

Warszawa, dn. 15 października 2026 r.

Anonimowe, ale dla kogo? Projekt wytycznych EROD o anonimizacji

7 lipca 2026 r. EROD przyjął do konsultacji projekt wytycznych 02/2026 o anonimizacji. Aktualizuje on opinię WP29 05/2014 i zmienia sposób myślenia o często powtarzanym zdaniu „dane są zanonimizowane”, które tak chętnie wpisujemy do umów. Według wytycznych anonimowość nie jest cechą zbioru. Anonimowość musi być uzasadniona wobec konkretnych podmiotów, które przetwarzają dane.

Najpierw oczywistość – dane anonimowe nie są danymi osobowymi w myśl RODO. Autorzy wytycznych jednak podkreślają, że ta anonimowość jest względna. Ta sama tabela może być daną osobową dla administratora, który ma dodatkowe informacje, a równocześnie anonimową dla odbiorcy, który nimi nie dysponuje. Dlatego pierwszym krokiem jest ustalenie, dla kogo dane mają być anonimowe i dokonanie oceny z jego perspektywy. Trzeba dodać, że podmiot przetwarzający patrzy na dane oczami zleceńodawcy, czyli jeśli dla administratora dane można powiązać z konkretnymi osobami, to dla procesora będą to również dane osobowe.

Dobrze obrazuje to wyrok TSUE w sprawie SRB (C-413/23 P). Jednolita Rada ds. Restrukturyzacji i Uporządkowanej Likwidacji przekazała firmie Deloitte uwagi akcjonariuszy i wierzycieli Banco Popular, a ich nazwiska zastąpiła kodami. Uznała, że skoro Deloitte i tak nikogo nie rozpozna, nie musi informować o takim odbiorcy. TSUE potwierdził, że dane pseudonimizowane nie zawsze i nie dla każdego są danymi osobowymi, ale Radzie to nie pomogło. Obowiązek informacyjny ocenia się z perspektywy administratora w chwili zbierania danych, a Rada wiedziała przecież, czyje są to uwagi. Wytyczne dodają do tego regułę działającą w drugą stronę. Jeśli dane mogą realnie trafić do odbiorcy, a nie da się wykluczyć, że ma on rozsądnie prawdopodobne sposoby identyfikacji, są osobowe także dla przekazującego.

Co to znaczy, że ktoś jest identyfikowalny?

Zgodnie z wytycznymi zidentyfikowanie kogoś możemy tłumaczyć jako wyróżnienie go spośród innych na tyle, aby dało się go potraktować inaczej niż resztę. Nazwisko nie jest do tego potrzebne. EROD ilustruje to na przykładzie ulicy, na której jest sporo osób w czarnych garniturach i sporo osób z psami, ale tylko jedna ma i garnitur, i psa. Nikt nie zna jej nazwiska, a jednak da się ją bez trudu wskazać. Ryzyko nie musi być przy tym zerowe, ma być w rzeczywistości nieistotne. Pod uwagę bierze się tylko rozsądnie prawdopodobne sposoby identyfikacji, o których mówi motyw 26 RODO. Za orzecznictwem TSUE nie są nimi sposoby niemożliwe do zastosowania, wymagające nieproporcjonalnego wysiłku albo zakazane prawem.

Ciekawsze jest to, czego EROD nie uznaje za argument. Zapewnienie, że nikt nie będzie chciał nikogo identyfikować, nie wystarczy, gdyż motywacja może się zmienić, a do identyfikacji może dojść przez przypadek. Zakaz ustawowy co do zasady wystarcza, bo można założyć, że ludzie przestrzegają prawa. To założenie da się jednak obalić, jeśli są dowody na konkretne ryzyko, że zakaz zostanie złamany. Świadczyć o tym może to, że nikt go nie egzekwuje, jego łamanie się opłaca, dane są słabo zabezpieczone albo w podobnych sytuacjach zakaz był już obchodzony. Najstabilniej wypada zakaz wpisany do umowy. Według EROD to nie jest zakaz prawny, a umowa może jedynie uzupełniać środki techniczne. Myśleć trzeba przy tym o każdym, kto może dostać dane do ręki: od nieuczciwego pracownika, przez sąsiada i dziennikarza śledczego, po zagraniczne służby i cyberprzestępców.

Do tego dochodzi czas. Zbiór dziś anonimowy nie musi takim pozostać, bo techniki identyfikacji tanieją, a danych w obiegu przybywa. Gdy ryzyko przestanie być nieistotne, dane znów stają się osobowe, a administrator odpowiada za ich przetwarzanie. Rozwój AI, szczególnie agentów,

prawdopodobnie obniży czas i koszty stosowania technik ponownej identyfikacji, w tym wyszukiwania i łączenia informacji z wielu źródeł.

Trzy pytania do każdego zbioru

Do oceny czy dane rzeczywiście są anonimowe, EROD proponuje trzy kryteria. Kto pamięta opinię WP29 z 2014 r., rozpozna w nich dawne wyodrębnienie, możliwość powiązania i wnioskowanie, tyle że zostały teraz sformułowane na nowo jako: brak izolacji rekordu, brak powiązań i brak wnioskowania (*No Record Isolation, No Linkage, No Inference*). W praktyce sprowadza się to do trzech pytań. Czy któryś rekord jest jedyny w swoim rodzaju? Czy da się go połączyć z innym zbiorem, w którym występuje ta sama osoba? Czy z danych można wyciągnąć wniosek o konkretnym człowieku? Trzy razy „nie” oznacza, że dane można uznać za anonimowe.

Odpowiedź „tak” nie kończy jednak sprawy. Unikalny rekord jest tylko sygnałem ostrzegawczym. Rozstrzyga to, czy da się go dopasować do konkretnego człowieka, na przykład przez cechy z publicznych rejestrów. Ankieta w kinie z trzydziestoma pytaniami o film da prawie same unikalne zestawy odpowiedzi, a mimo to może pozostać anonimowa, jeśli nie ma ich z czym zestawić. Dawne wyodrębnienie nie zniknęło więc z wytycznych, tylko przesunęło się o krok dalej.

Najwięcej niuansów kryje wnioskowanie. Anonimowość psuje tylko taki wniosek, który dotyczy jednej osoby i jest istotny, czyli wpływa na jej prawa i wynika z jej danych, a nie z wiedzy ogólnej. Inny przykład z wytycznych, Bank uczy się na zanonimizowanej historii kredytów, że klienci o pewnym profilu częściej nie spłacają rat, i na tej podstawie ocenia nowego wnioskodawcę. Wniosek dotyczy konkretnej osoby, ale nie wynika z jej danych, bo nie było jej w zbiorze, więc zbiór może pozostać anonimowy. Inaczej jest ze ścieżkami lokalizacji. Jeśli ktoś w dni robocze od 20 do 8 rano przebywa w jednym budynku, a od 9 do 17 w innym, z punktów na mapie robi się jego adres domowy i miejsce pracy.

Samą ocenę można prowadzić na dwa sposoby. Podejście uproszczone zakłada, że jeżeli ktokolwiek mógłby kogoś zidentyfikować, dane mają charakter osobowy. Podejście kontekstowe sprawdza, kto realnie może dostać dane i co potrafi z nimi zrobić. Jest dokładniejsze, ale łatwiej w nim coś przeoczyć. Wytyczne wskazują, że często pomaga połączenie obu podejść. Można zacząć od uproszczonego i w razie potrzeby przejść do kontekstowego. Jeśli jednak już wstępne mapowanie danych pokazuje, że większość teoretycznych sposobów identyfikacji jest w praktyce nierealna, można od razu sięgnąć po kontekstowe.

Gdzie zawodzą popularne metody?

W opinii z 2014 r. WP29 opisywała konkretne techniki, takie jak randomizacja czy uogólnianie. Nowy projekt takiego katalogu nie przewiduje. Podaje za to ogólną regułę: im więcej atrybutów i im są dokładniejsze, tym łatwiej kogoś rozpoznać, a dane jednostkowe są dużo bardziej narażone niż zagregowane. Słabe punkty metod, po które sięgamy najczęściej, EROD pokazuje na przykładach.

Najprostsza z nich to usunięcie imion, nazwisk i adresów e-mail. Sklep z grami planszowymi z przykładu 11 dokładnie to robi, ale zestawy kupionych gier są na tyle charakterystyczne, że da się je dopasować do kolekcji, którymi klienci chwalą się w popularnym serwisie dla graczy. Brak nazwiska niczego tu nie zmienia, skoro dane same wskazują na konkretną osobę.

Pseudonim zamiast nazwiska nie pomoże, jeśli łatwo go odwrócić albo jeśli sam pozwala odróżnić osobę od innych. W przykładzie nr 6 firma reklamowa nie zna niczyjego nazwiska, ale pseudonim zbudowany z cech urządzenia wystarcza jej, żeby śledzić użytkownika między stronami i dobierać mu reklamy. Szyfrowanie z kolei ogranicza dostęp do danych, ale z założenia jest odwracalne, więc samo w sobie anonimizacją nie jest.

Uogólnianie zmniejsza szczegółowość danych: zamiast daty urodzenia zostaje rok, zamiast kodu pocztowego - jego pierwsze cyfry. W przykładzie 12 żaden rekord nie jest już unikalny, ale obaj mężczyźni urodzeni w 1946 z kodem zaczynającym się od 54, mają tę samą diagnozę. Kto wie, że jego sąsiad z tego samego rocznika trafił do zbioru, pozna jego chorobę. To znana słabość k-anonimowości: rekordy przestają się wyróżniać, ale wszyscy w grupie mają tę samą cechę, więc wystarczy wiedzieć, że ktoś do niej też przynależy.

Szum, czyli celowe zniekształcanie wartości, też nie daje pewności, bo rekordy da się łączyć nawet wtedy, gdy są niedokładne. EROD powołuje się tu m.in. na badanie Narayana i Shmatikova z 2008 r. Autorzy pokazali w nim, że część użytkowników z rzekomo anonimowego zbioru ocen filmów, który pewny popularny dostawca udostępnił w ramach konkursu, da się rozpoznać, zestawiając ich oceny z tymi, które publicznie wystawili w serwisie.

Najlepiej wypada zwykle agregacja, bo sumy i średnie dotyczą grup, ale i tu są pułapki. Firma z przykładu 16 publikuje tylko łączną kwotę wynagrodzeń inżynierów, a kto ma też jej wewnętrzny raport, wyliczy, że szósty inżynier zarabia 70 tys. euro rocznie. W przykładzie 17 menedżer dostaje zbiorcze wyniki ankiety, ale skoro wypełnił ją cały zespół i nikt nie dał oceny wyższej niż 3, wie, że wszyscy krytycznie oceniają kulturę firmy. Czasem brak wartości mówi więcej niż sama wartość.

Sama anonimizacja też podlega RODO

Łatwo zapomnieć, że anonimizacja jest przetwarzaniem danych osobowych. Potrzebuje więc podstawy z art. 6 RODO, a przy danych szczególnych kategorii także wyjątku z art. 9 ust. 2. Jeśli jest częścią tej samej czynności przetwarzania i służy tym samym celom, przyjmuje się, że podstawy są takie same jak wcześniej. Dla tych, którzy opierają się na prawnie uzasadnionym interesie, jest też dobra wiadomość: szybka anonimizacja połączona z usunięciem danych źródłowych przemawia na ich korzyść w teście równowagi.

Uwaga też na słownictwo. Ludzie powinni wiedzieć, że z ich danych powstaną dane anonimowe, a słów „anonimowe” czy „zdepersonalizowane” nie wolno używać, jeśli kogoś nadal da się rozpoznać. Cały proces, razem z testami, trzeba udokumentować i zachować tę dokumentację także po zakończeniu anonimizacji. Jeśli anonimizacja zadziałała tylko wobec części osób, a części zbioru nie przetwarza się osobno, całość trzeba traktować jak dane osobowe. A jeśli wyciekną informacje, na których poufności opierała się anonimowość, ocenę trzeba powtórzyć. Kto oceniał swoje zbiory według opinii z 2014 r., nie musi robić tego od nowa, choć okresowy przegląd z pewnością nie zaszkodzi 😊

Czego moim zdaniem brakuje w projekcie?

Najbardziej brakuje mi odpowiedzi na pytanie, które w praktyce pada często: czy zamiast usuwać dane po upływie okresu przechowywania, można je po prostu zanonimizować?

Projekt wprost zastrzega, że nie wyjaśnia przepisów nakazujących usunięcie danych. Wynika z niego tylko tyle, że jeśli oryginał zostaje w systemie, anonimizacja kopii nie pomaga w realizacji zasady ograniczenia przechowywania.

Do tego rzetelne podejście kontekstowe wymaga kompetencji, którymi wiele firm nie dysponuje. EROD zachęca do rozwijania i wykorzystywania wiedzy eksperckiej, ale mniejszym administratorom nie daje prostych narzędzi ani progów, więc w praktyce zostaje im podejście uproszczone.

Najmniej konkretność jest przy modelach AI i danych syntetycznych. Projekt wskazuje, że odpowiednio zadanymi pytaniami albo promptami można czasem wydobyć z nich informacje o konkretnej osobie, ale nie daje osobnej procedury ani progów oceny dla takich modeli i zbiorów. Ogólne metody badania wnioskowania trzeba do nich dopasować samemu.

Projekt nie jest ostateczny, konsultacje trwają do 30 października 2026 r. Jeśli któraś z tych luk dotyczy Waszych procesów, to dobry moment na zgłoszenie uwag.

Jakub Brzozowski – młodszy specjalista ds. ochrony danych osobowych w iSecure Sp. z o.o.