarzanie odbywa się na podstawie art. 6 ust. 1 lit. f) – prawnie uzasadnione interesy realizowane przez administratora lub przez stronę trzecią;
c) informacje o prawie do żądania od administratora dostępu do danych osobowych dotyczących osoby, której dane dotyczą, ich sprostowania, usunięcia lub ograniczenia przetwarzania oraz o prawie do wniesienia sprzeciwu wobec przetwarzania, a także o prawie do przenoszenia danych;
d) jeżeli przetwarzanie odbywa się na podstawie art. 6 ust. 1 lit. a) lub art. 9 ust. 2 lit. a) – informacje o prawie do cofnięcia zgody w dowolnym momencie bez wpływu na zgodność z prawem przetwarzania, którego dokonano na podstawie zgody przed jej cofnięciem;
e) informacje o prawie wniesienia skargi do organu nadzorczego;
f) źródło pochodzenia danych osobowych, a gdy ma to zastosowanie – czy pochodzą one ze źródeł publicznie dostępnych;
g) informacje o zautomatyzowanym podejmowaniu decyzji, w tym o profilowaniu, o którym mowa w art. 22 ust. 1 i 4, oraz – przynajmniej w tych przypadkach – istotne informacje o zasadach ich podejmowania, a także o znaczeniu i przewidywanych konsekwencjach takiego przetwarzania dla osoby, której dane dotyczą.
3. Informacje, o których mowa w ust. 1 i 2, administrator podaje:
a) w rozsądnym terminie po pozyskaniu danych osobowych – najpóźniej w ciągu miesiąca – mając na uwadze konkretne okoliczności przetwarzania danych osobowych;
b) jeżeli dane osobowe mają być stosowane do komunikacji z osobą, której dane dotyczą – najpóźniej przy pierwszej takiej komunikacji z osobą, której dane dotyczą; lub

- osób pełniących funkcje publiczne mających związek z pełnieniem tych funkcji, w tym o warunkach powierzenia i wykonywania tych funkcji,
- osób fizycznych reprezentujących osoby prawne, w tym ich dane kontaktowe,
- obejmujących nazwę (firmę), numer identyfikacji podatkowej (NIP) albo imię i nazwisko kontrahenta podmiotu zobowiązanego

c) Dane podlegające nieograniczonemu ponownemu wykorzystywaniu:

- dane wykorzystywane na podstawie rezygnacji osoby fizycznej lub przedsiębiorcy z przysługującego im prawa,
- dane zanonimizowane,
- publicznie dostępne dane statystyczne, pozyskiwane np. w oparciu o ustawę o statystyce publicznej.

c) jeżeli planuje się ujawnić dane osobowe innemu odbiorcy – najpóźniej przy ich pierwszym ujawnieniu.
4. Jeżeli administrator planuje dalej przetwarzać dane osobowe w celu innym niż cel, w którym te dane zostały pozyskane, przed takim dalszym przetwarzaniem informuje on osobę, której dane dotyczą, o tym innym celu oraz udziela jej wszelkich innych stosownych informacji, o których mowa w ust. 2.

3.

Dane podlegające anonimizacji i pseudonimizacji oraz sposoby doboru technik

1. Przykłady zastosowania technik dla grup danych

a) Dane osobowe

Dane osobowe niedostępne publicznie zgodnie z obowiązującymi przepisami.

Technika: w celu udostępniania danych do ponownego wykorzystywania należy dokonać pełnej anonimizacji, dla zapewnienia odpowiedniego poziomu bezpieczeństwa wskazane jest jednoczesne zastosowanie technik randomizacji i uogólniania.

b) Szczególne kategorie danych osobowych

Dotyczy danych osobowych ujawniających pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub światopoglądowe, przynależność do związków zawodowych oraz dane genetyczne, dane biometryczne przetwarzane w celu jednoznacznego zidentyfikowania osoby fizycznej lub dane dotyczące zdrowia, seksualności lub orientacji seksualnej tej osoby.

Są to dane osobowe niedostępne powszechnie zgodnie z obowiązującymi przepisami. Ich udostępnianie w celu ponownego wykorzystywania jest zakazane.

Technika: w celu udostępniania danych do ponownego wykorzystywania należy dokonać pełnej anonimizacji, dla zapewnienia odpowiedniego poziomu bezpieczeństwa wskazane jest jednoczesne zastosowanie technik randomizacji i uogólniania.

c) Dane powstałe w wyniku agregacji lub anonimizacji danych osobowych

W przypadku udostępniania do ponownego wykorzystywania zbiorów danych powstałych w wyniku anonimizacji danych osobowych konieczne jest każdorazowo przeprowadzenia oceny ryzyka ich deanonimizacji.

Technika: w celu udostępniania danych do ponownego wykorzystywania oraz w razie ustalenia wystąpienia ryzyka identyfikacji danych osobowych należy powtórzyć proces anonimizacji.

2.Przykłady zastosowania technik dla typów danych

Każdy typ danych może podlegać obowiązkowi anonimizacji, jeśli może on zostać powiązany z konkretną osobą fizyczną.

Konieczność anonimizacji może wynikać z tego, że typ danych stanowi sam w sobie daną osobową (np. numer PESEL) lub z tego, że zbiór atrybutów jednoznacznie wskazuje na daną osobę lub istnieje szczątkowe ryzyko identyfikacji dla osoby, których dane dotyczą. Przykładem może być tu wyznanie, które nie powinno być ujawnianie w kontekście konkretnej osoby, ale jeśli zbiór danych jest odpowiednio uogólniony, możemy zostawić takie pole informacyjne bez obawy, że dla osoby fizycznej zostaną ujawnione konkretne dane.

a)Imię i nazwisko

Technika: pełna anonimizacja lub wymazanie (np. przez maskowanie, aby nie było możliwości odczytu). Pseudonimizacja w przypadku, jeśli może wystąpić konieczność identyfikacji na indywidualny, uzasadniony prawnie wniosek.

[15] Chodzi o szczególne kategorie danych osobowych w rozumieniu art. 9 i 10 RODO (czyli tzw. dane wrażliwe), tj. dane osobowe: ujawniające pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub światopoglądowe, przynależność do związków zawodowych oraz przetwarzania danych genetycznych, danych biometrycznych w celu jednoznacznego zidentyfikowania osoby fizycznej lub danych dotyczących zdrowia, seksualności lub orientacji seksualnej tej osoby; dotyczące wyroków skazujących oraz naruszeń prawa lub powiązanych środków bezpieczeństwa.

b) Adres (miasto, ulica, adres domu/mieszkania)

Technika: pełna anonimizacja lub wymazanie (np. przez maskowanie, aby nie było możliwości odczytu). W przypadku braku ryzyka deanonimizacji można stosować uogólnienie, poprzez ujawnienie przykładowo województwa.

c) Kod pocztowy

Technika: pełna anonimizacja lub wymazanie (np. przez maskowanie, aby nie było możliwości odczytu). W przypadku braku ryzyka deanonimizacji można stosować uogólnienie poprzez maskowanie kilku ostatnich cyfr kodu.

d) Dane liczbowe (zarobki, waga, wzrost lub inne dane opisujące osobę)

Technika: jednoczesne zastosowanie metod randomizacji (np. dodawanie zakłóceń +/- 20%) oraz uogólniania (np. agregacja przez wprowadzenie przedziałów wartości, przy założeniu, że poszczególne przedziały będą wystarczająco licznie reprezentowane).

e) Identyfikatory będące samodzielnie danymi osobowymi (np. PESEL)

Technika: pełna anonimizacja lub wymazanie (np. przez maskowanie, aby nie było możliwości odczytu). Pseudonimizacja w przypadku, jeśli może wystąpić konieczność identyfikacji na indywidualny, uzasadniony prawnie wniosek. Jeśli nie występuje ryzyko szczątkowe identyfikacji dla osoby, której dane dotyczą, możliwe jest udostępnienie części takiego pola, np. dwóch pierwszych cyfr numeru PESEL, które oznaczają rok urodzenia. Jeśli po takim ograniczeniu zawartości pola istnieje ryzyko identyfikacji, należy zastosować jednoczesne zastosowanie metod randomizacji i uogólniania, tak jak to opisano powyżej.

4.

Zmiana celu przetwarzania danych osobowych z zasobów publicznych

Ponowne wykorzystywanie co do zasady będzie związane ze zmianą celu przetwarzania danych osobowych w stosunku do celu, dla jakiego były one zebrane.

Z uwagi na charakter danych osobowych znajdujących się w zasobach publicznych należy uznać za mało prawdopodobne, że osoby, które przekazują do nich dane, nie wyrażały do tej pory zgody na ich ponowne wykorzystywanie.

Zgodnie z art. 6 ust. 4 RODO[16] dopuszcza się zmianę celu przetwarzania danych osobowych w stosunku do celu, w jakim zostały one pierwotnie zebrane przy spełnieniu jednej z poniższych przesłanek[17]:

1) **Kiedy osoba, której dane dotyczą, wyraziła na to zgodę;**
2) **Kiedy nowy cel przetwarzania danych jest zgodny z pierwotnym celem ich zebrania,** przy czym dalsze przetwarzanie dla celów archiwalnych w interesie publicznym, do celów badań naukowych lub historycznych lub do celów statystycznych nie jest uznawane za niezgodne z pierwotnymi celami.

W celu ustalenia, czy ponowne wykorzystywanie danych osobowych jest dopuszczalne, konieczne jest przeprowadzenie **testu zgodności ponownego wykorzystywania z pierwotnym celem zebrania danych.** Przykładowe okoliczności, które dysponent danych musi wziąć pod uwagę:

---

[16]Art. 6 ust. 4. Jeżeli przetwarzanie w celu innym niż cel, w którym dane osobowe zostały zebrane, nie odbywa się na podstawie zgody osoby, której dane dotyczą, ani prawa Unii lub prawa państwa członkowskiego stanowiących w demokratycznym społeczeństwie niezbędny i proporcjonalny środek służący zagwarantowaniu celów, o których mowa w art. 23 ust. 1, administrator - aby ustalić, czy przetwarzanie w innym celu jest zgodne z celem, w którym dane osobowe zostały pierwotnie zebrane - bierze pod uwagę między innymi:
a) wszelkie związki między celami, w których zebrano dane osobowe, a celami zamierzonego dalszego przetwarzania;
b) kontekst, w którym zebrano dane osobowe, w szczególności relację między osobami, których dane dotyczą, a administratorem;
c) 22) charakter danych osobowych, w szczególności czy przetwarzane są szczególne kategorie danych osobowych zgodnie z art. 9 lub dane osobowe dotyczące wyroków skazujących i czynów zabronionych zgodnie z art. 10;
d) ewentualne konsekwencje zamierzonego dalszego przetwarzania dla osób, których dane dotyczą;
e) istnienie odpowiednich zabezpieczeń, w tym ewentualnie szyfrowania lub pseudonimizacji.

[17]Art. 23 ust. 1 RODO dopuszcza ponadto zmianę w następujących celach: a) ochrona bezpieczeństwa narodowego; b) zapewnienie obrony; c) zapewnienie bezpieczeństwa publicznego; d) zapobieganie przestępczości, prowadzeniu postępowań przygotowawczych, wykrywaniu lub ściganiu czynów zabronionych lub wykonywaniu kar, w tym ochronie przed zagrożeniami dla bezpieczeństwo publicznego i zapobieganiu takim zagrożeniom; e) inne ważne cele leżące w ogólnym interesie publicznym Unii lub państwa członkowskiego, w szczególności ważnemu interesowi gospodarczemu lub finansowemu Unii lub państwa członkowskiego, w tym kwestiom pieniężnym, budżetowym i podatkowym, zdrowiu publicznemu i zabezpieczeniu społecznemu; f) ochronie niezależności sądów i postępowania sądowego; g) zapobieganiu naruszeniom zasad etyki w zawodach regulowanych, prowadzeniu postępowań takich sprawach, ich wykrywaniu oraz ściganiu; h) cel kontrolny, inspekcyjny lub regulacyjny związany, nawet sporadycznie ze sprawowaniem władzy publicznej w przypadkach o których mowa w lit a-e) powyżej oraz g); i) ochronny wobec osoby, której dane dotyczą, lub praw i wolności tych osób oraz j) dla celów egzekucji roszczeń cywilnoprawnych. Analiza tych celów jasno wskazuje również, że przepis ten nie będzie mieć zastosowania do ponownego wykorzystywania danych publicznych.

1) Wszelkie **związki między celami**, w których zebrano dane osobowe, a celami zamierzonego dalszego przetwarzania;
2) **Kontekst**, w którym zebrano dane osobowe, w szczególności relację między osobami, których dane dotyczą, a dysponentem danych;
3) Ewentualne **konsekwencje zamierzonego dalszego przetwarzania** dla osób, których dane dotyczą;
4) **Istnienie odpowiednich zabezpieczeń**, w tym ewentualnie szyfrowanie lub pseudonimizacja;
5) **Konsekwencje i zagrożenia wynikające** z utraty kontroli nad udostępnionymi danymi po ich udostępnieniu;
6) **Szczególną kategorię danych.**

Jeśli w wyniku przeprowadzonego testu zgodności ponownego wykorzystywania z pierwotnym celem, dla którego dane zebrano, uzyskano odpowiedź negatywną dla udostępnienia danych osobowych zawartych w danych publicznych, konieczna będzie ich depersonalizacja z wykorzystaniem technik opisanych w **Załączniku nr 8.**

Test zgodności celów przeprowadza się wyłącznie, gdy zmiana celu nie może być uzasadniona w oparciu o zgodę lub przepis prawa regulujący kwestie tej zmiany.

W przypadku przetwarzania danych osobowych osób pełniących funkcje publiczne w związku z ich pełnieniem przepis art. 6 ust. 2 upw[18] stanowi podstawę (zawartą w prawie krajowym) zmiany celu przetwarzania danych.

Przeprowadzenie testu zgodności celów nie będzie wymagane także w przypadku pozyskania zgody, o której mowa w treści tego przepisu („rezygnacji z przysługujących praw" zgodnie z art. 6 ust. 2 upw).

Należy zwrócić uwagę, że na pozytywny wynik testu zgodności celów wpływać będzie wprowadzenie mechanizmów ograniczających identyfikowalność (np. pseudonimizacja). Świadczy to o ścisłym związku testu zgodności celów (z art. 6 ust. 4 RODO[19]) i analizy ryzyka naruszenia praw lub wolności.

Chociaż nie wynika to wprost z RODO, to należy przyjąć, że w ramach testu zgodności celów należy przeprowadzić analizę ryzyka, która może wpłynąć na odpowiedź na pytanie, czy i w jakim zakresie

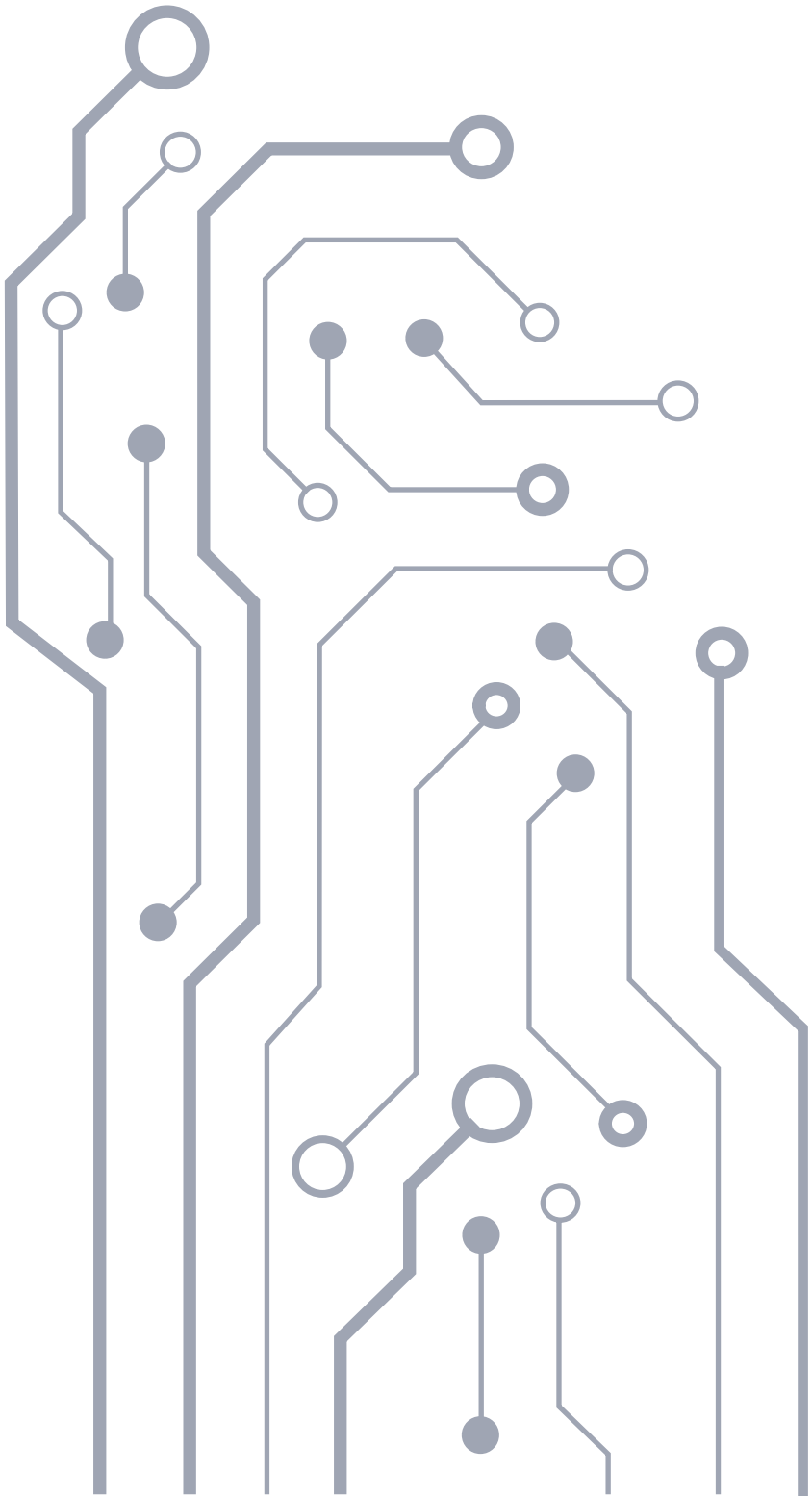
---

[18] Art. 6 ust. 2. Prawo do ponownego wykorzystywania podlega ograniczeniu ze względu na prywatność osoby fizycznej lub tajemnicę przedsiębiorcy. Ograniczenie to nie dotyczy informacji o osobach pełniących funkcje publiczne, mających związek z pełnieniem tych funkcji, w tym o warunkach powierzenia i wykonywania funkcji, oraz przypadku gdy osoba fizyczna lub przedsiębiorca rezygnują z przysługującego im prawa.

[19] Art. 6 ust. 4. Jeżeli przetwarzanie w celu innym niż cel, w którym dane osobowe zostały zebrane, nie odbywa się na podstawie zgody osoby, której dane dotyczą, ani prawa Unii lub prawa państwa członkowskiego stanowiących w demokratycznym społeczeństwie niezbędny i proporcjonalny środek służący zagwarantowaniu celów, o których mowa w art. 23 ust. 1, administrator - aby ustalić, czy przetwarzanie w innym celu jest zgodne z celem, w którym dane osobowe zostały pierwotnie zebrane - bierze pod uwagę między innymi:
a) wszelkie związki między celami, w których zebrano dane osobowe, a celami zamierzonego dalszego przetwarzania;
b) kontekst, w którym zebrano dane osobowe, w szczególności relację między osobami, których dane dotyczą, a administratorem;
c) 22) charakter danych osobowych, w szczególności czy przetwarzane są szczególne kategorie danych osobowych zgodnie z art. 9 lub dane osobowe dotyczące wyroków skazujących i czynów zabronionych zgodnie z art. 10;
d) ewentualne konsekwencje zamierzonego dalszego przetwarzania dla osób, których dane dotyczą;
e) istnienie odpowiednich zabezpieczeń, w tym ewentualnie szyfrowania lub pseudonimizacji.

może dojść do ujawnienia danych osobowych (zobacz część **VI. Standardu**).Test zgodności przeprowadza się w oparciu o **Załącznik nr 4.**

Dysponent danych musi w ramach testu zgodności celów zidentyfikować i ocenić pierwotny cel zebrania danych. Dlatego też należy rekomendować, aby użytkownik – weryfikując swoją podstawę przetwarzania danych - znał pierwotny cel zebrania danych osobowych. Dysponent danych powinien zatem komunikować użytkownikowi jaki był pierwotny cel zebrania danych osobowych zawartych w danych publicznych, np. poprzez wskazanie go w warunkach ponownego wykorzystywania.



5.

Środki techniczno-organizacyjne służące zapewnieniu bezpieczeństwa przetwarzanych danych osobowych

Dysponent danych i podmiot przetwarzający dane osobowe wdrażają odpowiednie środki techniczne i organizacyjne, aby zapewnić stopień bezpieczeństwa odpowiadający danemu ryzyku .

Przykładowe środki techniczne lub organizacyjne, które mogą być odpowiednie do redukcji określonych kategorii ryzyk w związku z ponownym wykorzystywaniem to:

- Anonimizacja;
- Pseudonimizacja danych;
- Okresowa weryfikacja poprawności danych;
- Wskazanie na jaki dzień określony rejestr jest aktualny;
- Okresowa aktualizacja ujawnianych danych;
- Wprowadzenie warunków ponownego wykorzystania ISP (zob. poniżej);
- Określenie w warunkach ponownego wykorzystania ISP pierwotnego celu zebrania danych osobowych;
- Weryfikacja celów dla jakich dane osobowe mają być ponownie wykorzystane (w trybie wnioskowym, tj. gdy jest to możliwe);
- Wdrożenie mechanizmów zapewniających korektę czynników powodujących nieprawidłowości w danych osobowych i maksymalne zmniejszenie ryzyka błędów;
- Ograniczenie możliwości pobierania całej zawartości rejestru;
- Ograniczenie możliwości automatycznego zaciągania całego rejestru;
- Ograniczenie zapytań wyszukiwania, na przykład na podstawie imienia i nazwiska osoby.

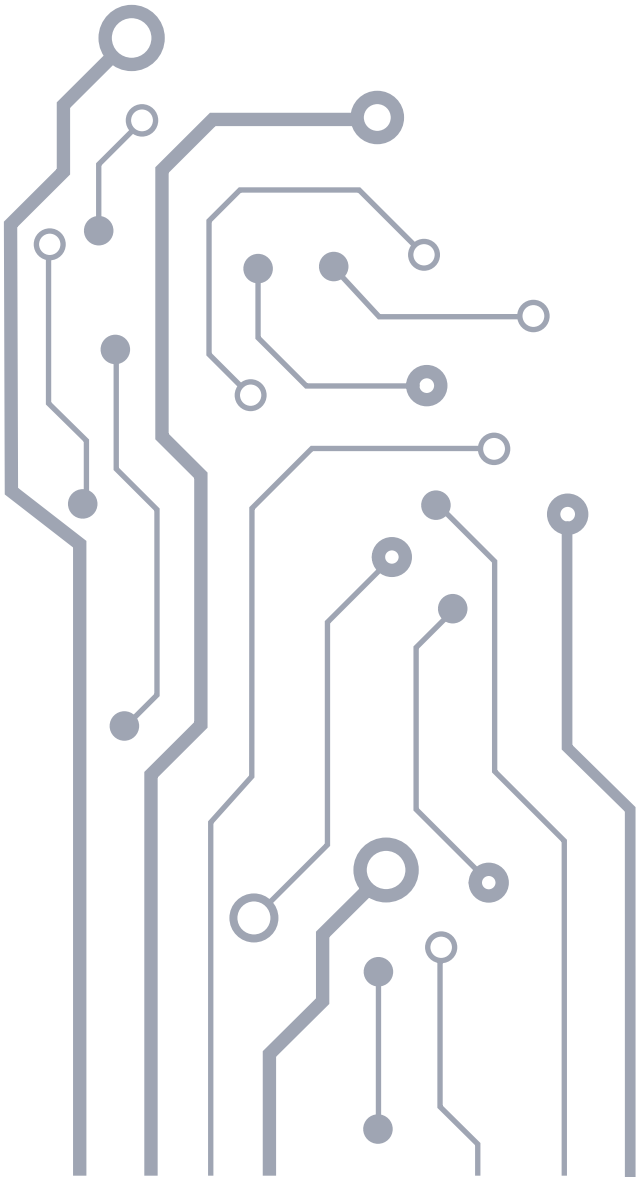
[20]Zgodnie z artykułem 32 RODO
Art. 32 ust. 1.Uwzględniając stan wiedzy technicznej, koszt wdrażania oraz charakter, zakres, kontekst i cele przetwarzania oraz ryzyko naruszenia praw lub wolności osób fizycznych o różnym prawdopodobieństwie wystąpienia i wadze, administrator i podmiot przetwarzający wdrażają odpowiednie środki techniczne i organizacyjne, aby zapewnić stopień bezpieczeństwa odpowiadający temu ryzyku, w tym między innymi w stosownym przypadku:
a) pseudonimizację i szyfrowanie danych osobowych;
b) zdolność do ciągłego zapewnienia poufności, integralności, dostępności i odporności systemów i usług przetwarzania;
c) zdolność do szybkiego przywrócenia dostępności danych osobowych i dostępu do nich w razie incydentu fizycznego lub technicznego;
d) regularne testowanie, mierzenie i ocenianie skuteczności środków technicznych i organizacyjnych mających zapewnić bezpieczeństwo przetwarzania.
2. Oceniając, czy stopień bezpieczeństwa jest odpowiedni, uwzględnia się w szczególności ryzyko wiążące się z przetwarzaniem, w szczególności wynikające z przypadkowego lub niezgodnego z prawem zniszczenia, utraty, modyfikacji, nieuprawnionego ujawnienia lub nieuprawnionego dostępu do danych osobowych przesyłanych, przechowywanych lub w inny sposób przetwarzanych.
3. Wywiązywanie się z obowiązków, o których mowa w ust. 1 niniejszego artykułu, można wykazać między innymi poprzez stosowanie zatwierdzonego kodeksu postępowania, o którym mowa w art. 40 lub zatwierdzonego mechanizmu certyfikacji, o którym mowa w art. 42.
4. Administrator oraz podmiot przetwarzający podejmują działania w celu zapewnienia, by każda osoba fizyczna działająca z upoważnienia administratora lub podmiotu przetwarzającego, która ma dostęp do danych osobowych, przetwarzała je wyłącznie na polecenie administratora, chyba że wymaga tego od niej prawo Unii lub prawo państwa członkowskiego.

Zgodnie z zasadami ochrony danych (odnoszących się do fazy projektowania oraz domyślnej ochrony danych), przy rozważaniu ich udostępnienia należy jak najwcześniej uwzględnić fakt, że niektóre z nich mogą zawierać dane osobowe.

Przy udostępnieniu danych dobór odpowiednich technik, pozwalających na zapewnienie prywatności, ale umożliwiających zachowanie wartości informacyjnej zbioru danych, jest kwestią kluczową. Technikami, które pozwalają na utrzymywanie korzyści z danych oraz minimalizują ryzyko w związku z utratą prywatności, jest anonimizacja lub pseudonimizacja.

Techniki anonimizacji i pseudonimizacji zostały opisane w **Załączniku nr 8.**

Proces anonimizacji informacji jest sposobem pogodzenia względów przemawiających za umożliwieniem ponownego wykorzystywania informacji sektora publicznego w jak najszerszym zakresie z obowiązkami wynikającymi z przepisów o ochronie danych. Nie należy jednak taktować tego procesu np. jako informacji przetworzonej nawet jeśli jest czasochłonny i wymaga zwiększonego nakładu środków osobowych. Czynności te stanowią proste czynności kancelaryjno-biurowe o charakterze technicznym. Tymczasem informacja przetworzona to informacja wytworzona specjalnie na potrzeby wnioskodawcy z dokumentów pozostających w dyspozycji podmiotów, wymagająca podjęcia działań przekraczających proste czynności (Wyrok Wojewódzkiego Sądu Administracyjnego w Łodzi z dnia 4 kwietnia 2019 r., II SA/Łd 134/19).



6.

Ryzyka dla ochrony danych osobowych zawartych w danych publicznych

Jeżeli dany rodzaj przetwarzania – w szczególności z użyciem nowych technologii – ze względu na swój charakter, zakres, kontekst i cele z dużym prawdopodobieństwem może powodować wysokie ryzyko naruszenia praw lub wolności osób fizycznych, dysponent danych przed rozpoczęciem przetwarzania powinien dokonać oceny skutków planowanych operacji przetwarzania dla ochrony danych osobowych . Rekomendowanym rozwiązaniem w przypadku udostępniania danych publicznych zawierających dane osobowe jest przeprowadzenie oceny skutków dla ochrony danych osobowych.

Ocena skutków dla ochrony danych zawiera co najmniej:

a) Systematyczny opis planowanych operacji przetwarzania i celów przetwarzania, w tym, gdy ma to zastosowanie – prawnie uzasadnionych interesów realizowanych przez dysponenta danych;

b) Ocenę, czy operacje przetwarzania są niezbędne oraz proporcjonalne w stosunku do celów;

c) Ocenę ryzyka naruszenia praw lub wolności osób, których dane dotyczą;

d) Środki planowane w celu zaradzenia ryzyku, w tym zabezpieczenia oraz środki i mechanizmy bezpieczeństwa mające zapewnić ochronę danych osobowych i wykazać przestrzeganie RODO, z uwzględnieniem praw i prawnie uzasadnionych interesów osób, których dane dotyczą i innych osób, których sprawa dotyczy.

Kluczowym elementem oceny skutków jest ocena ryzyka naruszenia praw lub wolności, na którą składa się identyfikacja i szacowanie ryzyka naruszenia praw lub wolności. **Ocena skutków jest pogłębionym mechanizmem oceny ryzyka.**

Analizę ryzyka należy przeprowadzić nawet, jeżeli wstępna ocena operacji realizowanych w ramach ponownego wykorzystania danych publicznych nie wskazuje, że dany rodzaj przetwarzania - w szczególności z użyciem nowych technologii - ze względu na swój charakter, zakres, kontekst i cele z dużym prawdopodobieństwem może powodować wysokie ryzyko naruszenia praw lub wolności osób fizycznych.

---

[21] Art. 35 RODO, ang. Data Protection Impact Assessment (tzw. DPIA).
Art. 35. 1. Jeżeli dany rodzaj przetwarzania – w szczególności z użyciem nowych technologii – ze względu na swój charakter, zakres, kontekst i cele z dużym prawdopodobieństwem może powodować wysokie ryzyko naruszenia praw lub wolności osób fizycznych, administrator przed rozpoczęciem przetwarzania dokonuje oceny skutków planowanych operacji przetwarzania dla ochrony danych osobowych. Dla podobnych operacji przetwarzania danych wiążących się z podobnym wysokim ryzykiem można przeprowadzić pojedynczą ocenę.
2. Dokonując oceny skutków dla ochrony danych, administrator konsultuje się z inspektorem ochrony danych, jeżeli został on wyznaczony.
3. Ocena skutków dla ochrony danych, o której mowa w ust. 1, jest wymagana w szczególności w przypadku:
a) systematycznej, kompleksowej oceny czynników osobowych odnoszących się do osób fizycznych, która opiera się na zautomatyzowanym przetwarzaniu, w tym profilowaniu, i jest podstawą decyzji wywołujących skutki prawne wobec osoby fizycznej lub w podobny sposób znacząco wpływających na osobę fizyczną;
b) 35) przetwarzania na dużą skalę szczególnych kategorii danych osobowych, o których mowa w art. 9 ust. 1, lub danych osobowych dotyczących wyroków skazujących i czynów zabronionych, o czym mowa w art. 10; lub
c) systematycznego monitorowania na dużą skalę miejsc dostępnych publicznie.
4. Organ nadzorczy ustanawia i podaje do publicznej wiadomości wykaz rodzajów operacji przetwarzania podlegających wymogowi dokonania oceny skutków dla ochrony danych na mocy ust. 1. Organ nadzorczy przekazuje te wykazy Europejskiej Radzie Ochrony Danych, o której mowa w art. 68.

Ocenę ryzyka należy przeprowadzać w przypadku ujawniania danych osobowych w ramach ponownego wykorzystywania. Należy ją przeprowadzić także w przypadku ujawniania danych anonimowych – w tym przypadku chodzić będzie o ocenę wagi i prawdopodobieństwa możliwej re-identyfikacji.

Ogólna ocena ryzyka i jej szczególny przypadek tj. ocena skutków musi:

- uwzględniać wszystkie etapy przetwarzania danych osobowych (nie tylko ujawnienia ich innemu podmiotowi), ze szczególnym naciskiem położonym na fazę projektowania;
- uwzględniać wszystkie obszary wymogów (nie tylko bezpieczeństwo);
- uwzględnić identyfikację przypadków, gdy potrzebna będzie ocena skutków;
- nie ograniczać oceny ryzyka do przypadku oceny skutków;
- uwzględniać wykaz PUODO operacji wymagających oceny skutków[22];

Oczywiście nie oznacza to, że każdorazowe (np. dla wszystkich operacji na danych) prowadzenie oceny skutków jest zakazane. Zwykle jednak byłoby jednak nadmierne i zbyt czasochłonne. Zwłaszcza, że w ramach udostępnienia danych publicznych realizowane mogą być różne operacje na danych osobowych, z którymi może łączyć się różne ryzyko naruszenia praw lub wolności.

Co istotne, analiza ryzyka nie może być też argumentem za nieudostępnieniem danych publicznych w przypadku, gdy istnieje w tym zakresie wyraźny wymóg prawny. Dotyczyć to będzie np. informacji o osobach pełniących funkcje publiczne w związku z pełnieniem tych funkcji. Natomiast także w tym przypadku analiza ryzyka może mieć znaczenie np. w kontekście wdrożenia środków zapewniających poprawności danych.

Ocenę skutków należy przeprowadzić wtedy, gdy:

- wstępna ocena realizowanych operacji (**załącznik nr 1 i załącznik nr 2**) wykaże, że jedna z operacji na danych mieści się w kryteriach z wykazu PUODO;

w tym przypadku z góry wiadomo, że przeprowadzona będzie ocena skutków;

- dopiero ocena ryzyka naruszenia praw lub wolności (**załącznik nr 5**) wykaże konieczność przeprowadzenia oceny skutków;

w tym przypadku z góry nie wiadomo, że będzie prowadzona ocena skutków, dlatego po zidentyfikowaniu dużego prawdopodobieństwa wysokiego ryzyka (w oparciu o załącznik nr 5) należy uzupełnić ocenę wypełniając treść załącznika nr 6.

Biorąc pod uwagę charakter realizowanych operacji (**załącznik nr 1-2**) należy przyjąć, że konieczność przeprowadzenia oceny skutków (pogłębionej oceny ryzyka) w praktyce dotyczyć może przypadków, gdy przetwarzane będą dane innych osób niż pełniących funkcje publiczne.

[22] Wykaz stanowi załącznik do Komunikatu Prezesa Urzędu Ochrony Danych z dnia 17 czerwca 2019 r. w sprawie wykazu rodzajów operacji przetwarzania danych osobowych wymagających oceny skutków przetwarzania dla ich ochrony. Osobowych http://monitorpolski.gov.pl/MP/2018/827/

W przypadku oceny skutków należy wypełnić obligatoryjnie **załączniki 1-7**, z tym, że **załącznik nr 4** wyłącznie, gdy podstawa przetwarzania danych wskazywać będzie na konieczność przeprowadzenia testu zgodności (patrz część **IV. Standardu**).

Dysponenci danych i użytkownicy dla przeprowadzenia ceny skutków mogą ponadto posiłkować się dostępnymi wskazówkami (Grupa Robocza Art. 29, Wytyczne dotyczące oceny skutków dla ochrony danych oraz ustalenia czy przetwarzanie „z dużym prawdopodobieństwem może powodować wysokie ryzyko"), dobrymi praktykami i normami, jak np. norma ISO/IEC 29134:2017 (Information technology -- Security techniques -- Guidelines for privacy impact assessment).

Przy analizie ryzyka na potrzeby oceny ryzyka można posiłkować się normą PN-ISO/IEC 27005 (Technika informatyczna -- Techniki bezpieczeństwa -- Zarządzanie ryzykiem w bezpieczeństwie informacji) oraz normą PN-ISO 31000 (Zarządzanie ryzykiem -- Zasady i wytyczne). Analiza ryzyka powinna zostać udokumentowana oraz regularnie przeglądana.

7.

Ocena ryzyka i ocena skutków

Z przetwarzaniem danych osobowych zawartych w zasobach publicznych, udostępnianych przez systemy teleinformatyczne do ponownego wykorzystywania lub przekazywanych na wniosek, może wiązać się ryzyko naruszenia praw lub wolności osób fizycznych o różnym prawdopodobieństwie i wadze zagrożenia.

Prawdopodobieństwo ryzyka związanego z naruszeniem praw osób, których dane dotyczą, różnić się będzie w zależności od kategorii przypadku opisanego w **załączniku nr 1**, i operacji realizowanych w ramach każdego z tych przypadków (zob. **załącznik nr 2**).

To jakie operacje będą mieć miejsce zależy od przesądzenia, którego przypadku z **załącznika nr 1** dotyczy ujawnienie danych publicznych.

Na tej podstawie w ramach **załącznika nr 2** stworzono wykaz typowych zagrożeń (źródeł ryzyka naruszenia praw lub wolności), które dotyczą poszczególnych operacji na danych typowych dla ponownego wykorzystania.

Można wymienić przykładowe ryzyka związane z udostępnianiem danych:

1. Możliwość szczątkowej identyfikacji z uwagi na zbyt powierzchowną anonimizację

Użytkownicy postronni mogą dokonać identyfikacji danych konkretnej osoby, które w ocenie dysponenta danych zostały w pełni zanonimizowane bądź zagregowane. Problem ten dotyczy w szczególności danych statystycznych, które w wyniku zbyt dużej szczegółowości oraz w połączeniu ze zbyt małą próbą powodują, że osoby wchodzące w skład danej wspólnoty mogą przy pomocy powszechnie znanych im informacji o jej członkach dokonać odkodowania anonimowych informacji. Ryzyko identyfikacji danych może nastąpić także poprzez połączenie zanonimizowanego zbioru danych z innymi zbiorami, co w konsekwencji także pozwala na odkodowanie anonimowych informacji.

2. Łączenie danych pochodzących z różnych źródeł

W powiązaniu z danymi publicznie dostępnymi w innych zbiorach, dane mogą posłużyć do stworzenia kompleksowych zbiorów danych osobowych osób, których dane dotyczą, co może prowadzić do naruszenia ich prywatności.

3. Przetwarzanie danych osobowych po ich usunięciu z zasobu publicznego (wykreśleniu z zasobu)

Dane przetwarzane w zasobach publicznych często są w nich ujawniane na zadany okres, po czym podlegają wykreśleniu. W efekcie, po wykreśleniu danych osoba, której dane dotyczą, ma zagwarantowane zaprzestanie ich udostępniania po upływie okresu, przez jaki miały być one zgodnie z prawem przetwarzane. Po udostępnieniu danych do ponownego wykorzystywania dysponent danych traci wpływ na sposoby jego wykorzystywania. Ponowne udostępnienie danych osobowych rodzi ryzyko ich przetwarzania w okresie, w jakim zostały one już wykreślone z zasobów publicznych.

4. Zagrożenie dla ochrony szczególnych kategorii danych

W przypadku udostępniania szczególnych kategorii danych zawartych w zbiorach publicznych istnieje ryzyko upublicznienia tych danych do nieograniczonego kręgu adresatów. Zasadniczo ich ponowne wykorzystywanie w celach innych, niż cele dla których zostały zebrane jest niedozwolone.

5. Możliwość szerszego wykorzystywania danych

Przetworzenie danych do formatu odczytu maszynowego powoduje możliwość szerszego wykorzystywania danych osobowych np. stosowanie profilowania, marketing bezpośredni, tworzenie baz adresowych oraz danych komunikacyjnych.

8.

Postępowanie na wypadek ryzyka identyfikacji danych osobowych

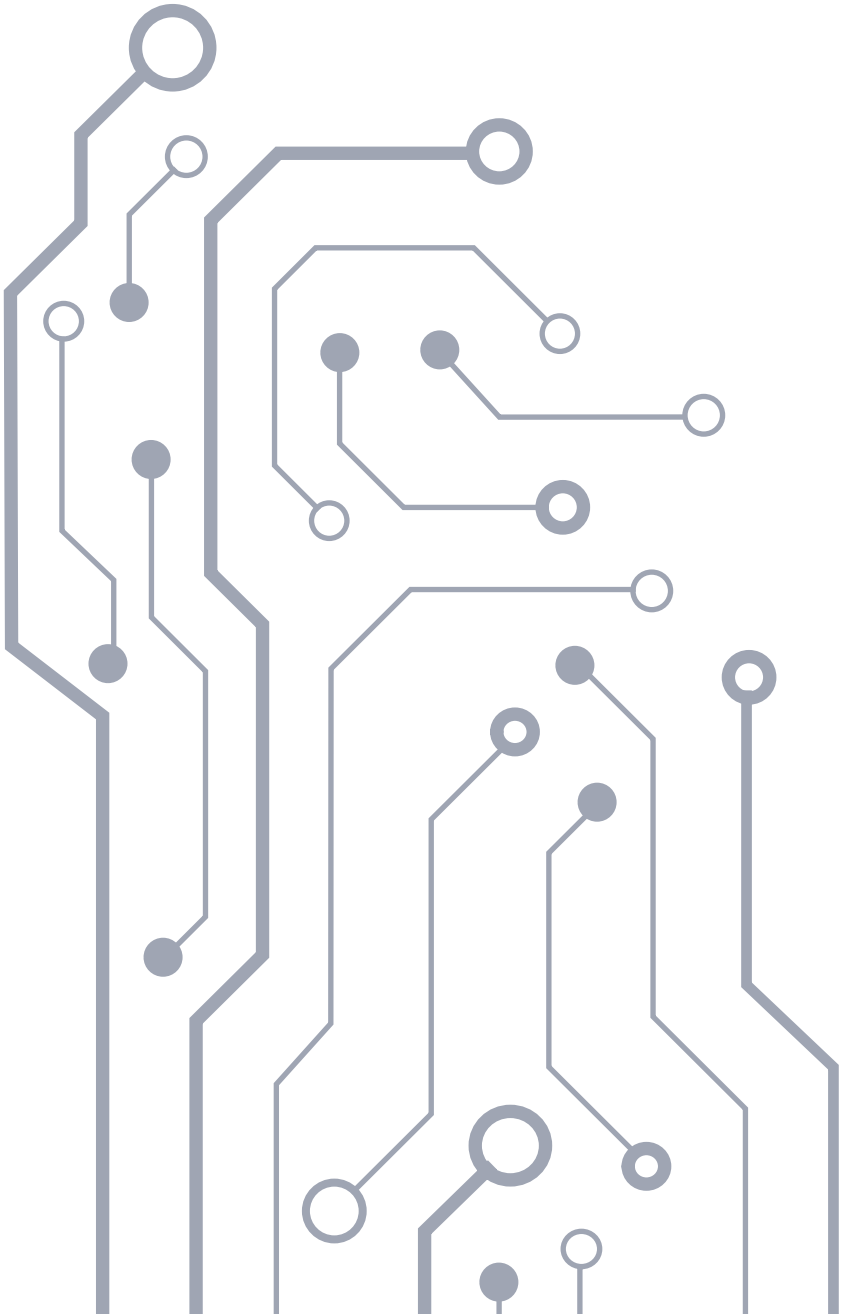
Techniki anonimizacji wiążą się z konkretnymi ograniczeniami. Dysponenci danych muszą rozważyć te ograniczenia, zanim zastosują daną technikę w celu przeprowadzenia procesu anonimizacji. Muszą zwrócić uwagę na cele, jakie należy osiągnąć przez anonimizację – takie jak ochrona prywatności osób fizycznych przy publikowaniu zbioru danych lub dopuszczaniu wyszukania części informacji ze zbioru danych.

Nawet po zastosowaniu anonimizacji nie można wykluczyć innych zagrożeń, takich jak możliwość wyodrębnienia określonej osoby fizycznej, możliwość tworzenia powiązań między zapisami dotyczącymi określonej osoby, czy możliwość wnioskowania w odniesieniu do określonej osoby. Niektóre z tych zagrożeń można jednak wyeliminować lub ograniczyć za pomocą konkretnej techniki, przy czym zalecane jest łączenie różnych technik, co zwiększa szanse na osiągnięcie skutecznej ochrony danych osobowych.

W celu ograniczenia ryzyka identyfikacji danych osobowych należy dodatkowo uwzględnić następujące kwestie:

1. Należy jasno określić cele, jakie planuje się osiągnąć poprzez zanonimizowanie zbioru danych, ponieważ odgrywają one ważną rolę w określaniu ryzyka identyfikacji;
2. Ze względu na ryzyko szczątkowej identyfikacji dysponenci danych powinni:
   - **a)** identyfikować nowe rodzaje ryzyka i regularnie przeprowadzać ponowne oceny ryzyka szczątkowego,
   - **b)** ocenić, czy kontrole w odniesieniu do zidentyfikowanego ryzyka są wystarczające i odpowiednio dostosowywane
   - **c)** monitorować i kontrolować ryzyko.
3. W ramach ryzyka szczątkowego należy ocenić, czy istnieje możliwość identyfikacji w niezanonimizowanej części zbioru (jeżeli taka istnieje), w szczególności jeżeli została ona połączona z częścią zanonimizowaną, oraz możliwe korelacje między atrybutami, np. między danymi dotyczącymi lokalizacji geograficznej a danymi dotyczącymi poziomu zamożności;
4. Jednocześnie należy uwzględnić wszystkie odpowiednie elementy kontekstowe, np. charakter danych pierwotnych, istniejące mechanizmy kontroli (w tym środki bezpieczeństwa służące ograniczeniu dostępu do zbiorów danych), liczebność próby (cechy ilościowe), dostępność zasobów informacji publicznych (na których mogą opierać się odbiorcy), przewidziane udostępnienie danych osobom trzecim (ograniczone, nieograniczone np. w Internecie itp.);
5. Należy zwrócić uwagę na potencjalnych atakujących, uwzględniając atrakcyjność danych z perspektywy ukierunkowanych ataków (w tym kontekście ponownie najważniejszymi czynnikami będą szczególna ochrona informacji i charakter danych);
6. Ze zbioru danych należy usunąć oczywiste (np. rzadkie) atrybuty / *quasi-identyfikatory*;
7. Jeżeli stosuje się technikę dodawania zakłóceń (w randomizacji), poziom zakłóceń dodanych do zapisów należy określić jako funkcję wartości atrybutu (tj. nie należy dodawać żadnych zakłóceń wykraczających poza skalę), wpływu atrybutów, które mają podlegać ochronie, na osoby, których dane dotyczą, lub rozproszenie zbioru danych;

8. W przypadku opierania się na prywatności różnicowej (w randomizacji) należy uwzględnić konieczność monitorowania zapytań, aby wykrywać zapytania naruszające prywatność, ponieważ naruszenia w ramach zapytań mają charakter kumulacyjny;
9. Jeżeli wdrożono techniki uogólniania, bardzo istotne jest, aby dysponent danych nie ograniczał się do jednego kryterium uogólniania nawet w odniesieniu do tego samego atrybutu. Oznacza to, że należy wybierać różne poziomy szczegółowości lub różne przedziały czasowe. Wybór kryteriów, które należy stosować, musi zależeć od dystrybucji wartości atrybutu w danej populacji. Nie wszystkie dystrybucje nadają się do uogólniania, tj. w przypadku uogólniania nie można zastosować podejścia uniwersalnego. Należy zapewnić zmienność w ramach klas równoważności: na przykład wybrać określony próg na podstawie „elementów kontekstowych", o których mowa powyżej (liczebność próby itp.), i jeżeli próg ten nie zostanie osiągnięty, wówczas należy odrzucić określoną próbę (lub należy określić inne kryterium uogólniania);



9.

Współpraca

Ze względu na to, że otwieranie danych może wiązać się z przetwarzaniem danych osobowych konieczna jest stała współpraca i wymiana dobrych praktyk pomiędzy pracownikami odpowiedzialnymi za (np. tzw. pełnomocników ds. otwartości danych powołanych w administracji centralnej) ze służbami odpowiedzialnymi za ochronę danych osobowych (w szczególności inspektorem ochrony danych i jego zespołem)

Obszary współpracy:

- **Ocena ryzyka i ocena skutków**

Współpraca pomiędzy zespołami zajmującymi się otwartością danych i ochroną danych osobowych konieczna będzie w szczególności na etapie poprzedzającym udostępnienie danych, a więc przeprowadzania analizy ryzyka i oceny skutków.

Za przeprowadzenie oceny skutków odpowiedzialny jest formalnie administrator danych. Odpowiedzialny za ocenę ryzyka jest zarówno dysponent (ujawniający dane publiczne), jak i odbiorca (użytkownik), który będzie wykorzystywał dane.

RODO nie określa ram wewnątrzorganizacyjnych reguł prowadzenia oceny skutków, dlatego administrator danych (np. w wewnętrznej polityce ochrony danych osobowych) musi w praktyce przesądzić:

a) kto w organizacji będzie odpowiedzialny za wstępną ewaluacje co do potrzeby oceny skutków;

b) kto w organizacji będzie odpowiedzialny za prowadzenie oceny skutków;

c) kto w organizacji (np. bezpośrednio administrator lub osoba przez niego umocowana) będzie podejmował decyzje co do akceptacji wyników oceny skutków (np. akceptacji zaplanowanych środków służących zaradzeniu ryzyku);

d) metodologię oceny skutków;

e) wykorzystywane formularze, etc.

Dysponent danych musi podjąć decyzję co do sposobu prowadzenia analizy ryzyka naruszenia praw lub wolności. W szczególności czy analiza ryzyka będzie prowadzona w ramach każdego przypadku ujawniania danych publicznych czy np. w ramach określonych (powtarzających się) przypadków, które składają się z takich samych lub podobnych operacji. Pomocny w tym zakresie jest art. 35 ust. 1 RODO, który wskazuje, że dla podobnych operacji przetwarzania danych wiążących się z podobnym wysokim ryzykiem można przeprowadzić pojedynczą ocenę. W tym przypadku należy uwzględnić także elementy „tła" analizy ryzyka jak kontekst, zakres, charakter, skalę przetwarzania.

- **Depersonalizacja danych osobowych**

Konieczność przeprowadzenia anonimizacji i pseudnonimizacji (zob. **załącznik nr 8.**) danych osobowych udostępnianych w ramach ponownego wykorzystywania wymagać może również konsultacji z działem IT w instytucji.

- **Projektowanie przepisów**

Innym obszarem współpracy jest prowadzenie prac legislacyjnych. Tutaj konieczne jest włącznie również służb prawno-legislacyjnych urzędu obsługującego dysponenta danych.

Dokonanie oceny skutków w toku procesu legislacyjnego (oceny skutków regulacji) może zwolnić z tego obowiązku administratora danych. Co istotne, nie każda ocena skutków (w rozumieniu art. 35 rodo) przeprowadzona w toku postępowania legislacyjnego (w odniesieniu do projektowanych przepisów), może zwolnić administratora z tego obowiązku.

Ocena skutków na etapie projektowania przepisów powinna:

a) dotyczyć danych publicznych;

b) zawierać uzasadnienie, że zasób (wykazy, rejestry, etc.), którego dotyczą projektowane przepisy obejmuje dane publiczne;

c) zawierać uzasadnienie, że weryfikowany zasób zawierać będzie lub może zawierać dane osobowe;

d) określać - w ramach tła oceny ryzyka - jakich operacji dotyczy, ze szczególnym uwzględnieniem w jaki sposób dane osobowe (w ramach upw) będą ujawniane;

e) przesądzać, czy dotyczy każdego sposobu ujawniania danych publicznych (np. API);

f) przesądzać, czy dotyczy każdego trybu ujawniania danych publicznych („udostępnianie" / „przekazanie" w ramach art. 5 upw);

g) przesądzać, czy dotyczy skonkretyzowanych celów ponownego wykorzystania;

h) precyzować czy oceniane operacje wypełniać będą kryteria wykazu PUODO.

i) precyzować jakich kategorii osób i kategorii danych dotyczy;

j) zawierać ocenę zgodności celów przetwarzania: pierwotnego celu zebrania danych i celu w postaci udostępniania danych publicznych do ponownego wykorzystywania;

k) zawierać ocenę proporcjonalności i niezbędności ocenianych operacji, w szczególności ujawnienia danych w ramach upw;

l) zawierać identyfikację ryzyka naruszenia praw lub wolności (źródeł i potencjalnych negatywnych konsekwencji) a także szacowanie poziomu ryzyka;

m) wyjaśniać przyjętą metodologię oceny ryzyka naruszenia praw lub wolności;

n) określać środki planowane w celu zaradzenia ryzyku, w tym zabezpieczenia oraz środki i mechanizmy bezpieczeństwa mające zapewnić ochronę danych osobowych i wykazać przestrzeganie niniejszego rozporządzenia, z uwzględnieniem praw i prawnie uzasadnionych interesów osób, których dane dotyczą, i innych osób, których sprawa dotyczy.

W ramach powyższej oceny wykorzystane mogą być **załączniki nr 3-7**.

- **Konsultacje z PUODO**

Wynik przeprowadzonej oceny skutków może prowadzić do konieczności uprzednich konsultacji z Prezesem Urzędu Ochrony Osobowych. Dotyczy to sytuacji, gdy pomimo przeprowadzonej oceny skutków i zaplanowanych mechanizmów ograniczających ryzyko naruszenia praw lub wolności, w dalszym ciągu występuje wysokie ryzyko naruszenia praw lub wolności.

Niezależnie od uprzednich konsultacji z PUODO należy konsultować projekty aktów normatywnych wkraczających w sferę ochrony danych osobowych w ramach procedury opiniowania projektu zgodnie z Regulaminem pracy Rady Ministrów.

[23] Art. 5. Każdemu przysługuje prawo do ponownego wykorzystywania informacji sektora publicznego:
1) udostępnionych w systemie teleinformatycznym, a w szczególności na stronie podmiotowej Biuletynu Informacji Publicznej podmiotu zobowiązanego lub w centralnym repozytorium informacji publicznej, o którym mowa w art. 9a ustawy z dnia 6 września 2001 r. o dostępie do informacji publicznej (Dz. U. z 2018 r. poz. 1330 i 1669), zwanym dalej „centralnym repozytorium", lub w inny sposób;
2) przekazanych na wniosek o ponowne wykorzystywanie.

10.

Zakończenie

Z przetwarzania danych osobowych zawartych w zasobach publicznych, udostępnianych poprzez systemy teleinformatyczne do ponownego wykorzystywania, może wynikać ryzyko naruszenia praw lub wolności osób fizycznych. Ryzyko to może mieć różne prawdopodobieństwo oraz wagę. Dysponent danych i podmiot przetwarzający dane osobowe powinien wdrożyć odpowiednie środki techniczne i organizacyjne, aby zapewnić stopień bezpieczeństwa odpowiadający temu ryzyku. Kluczowym jest przeprowadzenie oceny skutków dla ochrony danych osobowych. Dobór odpowiednich technik pozwalających na ochronę prywatności, ale umożliwiających zachowanie wartości informacyjnych danego zbioru, w wypadku udostępniania danych do ponownego wykorzystywania jest bardzo ważną kwestią. Technikami, które pozwalają na uzyskiwanie korzyści z danych oraz minimalizują ryzyko w związku z utratą prywatności, jest anonimizacja lub pseudonimizacja.

Należy podkreślić, że konieczność zapewnienia ochrony danych osobowych nie powinna stać na przeszkodzie otwieraniu danych publicznych. Zastosowanie odpowiedniej techniki depersonalizacji danych lub kliku technik połączonych pozwala na pogodzenie obu wartości, tj. prywatności osób fizycznych i prawa do ponownego wykorzystywania danych publicznych.

Załączniki

NR 1. FORMULARZ OCENY PODSTAWY PRZETWARZANIA DANYCH

Formularz oceny podstawy przetwarzania danych						
				Realizowane operacje		
NR	Zagadnienie	Odpowiedź Tak/Nie	Czy wymagana jest anonimizacja?	Przed ujawnieniem danych publicznych	Ujawnienie danych Publicznych	Uwagi
1.	Czy określony zasób (np. rejestr) obejmuje dane osobowe?	Nie Np. dane statystyczne, anonimowe, etc.	Nie, chyba, że analiza ryzyka wykaże, że istnieje ryzyko, którego źródłem może być re-identyfikacja (należy pogłębić anonimizację)	Operacje realizowane na informacjach nie mających statusu danych osobowych		W przypadku danych anonimowych, zanonimizowanych, statystycznych: należy wypełnić załącznik nr 3 (tło oceny ryzyka) i dokonać oceny ryzyka w ramach załącznika nr 5 (chodzi o ryzyko re-identyfikacji)
		Tak	Zależy od odpowiedzi na poniższe pytania	zob. poniżej		

2.	Czy dane publiczne obejmują dane osobowe wyłączone spod rodo?	Tak	Nie (tj. możliwe jest ujawnienie informacji np. wskazanych w motywie 14 rodo, osoby zmarłe)	Operacje realizowane na danych wyłączonych spod rodo		Należy wypełnić załącznik nr 3 (tło oceny ryzyka) i dokonać oceny ryzyka w ramach załącznika nr 5
		Nie		zob. poniżej		
3.	Czy dane publiczne obejmują dane szczególnych kategorii?	Tak	Tak	Na etapie przed ujawnieniem danych publicznych dochodzi do przetwarzania danych osobowych (szczególnych kategorii)	Ujawnienie danych zanonimizowanych chyba, że istnieje przepis prawa umożliwiający ujawnienie takich danych do ponownego wykorzystania lub odebrano wyraźną zgodę, jeżeli spełnia wymóg minimalizacji	w przypadku ujawniania danych zanonimizowanych należy wypełnić załącznik nr 3 (tło oceny ryzyka) i dokonać oceny ryzyka w ramach załącznika nr 5
		Nie	Zależy od odpowiedzi na poniższe pytania	zob. poniżej		
4.	Czy dane publiczne obejmują dane osobowe osób pełniących funkcje publiczne w związku z pełnieniem funkcji?	Tak	Nie (tj. możliwe jest ujawnienie danych osobowych)	Na etapie przed ujawnieniem ISP dochodzi do przetwarzania danych osobowych	Ujawnienie ISP obejmuje dane osobowe	Należy wypełnić załącznik nr 3 (tło oceny ryzyka) i dokonać oceny ryzyka w ramach załącznika nr 5 Jeżeli wymagane będzie przeprowadzenie oceny skutków – należy wypełnić także załącznik nr 6
		Nie	Zależy od odpowiedzi na poniższe pytania	zob. poniżej		

5.	Czy dane publiczne obejmuje dane osób, które zrezygnowały z przysługującego prawa?	Tak	Nie (tj. możliwe jest ujawnienie danych osobowych)	Na etapie przed ujawnieniem danych publicznych dochodzi do przetwarzania danych osobowych	Ujawnienie ISP obejmuje dane osobowe	Należy wypełnić załącznik nr 3 (tło oceny ryzyka) i dokonać oceny ryzyka w ramach załącznika nr 5 Jeżeli wymagane będzie przeprowadzenie oceny skutków – należy wypełnić także załącznik nr 6
		Nie	Zależy od odpowiedzi na poniższe pytania			
6.	Czy dane publiczne obejmują dane informacje dotyczących innych osób niż pełniących funkcje publiczne, ale możliwe do ujawnienia w oparciu o art. 6 ust. 4 rodo?	Tak	Nie	Na etapie przed ujawnieniem danych publicznych dochodzi do przetwarzania danych osobowych (szczególnych kategorii)	Ujawnienie danych publicznych obejmuje dane osobowe	Należy wypełnić załącznik nr 3-5 (tło oceny ryzyka) i dokonać oceny ryzyka w ramach załącznika nr 5 Jeżeli wymagane będzie przeprowadzenie oceny skutków – należy wypełnić także załącznik nr 6
		Nie	Tak		Ujawnienie danych zanonimizowanych	Należy wypełnić załącznik nr 3 (tło oceny ryzyka) i dokonać oceny ryzyka w ramach załącznika nr 5

NR 2. WYKAZ OPERACJI I TYPOWYCH ŹRÓDEŁ RYZYKA NARUSZENIA PRAW LUB WOLNOŚCI

Wykaz operacji i typowych źródeł ryzyka naruszenia praw lub wolności			
Etap	**Operacje**	**Kategoria informacji**	**Przykładowe zdarzenia, które może być źródłem ryzyka naruszenia praw lub wolności**
Czynności przygotowawcze	Przeglądanie i analiza danych Strukturyzowanie danych Łączenie danych	Dane osobowe	• Nieuprawnione połączenie danych zebranych dla różnych celów; • Brak zidentyfikowania pierwotnego celu przetwarzania w ramach oceny zgodności celów; • Brak przeprowadzenia oceny zgodności celów w przypadkach wskazanych w art. 6 ust. 4 rodo; • Naruszenie zasady minimalizacji danych poprzez uwzględnienie danych osób pełniących funkcje publiczne poza zakresem pełnienia tych funkcji; • Naruszenie wymogu dobrowolności zgody, o której mowa w art. 6 ust. 2 upwisp; • Naruszenie zasady minimalizacji danych poprzez nieuwzględnienie, że zasób obejmuje nie tylko dane osób pełniących funkcje publiczne, ale także innych; • Ujawnienie danych wrażliwych ze względu na połączenie danych z różnych zasobów.
	Przeglądanie i analiza danyc Strukturyzowanie informacji Łączenie informacji	Dane nie-osobowe (anonimowe, statystyczne, wyłączone spod rodo, etc.)	• Możliwość szczątkowej identyfikacji; • Możliwość identyfikacji wskutek łączenia danych pochodzących z różnych źródeł (re-identyfikacja); • Błędna ocena co do statystycznego charakteru danych; • Błędnej oceny możliwości przypisania informacji do osoby fizycznej (np. błędna kwalifikacji potrzebnych środków i nakładów na identyfikację – zob. motyw 26 rodo); • Błędna kwalifikacja, iż rejestr jest objęty motywem 14 rodo; • Błędnej oceny w zakresie istnienia/nieistnienia zbioru danych osobowych – jeżeli będzie to istotne dla kwalifikacji (art. 2 rodo).
Etap decyzji co do ujawnienia danych	Anonimizacja	Dane osobowe poddane anonimizacji -> dane zanonimizowane	• Re-identyfikacja

	Pseudonimizacja	Dane osobowe poddane pseudonimizacji -> dane speduonimizowane (Uwaga! dane speudonimizowane to w dalszym ciągu dane osobowe)	• odwróceniem pseudonimizacji
Ujawnienie danych	Ujawnienie danych osobowych (w trybie wnioskowym albo bezwnioskowym	Dane osobowe poddane anonimizacji -> dane zanonimizowane	• Naruszenie zasady minimalizacji i ujawnienie zbyt szerokiego zakresu danych o osobach pełniących funkcje publiczne; • błędnej oceny, że w rejestrze nie występują dane wrażliwe (art. 9-10 rodo); • re-identyfikacja (możliwość szczątkowej identyfikacji / możliwość identyfikacji wskutek łączenia danych pochodzących z różnych źródeł; • możliwość szerszego wykorzystywania danych osobowych niż cele udostępniania; • Możliwość przetwarzania danych osobowych po ich usunięciu z zasobu publicznego; • Wykorzystanie danych niezgodnie z pierwotnym celem zebrania danych.
	Ujawnienie danych osobowych (w trybie wnioskowym albo bezwnioskowym)	Dane osobowe (poddane pseudonimizacji)	• j/w oraz: • odwróceniem pseudonimizacji.
	Ujawnienie danych anonimowych / statystycznych	Dane osobowe anonimowe / zanonimizowane / statystyczne	• re-identyfikacja;
Etap końcowy	Przechowywanie	Dane osobowe	• Przechowywanie danych dłużej niż to niezbędne do realizacji zadań wynikających z ustawy o ponownym wykorzystaniu ISP; • Przechowywanie danych nieaktualnych (nieprawidłowych); • Naruszenie wymogu poprawności danych (np. ze względu na brak wskazania, że określony zasób) jest aktualny na dany dzień
	Usuwanie		• Brak usunięcia danych pomimo spełnienia przesłanek z art. 17 rodo.

NR 3. OKREŚLENIE TŁA ANALIZY RYZYKA NARUSZENIA PRAW LUB WOLNOŚCI

Określenie tła analizy ryzyka naruszenia praw lub wolności						
• Tryb ujawnienia ISP: • czy dane będą mogły być wykorzystane także w przypadku braku określenia warunków ponownego wykorzystania? • pierwotny cel zebrania danych: • Ujawniony przez użytkownika cel ponownego wykorzystania ISP: • Sposób ujawniania danych publicznych (np. BIP): • czy wymagana będzie ocena zgodności celów przetwarzania? (TAK/NIE): • czy któraś z operacji mieści się w wykazie PUODO (wymagających oceny skutków)?						
Wskaż i uzasadnij o jaki przypadek ujawniania danych publicznych chodzi	Etap	Opis operacji (poniżej przykłady)	Czy operacja obejmuje przetwarzanie danych osobowych?	Kategorie osób (ze szczególnym wskazaniem os. pełniących funkcje publiczne) Czy chodzi o osoby pełniące funkcje publiczne?	Kategorie danych (w tym ze wskazaniem danych anonimowy)	Czy zakres danych jest niezbędny i zgodny z zasadą minimalizacji?
- przetwarzanie danych osobowych anonimowych (wykorzystanie i ujawniane będą dane anonimowe);	Przed ujawnieniem danych publicznych	Przeglądanie i analiza danych Strukturyzowanie danych Łączenie danych				

- przetwarzanie danych osobowych natomiast ujawnione będą dane zanonizowane; - przetwarzanie danych osobowych wyłączonych spod rodo (motyw 14); - wykorzystanie danych osobowych szczególnych kategorii – ujawniane będą dane, które zostały zanonimizowane; - przetwarzanie danych osób pełniących funkcje publiczne w związku z pełnieniem tych funkcji; - przetwarzanie danych osób, które zrezygnowały z przysługujących im praw (art. 6 ust. 2 upwisp); - przetwarzanie danych innych osób (niż wskazane w art. 6 ust. 2 upwisp) w celu ponownego wykorzystania – wstępnie założono, że ujawnione zostaną dane osobowe;	Ujawnienie ISP	Ujawnienie danych w trybie bezwnioskowym na stronie dane.gov.pl				
		Przechowywanie danych				
		Usuwanie				

NR 4. TEST ZGODNOŚCI CELÓW

Test zgodności celów (test z art. 6 ust. 4 rodo)	
Dodatkowe zagadnienia	**Odpowiedzi**
Jakich kategorii osób i danych dotyczy test zgodności celów?	
Czy chodzi o dane szczególnych kategorii (art. 9 ust. 1 rodo) lub o karalności (art. 10 rodo)?	
Jaki związek zachodzi pomiędzy celem ujawnienia danych a pierwotnym celem zebrania danych? Należy uwzględnić także cel ponownego wykorzystania, jeżeli został ujawniony przez użytkownika	
Jakie będzie poziom identyfikowalności osób, których dane dotyczą?	
Czy ujawniane będą dane już wcześniej dostępne publicznie? Jeżeli tak to w jaki sposób były one dostępne?	
Jaka będzie skala przetwarzania (ile orientacyjnie osób, rekordów, jaka będzie geograficzna skala, jak często dane będą pozyskiwane),	
Czy osoby, których dane mogą być ujawnione w ramach ponownego wykorzystania będą zaskoczone przetwarzaniem realizowanym w ten sposób?	
Czy przetwarzanie dotyczy grup osób wymagających szczególnej ochrony np. dzieci?	

Jakie będą konsekwencje zamierzonego dalszego przetwarzania dla osób, których dane dotyczą?	
Czy wdrożono lub zaplanowano określone mechanizmy ograniczenia identyfikowalności w stosunku do tych samych danych, które będą podlegały ujawnieniu? A jeśli tak, to jaki był powód zastosowania takich środków? (np. szyfrowanie lub pseudonimizację)	

NR 5. OCENA RYZYKA

Kontekst		**identyfikacja ryzyka**		**Szacowanie ryzyka**			**Postępowanie z ryzykiem**		**Ostateczna decyzja**
Operacje	Kategoria informacji	Źródło ryzyka Uwaga! Uwzględnij katalog typowych źródeł ryzyk dla poszczególnych operacji (załącznik nr 1-2)	Konsekwencje Jakie będą możliwe konsekwencje wystąpienia źródła ryzyka?	Prawdo podobieństwo ryzyka	Waga potencjalnych konsekwencji	Poziom ryzyka Czy jest on akceptowalny? Czy należy podjąć określone działań? Czy zidentyfikowano duże prawdopodobieństwo wysokiego ryzyka?	Środki wdrożone celem eliminacji lub redukcji ryzyka	Poziom ryzyka po wdrożeniu środków (uzasadnij, czy środki wpłynęły na wagę czy prawdo - podobieństwo)	Decyzja co do zidentyfikowanego poziomu ryzyka po uwzględnieniu zaplanowanych środków W szczególności czy: - konieczne jest uzupełnienie oceny poprzez dokonanie oceny skutków; - uprzednie konsultacje
Przeglądanie i analiza danych Strukturyzowanie danych Łączenie danych	Dane osobowe								

NR 6. OCENA NIEZBĘDNOŚCI I PROPORCJONALNOŚCI

Ocena	
Pytanie	**Odpowiedzi**
Czy operacje przetwarzania są niezbędne i proporcjonalne w stosunku do celów? W szczególności odpowiedz na pytanie: Jaka jest podstawa przetwarzania danych? Czy przetwarzanie faktycznie osiąga cel? Czy istnieje inny sposób na osiągnięcie tego samego rezultatu?	

NR 7. WYKAZ PODJĘTYCH CZYNNOŚCI

W szczególności wskaż:

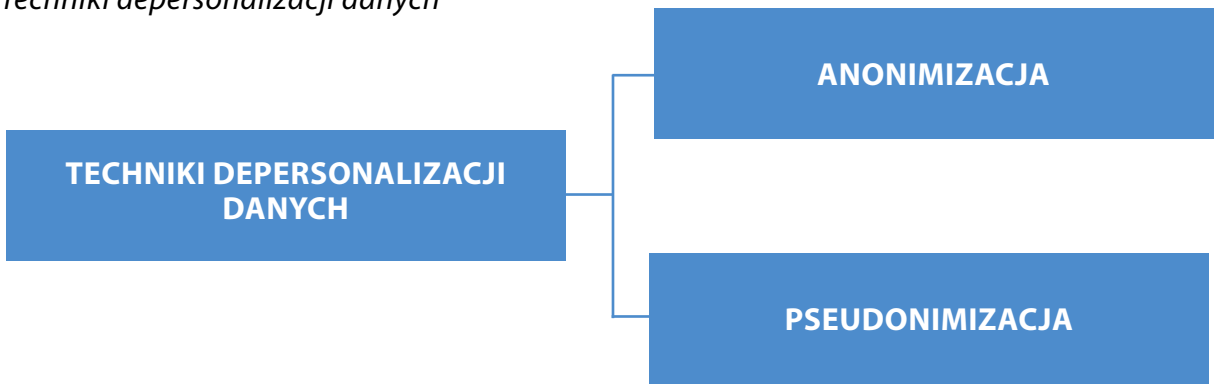
- kto oceniał ryzyko;
- kto decydował o akceptacji ryzyka;
- kto opiniował przeprowadzoną ocenę;
- kto dokonywał finalnej akceptacji przeprowadzonej oceny;

Formularz podjętych czynności			
Nr	**Podjęte czynności**	**Kto podejmował czynności**	**Ustalenia lub konsekwencje podjętych czynności**

NR 8. TECHNIKI DEPERSONALIZACJI DANYCH

Istnieją dwie główne techniki depersonalizacji danych osobowych. Jest to anonimizacja oraz pseudonimizacja. Z technicznego punktu widzenia największa różnica między nimi polega na tym, że pseudonimizacja jest procesem odwracalnym a anonimizacja już nie.

Rysunek 8.1. Techniki depersonalizacji danych



Techniki depersonalizacji danych posiadają zróżnicowany stopień odporności na czynniki ryzyka. Dla każdej z technik poniżej wskazane zostały podatności na trzy główne ryzyka zagrażające prywatności:

Wyodrębnienie: jest to możliwość wyizolowania niektórych lub wszystkich rekordów, które identyfikują daną osobę w zbiorze danych.

Tworzenie powiązań: jest to możliwość powiązania co najmniej dwóch rekordów dotyczących tego samego podmiotu danych (w tej samej bazie danych lub w dwóch różnych bazach danych).

Wnioskowanie: pozwala z dużym prawdopodobieństwem wydedukować wartość atrybutu z wartości zbioru innych atrybutów

1. Anonimizacja

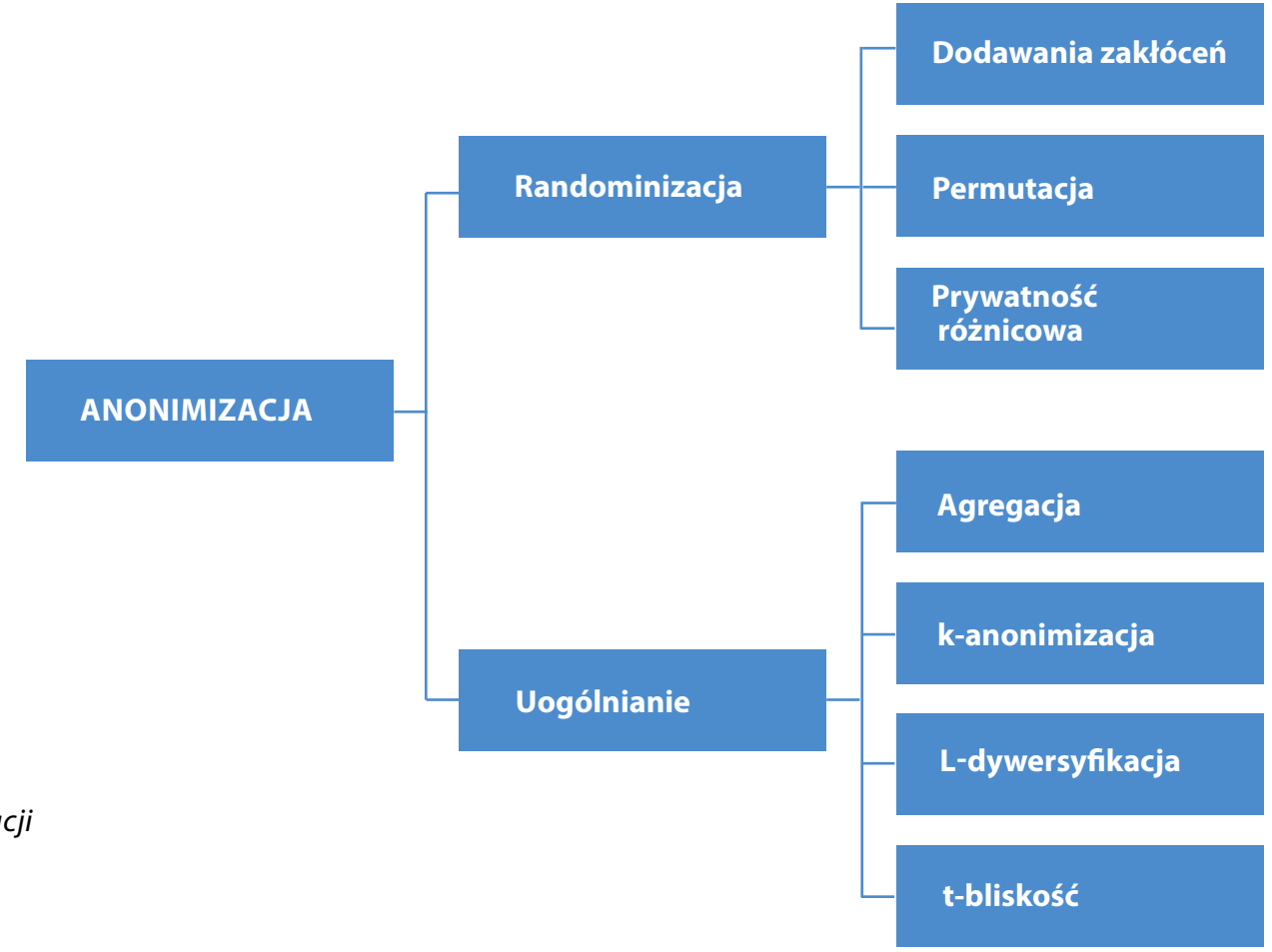
Anonimizacja jest to proces, w którym dane osobowe są trwale i nieodwracalnie przekształcone. Technika ta uniemożliwia (w rozsądnym wymiarze czasowym i finansowym) przyporządkowanie informacji o określonej lub możliwej do zidentyfikowania osobie fizycznej.

Randomizacja jest to rodzina technik, które zmieniają prawdziwość danych w celu wyeliminowania silnego związku między danymi a konkretną osobą. Dane charakteryzują się wystarczającą niepewnością i nie można ich już odnieść do konkretnej osoby. Randomizacja jednak sama w sobie nie zmniejsza osobliwości każdego rekordu, ponieważ każdy rekord będzie pochodził od pojedynczego podmiotu danych, ale może natomiast chronić przed zagrożeniami wynikającymi z wnioskowania. W celu zapewnienia większej ochrony zaleca się połączenie randomizacji z innymi technikami np. generalizacji.

Uogólnianie stanowi drugą rodzinę technik anonimizacji. Polega ono na uogólnianiu lub osłabieniu atrybutów dla pomiotów danych, których dane dotyczą, poprzez modyfikację skali lub rzędu wielkości (np.: podanie miast zamiast dzielnic, tygodnia zamiast dnia). Mimo, że uogólnianie może być skuteczne, aby zapobiec wyodrębnieniu, nie pozwala ono na skuteczną anonimizację we wszystkich przypadkach. Wymaga specyficznych i wyrafinowanych podejść ilościowych, aby zapobiec tworzeniom powiązań i wnioskowaniu.

a)dodawanie zakłóceń

Polega na modyfikowaniu atrybutów w zbiorze danych tak, aby były mniej dokładne przy zachowaniu ogólnej dystrybucji. Technika ta jest szczególnie użyteczna, gdy atrybuty mogą wywierać niekorzystny wpływ na poszczególne osoby. Zakłada się, że wartości danych będą prawdziwe (w znaczeniu wartości oryginalnych), ale tylko do pewnego zakresu. Przykładowo, jeżeli waga danej osoby została zmierzona z dokładnością do 0,5 kg, zbiór danych może zawierać wagę z dokładnością do +/-5 kg. Skuteczne zastosowanie tej techniki nie pozwoli osobie trzeciej na zidentyfikowanie konkretnej osoby. Osoba trzecia także nie będzie w stanie naprawić danych lub w inny sposób wykryć, w jaki sposób dane zostały zmodyfikowane.



Rysunek 8.1 Podział technik anonimizacji

Technika ta zwyczajowo jest łączona z innymi technikami anonimizacji takimi jak usunięcie oczywistych atrybutów oraz quasi-identyfikatorów. Poziom zakłóceń powinien zależeć od poziomu wymaganych informacji i wpływu na prywatność osób w wyniku ujawnienia chronionych atrybutów.

Tabela 8.1. Podatność na ryzyka metody dodawania zakłóceń

Wyodrębnienie	Tworzenie powiązań	Wnioskowanie
Możliwe jest wyodrębnianie zapisów danej osoby nawet, jeśli zapisy są mniej wiarygodne.	Można łączyć zapisy tej samej osoby, ale zapisy są mniej wiarygodne, a zatem prawdziwy zapis może być powiązany ze sztucznie dodanym (to znaczy z tzw. „szumem"). W niektórych przypadkach błędne przypisanie może narazić osobę, której dane dotyczą na znaczący, a nawet wyższy poziom ryzyka niż właściwy.	Wnioskowanie jest możliwe, ale wskaźnik sukcesu będzie niższy, a niektóre fałszywe trafienia (i fałszywe negatywy) są wiarygodne.

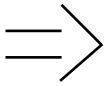
W technice tej często popełniane są dwa błędy. Pierwszym z nich jest dodawanie niespójnego zakłócenia. Jeżeli zakłócenie jest wykonane w sposób niesemantyczny (np. jest poza skalą lub jest nielogiczne) osoba trzecia będzie w stanie odfiltrować szum, a w niektórych przypadkach uzupełnić brakujące wpisy. Drugim błędem jest założenie, że samo dodanie zakłóceń jest wystarczające do zabezpieczenia danych. Technikę tę należy łączyć z innymi technikami anonimizacji, by w pełni zapewnić ochronę danych.

przedstawia przykład wprowadzenia metody zakłóceń do zbioru danych. Kolumny zaznaczone na czerwono (wiek, waga, wzrost) zostały poddane zakłóceniom, przez co określenie konkretnej osoby jest znacznie utrudnione.

Tabela 8.2. Zastosowanie metody dodawania zakłóceń

Wiek	Waga	Wzrost	Miasto
25	62	180	Lublin
20	80	160	Poznań
35	75	176	Opole
45	49	156	Kraków
25	82	162	Ełk

Tabela pierwotna



Wiek(+/-5)	Waga(+/-5)	Wzrost(+/-5)	Miasto
25	66	178	Lublin
22	79	165	Poznań
31	72	178	Opole
47	45	161	Kraków
24	79	162	Ełk

Tabela Zmodyfikowana

b)permutacja

Permutacja polega na przetasowaniu wartości atrybutów w tabeli tak, aby wartości dla jednego podmiotu danych były sztucznie przypisane do innego. Dane w zbiorze pozostają niezmienione, zmieniona jest natomiast korelacja między wartościami a poszczególnym podmiotem. Technika ta jest szczególnie przydatna, gdy ważne jest zachowanie dokładnego rozkładu atrybutów w zestawie danych.

Permutacja jest to szczególna forma dodawania zakłóceń. W klasycznej technice dodawania zakłóceń atrybuty są modyfikowane przy pomocy losowych wartości. Generowanie spójnego szumu może być trudnym zadaniem, gdyż modyfikowanie wartości atrybutów w nieznaczny sposób może nie zapewniać odpowiedniej prywatności. Alternatywnie, techniki permutacji zmieniają wartości wewnątrz zbioru danych, zamieniając je pomiędzy rekordami. Takie zamiany zapewniają, że zakres i rozkład wartości pozostaną niezmienione, a korelacje między wartościami a podmiotami nie będą występować. Należy jednak pamiętać, że jeśli dwa lub więcej atrybutów ma związek logiczny lub posiadają korelację statystyczną i są one permutowane niezależnie, relacja ta zostanie zerwana. Dlatego ważne jest, aby permutować zestaw powiązanych atrybutów, aby nie zerwać zależności logicznej. W przeciwnym razie osoba atakująca mogłaby zidentyfikować permutowane atrybuty i odwrócić permutację.

Tabela 8.3. Podatność na ryzyka metody permutacji

Wyodrębnienie	Tworzenie powiązań	Wnioskowanie
Możliwe jest wyodrębnienie zapisów danej osoby, nawet jeśli zapisy są mniej wiarygodne.	Może uniemożliwić „poprawne" powiązanie atrybutów zarówno wewnętrznie, jak i zewnętrznie z zestawem danych, ale nadal pozwala na „niepoprawną" powiązalność, ponieważ prawdziwy wpis może być powiązany z innym podmiotem danych.	Wnioskowanie jest możliwe, szczególnie jeśli atrybuty są skorelowane lub mają silne zależności logiczne; jednak nie wiedząc, które atrybuty zostały zmienione, atakujący musi wziąć pod uwagę, że jego wnioskowanie opiera się na błędnej hipotezie, a zatem możliwe jest jedynie wnioskowanie probabilistyczne

Najczęściej popełniane błędy w metodzie permutacji to:

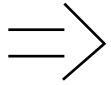
- **Wybór niewłaściwego atrybutu:** polega na wyborze do permutacji atrybutów niewrażliwych lub nieobarczonych ryzykiem, przez co nie poprawia się ochrony danych osobowych;
- **Osobna permutacja atrybutów:** jeśli atrybuty są silnie skorelowane, należy je permutować razem, wykonanie osobnej permutacji takich parametrów pozwala na ich identyfikację i odwrócenie permutacji;
- **Założenie, że sama permutacja jest wystarczająca:** permutacja sama w sobie nie zapewnia anonimowości i powinna być łączona z innymi technikami, takimi jak usuwanie oczywistych atrybutów.

Przykład zastosowania metody permutacji przedstawia Tabela 1-4 Kolumny zawierające wiek i płeć zostały poddane permutacji (osobno). Warto zauważyć, że w przypadku przeprowadzenia permutacji na atrybutach dotyczących zawodu i zarobków należy wykonać permutację razem, by uniknąć sprzeczności logicznych (np. sprzedawca zarabiałby więcej od menadżera).

Tabela 8-4. Zastosowanie metody permutacji

Zawód	Zarobki (tyś/ rok)	Rok ur.	Płeć
Sprzedawca	24	1990	K
Sprzedawca	48	1985	M
Kierowca	60	1992	K
Robotnik	72	1986	M
Menadżer	120	1975	M

Tabela pierwotna



Zawód	Zarobki (tyś/ rok)	Rok ur.	Płeć
Sprzedawca	24	1980	M
Sprzedawca	48	1990	M
Kierowca	60	1986	K
Robotnik	72	1985	M
Menadżer	120	1980	K

Tabela Zmodyfikowana

c)prywatność różnicowa

Prywatność różnicowa jest to technika randomizacji, która opiera się na innym podejściu niż wcześniej opisywane metody. Bazuje ona na wyborze odpowiedniego zanonimizowanego widoku danych przez dysponenta danych, przy czym zachowany jest oryginalny zbiór danych. Tak przygotowane zanonimizowane widoki danych przygotowywane są na podstawie zestawu zapytań generowanych przez osobę trzecią. Zbiór taki posiada celowo dodane zakłócenia po procesie anonimizacji. Metoda prywatności różnicowej pozwala dysponentowi danych na dodawanie zakłóceń do zbioru w odpowiedniej ilości i formie, który zapewnienia adekwatny poziom ochrony prywatności. Niezwykle istotne jest, by stale monitorować możliwości identyfikacji podmiotu danych w zbiorach, które powstają na podstawie zapytań. Należy jednak podkreślić, że metoda prywatności różnicowej zachowuje dane oryginalne w postaci niezmienionej i dysponent danych może zidentyfikować poszczególne osoby fizyczne.

Niewątpliwą korzyścią wynikająca z zastosowania prywatności różnicowej jest dostarczenie osobom trzecim zbioru danych na podstawie konkretnego zapytania, a nie poprzez udostepnienie całego zbioru danych. Dysponent danych może gromadzić takie zapytania, kontrolując tym samym dostęp osób trzecich do danych, do których nie mają upoważnienia. Samo zapytanie można również poddać anonimizacji, zwiększając tym samym poziom ochrony prywatności.

Nie należy udostępniać baz danych wykorzystujących metodę prywatności różnicowej w systemach wyszukiwania, które nie zapewniają identyfikacji podmiotów wprowadzających zapytanie. Wykorzystanie bardzo wielu zapytań może umożliwić identyfikację konkretnego podmiotu danych poprzez wnioskowanie lub tworzenie powiązań.

Tabela 8.5. Podatność na ryzyka metody prywatności różnicowej

Wyodrębnienie	Tworzenie powiązań	Wnioskowanie
W przypadku, gdy wynikiem zapytania są tylko statystyki, wyodrębnienie konkretnego podmiotu danych nie jest możliwe.	Wykorzystywanie dużej liczby zapytań może prowadzić do powiązania konkretnego podmiotu danych dwiema odpowiedziami.	Wykorzystywanie dużej liczby zapytań daje możliwość wnioskowania o konkretnym podmiocie danych.

Najczęściej popełnianym błędem w metodzie prywatności różnicowej to:

- **Wprowadzenie niewystarczającego zakłócenia:** w celu uniemożliwienia identyfikacji konkretnego podmiotu lub grupy podmiotów za pomocą dużej liczby zapytań, należy prawdziwe odpowiedzi uzupełnić o odpowiednią ilość zakłóceń.

d) agregacja i k-anonimizacja

Techniki agregacji i k-anonimizacji polegają na grupowaniu ze sobą danych o podmiotach. W tym celu wartości atrybutów są uogólnianie do takiego stopnia, że dla kilku (umownie k) podmiotów przypisana jest ta sama wartość. Można to osiągnąć poprzez np. zmniejszenie szczegółowości lokalizacji z gminy do województwa. Dla danych liczbowych (takich jak zarobki, wzrost, parametry medyczne) generalizację można wykonać poprzez podanie wartości przedziałowych (np. wzrost 180-190 cm). Techniki te mogą być stosowane w przypadku, gdy korelacja wartości punktowych atrybutów może utworzyć *quasi identyfikatory*.

Tabela 8-6. Podatność na ryzyka metody k-anonimizacji

Wyodrębnienie	Tworzenie powiązań	Wnioskowanie
Wyodrębnienie osoby z grupy k użytkowników nie powinno być już dłużej możliwe, ponieważ te same atrybuty są teraz udostępniane przez k użytkowników.	Pomimo ograniczenia tworzenia powiązań, możliwe jest łączenie rekordów według grup wpisów użytkowników. Prawdopodobieństwo, że dwa rekordy odpowiadają temu samemu pseudo-identyfikatorowi wynosi 1/k.	Wnioskowanie jest możliwe, zwłaszcza w przypadku gdy wiemy, że dana osoba należy do danej grupy.

Najczęściej popełniane błędy w metodzie k-anonimizacji to:

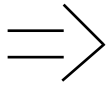
- **Brak niektórych quasi-identyfikatorów:** najważniejszym parametrem w przypadku techniki k-anonimizacji jest parametr k czyli wielkość określająca ile podmiotów danych powinno być w grupie. Im większa wartość parametru k tym silniejsze są gwarancje zachowania prywatności. Często spotyka się sytuację, że w celu zwiększenia wartości parametru k ogranicza się ilość quasi-identyfikatorów, co pozwala na budowanie większych klastrów k-użytkowników. Taki zabieg może prowadzić do tego, że istnieje możliwość identyfikacji jednostki w zbiorze poprzez wykorzystywanie innych dostępnych informacji dotyczących danego podmiotu.
- **Zbyt mała wartość k:** w przypadku zmniejszenia wartości parametru k waga każdego podmiotu danych w danej grupie jest znacząca. Istnieje wtedy podwyższone ryzyko identyfikacji podmiotu danych na podstawie wnioskowania.
- **Pomijanie grupowania podmiotów:** w przypadku grupowania zbioru danych o nierównej dystrybucji wpływ rekordu na daną grupę będzie się różnił w zależności od jego reprezentacji w danej grupie.
- **Grupowanie zestawu osób o nierównomiernym rozmieszczeniu atrybutów:** wpływ rekordu danej osoby na zbiór danych będzie się różnić: niektóre będą stanowić znaczną część wpisów, podczas gdy wkład innych osób pozostanie dość nieznaczny. Dlatego ważne jest, aby upewnić się, że k jest wystarczająco wysokie, aby żadne osoby nie stanowiły zbyt ważnej części wpisów w klastrze.

W przykładzie poniżej przedstawiono zastosowanie metody k-anonimizacji dla wartości k=2 (Tabela 1-). Zbiór został przekształcony w taki sposób, że dla minimum dwóch rekordów dane posiadają te same informacje. Uniemożliwia dokładną identyfikację konkretnej osoby.

Tabela 8-7. Zastosowanie metody k-anonimizacji (k=2)

Nazwisko	Wiek	Płeć	Powiat	Choroba
Nowak	29	K	Miński	Rak
Kowal	24	K	Otwocki	Psychiczna
Gajda	28	K	Miński	Astma
Halicki	27	M	Płocki	Brak
Górski	24	K	Otwocki	Serca
Golec	23	M	Płocki	Astma
Dłuski	19	M	Otwocki	Rak
Bogucki	29	M	Płocki	Serca
Leja	17	M	Otwocki	Serca
Nowak	19	M	Otwocki	Psychiczna

Tabela pierwotna



Nazwisko	Wiek	Płeć	Powiat	Choroba
*	20 < Wiek ≤ 30	K	Miński	Rak
*	20 < Wiek ≤ 30	K	Otwocki	Psychiczna
*	20 < Wiek ≤ 30	K	Miński	Astma
*	20 < Wiek ≤ 30	M	Płocki	Brak
*	20 < Wiek ≤ 30	K	Otwocki	Serca
*	20 < Wiek ≤ 30	M	Płocki	Astma
*	Wiek ≤ 20	M	Otwocki	Rak
*	20 < Wiek ≤ 30	M	Płocki	Serca
*	Wiek ≤ 20	M	Otwocki	Serca
*	Wiek ≤ 20	M	Otwocki	Psychiczna

Tabela Zmodyfikowana

Stosowane są także bardziej rozbudowane metody k-anonimizacji takie jak (X,Y) –anonimizacji, (α,k)-anonimizacji, (k,e)-anonimizacji[24].

e)l-dywersyfikacja i t-bliskość

Metoda L-dywersyfikacji to rozszerzenie metody k-anonimizacji. Przyjmuje się w niej, że dla każdej grupy wartości powinno występować co najmniej L różnych wartości. Rozumie się przez to odpowiednie dobranie wartości atrybutu tak, aby zapewnić odpowiednią liczność, a z drugiej strony poprawność dziedzinową.

[24] Liber, A. (2014). Problemy anonimizacji dokumentów medycznych. Część 1. Wprowadzenie do anonimizacji danych medycznych. Zapewnienie ochrony danych wrażliwych metodami f (a)-if (a, b)-anonimizacji. Higher School's Pulse, 8(1). Liber, A. (2014). Problemy anonimizacji dokumentów medycznych. Część 2 Anonimizacja zaawansowana oraz sterowana przez posiadacza danych wrażliwych.

Dzięki takiemu podejściu technika ta jest odporna na ataki wnioskowania. W celu poprawnego zastosowania, w metodzie tej należy ograniczyć klasy o niskiej zmienności atrybutów. Dzięki temu osoba atakująca ma zawsze znaczną niepewność odnośnie do konkretnego podmiotu danych, równą odwrotności dywersyfikacji wynoszącą 1/L.

Metoda L-dywersyfikacji skutecznie chroni przed atakami polegającymi na wnioskowaniu dla zbiorów o stosunkowo dobrze rozłożonych wartościach atrybutów. W przypadku zbioru danych o niewielkim zakresie lub o nierównym rozmieszczeniu metoda ta nie jest już odporna na ataki wnioskowania.

Metoda t-bliskości jest udoskonaleniem metody L-dywersyfikacji. Ma ona za zadanie stworzenie równoważnych klas, które przypominają początkowy rozkład atrybutów w tabeli. Nakłada ona ograniczenia na rozkłady prawdopodobieństwa występowania wartości atrybutów w grupach oraz w całej tabeli. Dąży się do tego, aby oba rozkłady były jak najbliższe oryginałowi.

Tabela 8-8. Podatność na ryzyka metody L-dywersyfikacji i t-bliskości

Wyodrębnienie	**Tworzenie powiązań**	**Wnioskowanie**
Wyodrębnienie konkretnego podmiotu danych nie powinno być już możliwe.	Pomimo ograniczenia tworzenia powiązań, możliwe jest łączenie rekordów według grup wpisów użytkowników	Nie jest możliwe ze 100% pewnością określenie danego podmiotu danych na podstawie wnioskowania.

Najczęściej popełniany błąd w metodzie L-dywersyfikacji i t-bliskości to:

- **Zabezpieczenie szczególnie chronionych wartości atrybutów poprzez zmieszanie ich z innymi atrybutami szczególnie chronionymi:** dwie wartości atrybutu w klastrze nie wystarczają do zapewnienia prywatności. W praktyce dystrybucja wartości szczególnie chronionych w każdym klastrze powinna przypominać dystrybucję tych wartości w całej populacji lub przynajmniej powinna być ona jednakowa w całym klastrze.

Tabela 8-8 przedstawia przykład metody L-dywersyfikacji trzeciego poziomu (L =3)., Oznacza to, że dla każdej grupy wartości pseudo-identyfikatora występują przynajmniej trzy „dobrze reprezentowane" rekordy danych źródłowych (tu konkretnie informacje o dochodzie i schorzeniu). Lewa strona tabeli jest oryginalną tabelą, prawa strona pokazuje wersję zanonimizowaną spełniającą 3-dywersyfikację. Za wrażliwe atrybuty przyjęto tutaj informacje dotyczące wynagrodzenia i choroby.

Tabela 8-9. Zastosowanie metody L-dywersyfikacji (L=3)

	Kod pocztowy	Wiek	Dochód	Schorzenie
1	47677	29	3 000	wrzody
2	47602	22	4 000	nieżyt żołądka
3	47678	27	5 000	rak żołądka
4	47905	43	6 000	nieżyt żołądka
5	47909	49	11 000	grypa
6	47906	47	8 000	zapalenie oskrzeli
7	47605	30	7 000	zapalenie oskrzeli
8	47673	36	9 000	zapalenie płuc
9	47607	32	10 000	rak żołądka
10	47909	52	9 000	nieżyt żołądka
11	47905	59	12 000	grypa
12	47908	61	10 000	rak żołądka

⇒

Kod pocztowy	Wiek	Dochód	Schorzenie
476**	29	3 000	wrzody
476**	22	4 000	nieżyt żołądka
476**	27	5 000	rak żołądka
4790*	43	6 000	nieżyt żołądka
4790*	49	11 000	grypa
4790*	47	8 000	zapalenie oskrzeli
476**	30	7 000	zapalenie oskrzeli
476**	36	9 000	zapalenie płuc
476**	32	10 000	rak żołądka
4790*	52	9 000	nieżyt żołądka
4790*	59	12 000	grypa
4790*	61	10 000	rak żołądka

Źródło: Li, N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 106-115). IEEE. https://www.cs.purdue.edu/homes/ninghui/papers/t_closeness_icde07.pdf

10 przedstawia przykład metody t-bliskości. Lewa strona tabeli jest oryginalną tabelą, prawa strona pokazuje wersję zanonimizowaną spełniającą odpowiednie założenia dotyczące rozkładu wynoszące dla dochodu t=0,167, a dla schorzenia t=0,278

Tabela 8-10. Zastosowanie metody t-bliskości

ID	Kod pocztowy	Wiek	Dochód	Schorzenie
1	47677	29	3 000	Wrzody
2	47602	22	4 000	nieżyt żołądka
3	47678	27	5 000	rak żołądka
4	47905	43	6 000	nieżyt żołądka
5	47909	52	11 000	Grypa
6	47906	47	8 000	zapalenie oskrzeli
7	47605	30	7 000	zapalenie oskrzeli
8	47673	36	9 000	zapalenie płuc
9	47607	32	10 000	rak żołądka

⇒

ID	Kod pocztowy	Wiek	Dochód	Schorzenie
1	476**	≤40	3000	wrzody
3	476**	≤40	5 000	rak żołądka
8	476**	≤40	9 000	zapalenie płuc
4	4790*	≥40	6 000	nieżyt żołądka
5	4790*	≥40	11 000	grypa
6	4790*	≥40	8 000	zapalenie oskrzeli
2	476**	≤40	4 000	nieżyt żołądka
7	476**	≤40	7 000	zapalenie oskrzeli
9	476**	≤40	10 000	rak żołądka

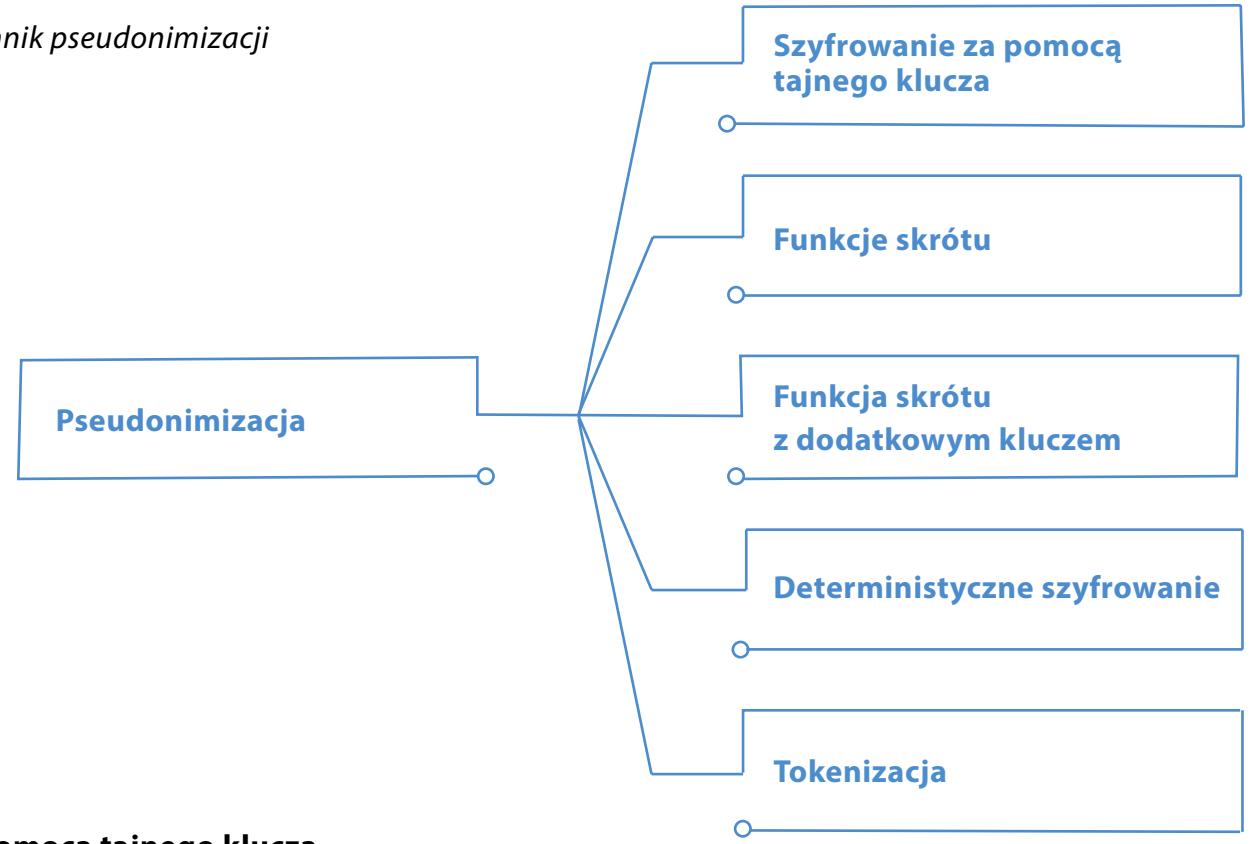
Źródło: Li, N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 106-115). IEEE. https://www.cs.purdue.edu/homes/ninghui/papers/t_closeness_icde07.pdf

2.Pseudonimizacja

Pseudonimizacja jest to odwracalny proces, polegający na zastąpieniu danej rzeczywistej nazwą przybraną, czyli nadanie jej tak zwanego pseudonimu. Pseudonimizacja utrudnia identyfikację, natomiast umożliwia przypisanie różnych czynności tej samej osobie (bez znajomości jej danych osobowych) oraz łączenie różnych zbiorów danych między sobą. Pseudonimizacja skutecznie podwyższa bezpieczeństwo przetwarzania danych, ale nie jest równoznaczna anonimizacji, w związku z czym dane poddane pseudonimizacji dalej podlegają pełnej ochronie.

Poniższy rysunek prezentuje podział pseudonimizacji na pięć głównych kategorii.

Rysunek 8-3. Podział technik pseudonimizacji



a)szyfrowanie za pomocą tajnego klucza

Dane osobowe są nadal przechowywane w zbiorze danych, ale w formie zaszyfrowanej. Posiadanie klucza szyfrującego pozwala na pełen dostęp do danych osobowych. Używając szyfrowania, które zachowuje aktualne standardy bezpieczeństwa, możliwość odszyfrowania danych jest możliwa, ale tylko z użyciem klucza szyfrującego.

b)funkcje skrótu

Polega na skróceniu każdej wielkości w danej liście atrybutów do stałej określonej wartości. Funkcji tej nie można odwrócić, tak jak w przypadku szyfrowania. Jakkolwiek, znając zakres wartości, jakie zostały poddane skracaniu oraz w jaki sposób zostało to wykonane, możliwe jest odtworzenie funkcji skrótu i uzyskanie prawidłowego zapisu poprzez tzw. atak siłowy (wypróbowanie wszystkich możliwych kombinacji w celu utworzenia tabel korelacji).

Funkcje skrótu można podzielić ze względu na wielkości bloku wyjściowego (ilość bitów). Obecne zalecenia amerykańskiej agencji NIST[25] dotyczące stosowania poszczególnych funkcji skrótu mówią, że do nowych aplikacji zalecane są funkcje skrótu z rodziny SHA-2[26], a w przyszłości funkcja SHA-3[27]. Do niedawna stosowane były funkcje skrótu MD5[28] oraz SHA-1[29], zostały on jednak wycofane ze względu na niewystarczający poziom zabezpieczeń. Tabela 2- prezentuje przykłady bloków wyjściowych dla różnych funkcji skrótu. Istnieją różnego rodzaju darmowe oprogramowania generujące funkcje skrótu: np. Gperf[30], CCP-Crypto[31].

Tabela 8-11. Funkcje skrótu dla wybranej frazy

Funkcja skrótu	Blok wyjściowy (wejściowa fraza: Jan Kowalski)
SHA-2*	e70207d004bdffb1702b9c50f5a8c70a49f6f8a-68404ca59993d3c39142162ba1bac5cc2c396460c205b0a-79676c736ae345257aae26c61444f3efc1efa822c0
SHA-3*	88f2a580725f980b96829210bf3e7a159a5ce23a13f289d-077c3e45ef0c5a3876a81618f70c6641f14cd4fc58a5cf80d-928909187c772108e9342051e632fdbb

*Wersja 512 bitowa

Źródło: Opracowanie własne na podstawie https://emn178.github.io/online-tools/

[25] National Institute of Standards and Technology
[26] SHA-2 składa się z zestawu czterech funkcji dających skróty wielkości 224, 256, 384 lub 512 bitów
[27] SHA-3 funkcja oparta o Algorytm Keccak charakteryzuje się wyższą wydajnością niż SHA-2 zarówno w implementacjach sprzętowych jak i programowych.
[28] MD5 – generuje z ciągu danych o dowolnej długości128-bitowy skrót
[29] SHA-1 tworzą 160-bitowy skrót z wiadomości o maksymalnym rozmiarze 264 bitów i jest oparty na podobnych zasadach co MD5
[30] https://www.gnu.org/software/gperf/
[31] https://github.com/torvalds/linux/blob/master/drivers/crypto/ccp/ccp-crypto-main.c

c) funkcja skrótu z dodatkowym kluczem

Jest to rozbudowana funkcja skrótu, która dodatkowo wykorzystuje do ochrony prywatności tajny klucz. Możliwość odszyfrowania danych nie jest praktycznie możliwa przez osoby trzecie, ze względy na bardzo dużą liczbę powstałych kombinacji do deszyfracji. Dysponent danych może natomiast bez problemu powrócić do pierwotnej formy danych wykorzystując tajny klucz. Najpowszechniejszej wykorzystywanymi funkcjami są HMAC i UMAC[32].

Tabela 8-12. Funkcje skrótu z dodatkowym kluczem dla wybranej frazy

Funkcja skrótu z kluczem	Blok wyjściowy (wejściowa fraza: Jan Kowalski z kluczem Cyfryzacja)
SHA-2*	cf7449fcdeea860bbfc503410520d6623dd40a1e
SHA-3*	2fd3e276606e45b474e07fcca1b605e3127f03ce9fb11aec09a5f3cda4f315b41a-6ecdfd94016edf49ab35fb1478f13247c7b23c7e5e509e68238ad9088bec70

Źródło: Opracowanie własne na podstawie https://www.freeformatter.com/hmac-generator

d) deterministyczne szyfrowanie

Technika ta polega na wygenerowaniu pseudonimu w postaci losowej liczby dla każdego atrybutu, a następnie usunięciu tabeli powiązań. Dzięki temu zabiegowi znacznie ograniczone jest ryzyko identyfikacji poszczególnego podmiotu danych na podstawie tworzenia powiązań z innymi tabelami, gdzie jest on wymieniony z innym pseudonimem.

e) tokenizacja

Metoda ta polega na wykorzystaniu jednokierunkowych mechanizmów szyfrujących opartych na przypisaniu identyfikatora (indeksu, sekwencji lub losowo wygenerowanej liczby) w żaden sposób niezwiązanego z pierwotnymi danymi. Technika ta jest często spotykana w sektorze finansowym do autoryzacji operacji bankowych.

[32] HMAC / UMAC- kod MAC z wmieszanym kluczem tajnym zapewniający zarówno ochronę integralności jak i autentyczności danych

Tabela 8-13. Podatność na ryzyka metod pseudonimizacji

Wyodrębnienie	Tworzenie powiązań	Wnioskowanie
Możliwe jest wyodrębnienie konkretnego podmiotu danych, ponieważ jest on identyfikowany przez unikalny pseudonimowany atrybut.	Tworzenie powiązań jest możliwe i proste na podstawie innych niepseudonimizowanych atrybutów. Ryzyko jest wykluczone tylko jeśli żaden inny atrybut w zestawie danych nie może zostać wykorzystany do zidentyfikowania podmiotu danych i jeśli wyeliminowano wszystkie powiązania między pierwotnym atrybutem a atrybutem pseudonimicznym (w tym przez usunięcie oryginalnych danych).	Wnioskowanie o konkretnym podmiocie danych jest możliwe, jeśli w jednym zbiorze lub w kilku zbiorach do pseudonimizacji wykorzystywane są te same atrybuty, lub jeśli pseudonimy nie ukrywają tożsamości w wystarczający sposób.

Najczęściej popełniane błędy w technikach pseudonimizacji to:

- **Uznanie pseudonimizacji za anonimizację:** powszechnie panuje opinia, że usunięcie jednego lub kilku atrybutów danych lub zastąpienie ich pseudonimizowanymi danymi jest wystarczające do ochrony prywatności danej osoby. Należy podjąć dodatkowe działania mające na celu zwiększenie anonimizacji zbioru danych poprzez uogólnianie lub agregację atrybutów w takim stopniu, by niemożliwa była identyfikacja danego podmiotu danych.
- **Użycie tego samego klucza w różnych bazach:** zastosowanie za każdym razem innego klucza do pseudonimizacji danych pozwala w znacznym stopniu ograniczyć możliwości tworzenia powiązań między bazami, a tym samym zmniejsza ryzyko identyfikacji podmiotu danych. Stosowanie tego samego klucza w różnych bazach znacznie podwyższa ryzyko naruszenia prywatności.
- **Używanie różnych kluczy w bazie:** zastosowanie w jednej bazie niejednolitego klucza (np. w przypadku dodania lub zmiany wpisów dla jednego z podmiotów danych) może spowodować powstawanie unikalnych wzorców, co skutkuje dodatkową informacją na temat podmiotu, która może umożliwić jego identyfikację.
- **Przechowywanie klucza:** w przypadku przechowywania w jednym miejscu klucza wraz z danymi źródłowymi, ewentualne naruszenie bezpieczeństwa (np. wykradzenie bazy wraz z kluczem) skutkuje bardzo szybkim powiązaniem danych i ujawnieniem tożsamości podmiotów danych, których dane dotyczą. Należy także pamiętać, że przechowywanie klucza w sposób niezapewniający wysokiej ochrony może wywołać ten sam skutek.

Podsumowanie technik

Obecnie techniki depersonalizacji danych osobowych są intensywnie rozwijane. Powyżej opisane metody są to najczęściej wykorzystywane i rekomendowane do ochrony prywatności. Należy jednak pamiętać, że każdy zbiór danych należy rozpatrywać osobno, a po analizie ryzyka dostosować wybraną technikę do zakresu danych i celu, w jakim dane mają być udostępnione.

Tabela 8-14. Podatność na ryzyka metod anonimizacji i pseudonimizacji danych

Technika depersonalizacji danych	Ryzyko wyodrębnia?	Ryzyko tworzenia powiązań?	Ryzyko związane z wnioskowaniem?
Dodawanie zakłóceń	Tak	Być może nie	Być może nie
Permutacja	Tak	Tak	Być może nie
Agregacja lub k-anonimizacja	Nie	Tak	Tak
L-dywersyfikacja	Nie	Tak	Być może nie
Prywatność różnicowa	Być może nie	Być może nie	Być może nie
Tokenizacja	Tak	Tak	Być może nie
Pseudonimizacja	Tak	Tak	Tak