2019

PubWeb - uwarunkowania prawne, zagadnienia bezpieczeństwa i niezawodności

GRUPA

DUBIŃSKI JAN
GRĘBOWSKI ŁUKASZ
KALETA JOANNA
OGONOWSKI ALEKSANDER
ONISZCZUK MARIA
SZYMCZYK KORNEL
WALKOWIAK PAWEŁ

Spis treści

1. Analiza zgodności z RODO

RODO jest skrótem od Rozporządzenia o Ochronie Danych Osobowych (ang. *General Data Protection Regulation* - GDPR). Jest to 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). Weszło ono w życie 25 maja 2018 r.

1.1. Definicje wykorzystywanych pojęć

Podczas analizy zgodności systemu z RODO konieczna jest znajomość wykorzystywanych w ustawie pojęć. Jest ona niezbędna do poprawnego zrozumienia wymagań stawianych systemowi wynikających z tego rozporządzenia. W tym celu

„dane osobowe” oznaczają 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

„***przetwarzanie***” oznacza operację lub zestaw operacji wykonywanych na danych osobowych lub zestawach danych osobowych w sposób zautomatyzowany lub niezautomatyzowany, taką jak zbieranie, utrwalanie, organizowanie, porządkowanie, przechowywanie, adaptowanie lub modyfikowanie, pobieranie, przeglądanie, wykorzystywanie, ujawnianie poprzez przesłanie, rozpowszechnianie lub innego rodzaju udostępnianie, dopasowywanie lub łączenie, ograniczanie, usuwanie lub niszczenie

„ograniczenie przetwarzania” oznacza oznaczenie przechowywanych danych osobowych w celu ograniczenia ich przyszłego przetwarzania

„profilowanie” oznacza dowolną formę zautomatyzowanego przetwarzania danych osobowych, które polega na wykorzystaniu danych osobowych do oceny niektórych czynników osobowych osoby fizycznej, w szczególności do analizy lub prognozy aspektów dotyczących efektów pracy tej osoby fizycznej, jej sytuacji ekonomicznej, zdrowia, osobistych preferencji, zainteresowań, wiarygodności, zachowania, lokalizacji lub przemieszczania się;

„pseudonimizacja” oznacza przetworzenie danych osobowych w taki sposób, by nie można ich było już przypisać konkretnej osobie, której dane dotyczą, bez użycia dodatkowych informacji, pod warunkiem że takie dodatkowe informacje są przechowywane osobno i są objęte środkami technicznymi i organizacyjnymi

uniemożliwiającymi ich przypisanie zidentyfikowanej lub możliwej do zidentyfikowania osobie fizycznej

„zgoda” osoby, której dane dotyczą oznacza dobrowolne, konkretne, świadome i jednoznaczne okazanie woli, którym osoba, której dane dotyczą, w formie oświadczenia lub wyraźnego działania potwierdzającego, przyzwala na przetwarzanie dotyczących jej danych osobowych

„naruszenie ochrony danych osobowych” oznacza naruszenie bezpieczeństwa prowadzące do przypadkowego lub niezgodnego z prawem zniszczenia, utracenia, zmodyfikowania, nieuprawnionego ujawnienia lub nieuprawnionego dostępu do danych osobowych przesyłanych, przechowywanych lub w inny sposób przetwarzanych

1.2. Prezentacja wymagań i rozwiązań

W tej części zawarto kolejny wymagania stawiane przed systemem, które wynikają z obowiązywania RODO oraz przyjęte rozwiązania odnoszące się do tych wymagań.

Wymaganie 1: Privacy by design

Zgodnie z 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 zagrożenia 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.

Wymaganie 2: Privacy by default

Zgodnie z Art. 25 ust.2:

Administrator wdraża odpowiednie środki techniczne i organizacyjne, aby domyślnie przetwarzane były wyłącznie te dane osobowe, które są niezbędne dla osiągnięcia każdego konkretnego celu przetwarzania. Obowiązek ten odnosi się do ilości zbieranych danych osobowych, zakresu ich przetwarzania, okresu ich przechowywania oraz ich dostępności. W szczególności środki te zapewniają, by domyślnie dane osobowe nie były udostępniane bez interwencji danej osoby nieokreślonej liczbie osób fizycznych

Realizacja 1+2: System nie zbiera na temat użytkowników danych nie mających zastosowania przy jego działaniu. Zminimalizowana jest ilość danych wymaganych przy tworzeniu konta użytkownika w serwisie. Podanie dodatkowe informacje

dotyczących użytkownika jest możliwe poprzez edycje profilu użytkownika i jest całkowicie dobrowolne. Dane użytkownika wykorzystywane są przez system rekomendacji, który przedstawia użytkownikowi spersonalizowane propozycje pubów do odpowiedzenia, jeżeli użytkownik wyrazi na to zgodę. 0 W bazie danych zawierającej dane osobowe użytkowników zastosowany jest mechanizm pseudonimizacji.

Wymaganie 3: Brak możliwości przetwarzania, zbierania i przechowywania danych osobowych bez jednej z podstawy wskazanych w przepisach

Zgodnie z Art. 6. Podstawy do przetwarzania danych to:

- umowa - przetwarzanie jest niezbędne do wykonania umowy
- zgoda wyrażona przez osobę, której dane dotyczą
- wypełnienie obowiązku prawnego
- realizacja prawnie uzasadnionych interesów administratora danych

Realizacja 3: Podstawą przetwarzania danych w systemie jest zgoda użytkownika. System przetwarza dane osobowe użytkownika jedynie w wypadku wyrażenia zgody przez użytkownika. Zgoda ta uzyskana jest w sposób prawidłowy, co opisuje *Wymaganie 4*.

Wymaganie 4: Pozyskanie zgody od użytkownika w sposób prawidłowy

Zgodnie z Art.7:

1. Jeżeli przetwarzanie odbywa się na podstawie zgody, administrator musi być w stanie wykazać, że osoba, której dane dotyczą, wyraziła zgodę na przetwarzanie swoich danych osobowych.

2. Jeżeli osoba, której dane dotyczą, wyrażą zgodę w pisemnym oświadczeniu, które dotyczy także innych kwestii, zapytanie o zgodę musi zostać przedstawione w sposób pozwalający wyraźnie odróżnić je od pozostałych kwestii, w zrozumiałej i łatwo dostępnej formie, jasnym i prostym językiem. Część takiego oświadczenia osoby, której dane dotyczą, stanowiąca naruszenie niniejszego rozporządzenia nie jest wiążąca.

3. Osoba, której dane dotyczą, ma prawo w dowolnym momencie wycofać zgodę. Wycofanie zgody nie wpływa na zgodność z prawem przetwarzania, którego dokonano na podstawie zgody przed jej wycofaniem. Osoba, której dane dotyczą, jest o tym informowana, zanim wyrazi zgodę. Wycofanie zgody musi być równie łatwe jak jej wyrażenie.

4. Oceniając, czy zgodę wyrażono dobrowolnie, w jak największym stopniu uwzględnia się, czy między innymi od zgody na przetwarzanie danych nie jest uzależnione wykonanie umowy, w tym świadczenie usługi, jeśli przetwarzanie danych osobowych nie jest niezbędne do wykonania tej umowy

Oznacz to, że:

• zgoda pozyskana musi być zawsze przed rozpoczęciem przetwarzania danych
• sposoby wyrażania zgody: pisemna, ustna; klarowne, potwierdzające czynność działanie
• automatyczne zaznaczanie zgody jest zabronione
• zgoda musi dotyczyć przetwarzania konkretnych danych, w konkretnym czasie, w konkretnym celu
• obowiązuje zakaz łączenia zgód na przetwarzanie danych
• zgoda musi być wyrażona bez przymusu
• zgoda musi być poparta rzetelną klauzulą informacyjną
• nie jest wyrażeniem zgody: milczenie, okienka domyślnie zaznaczone, niepodjęcie działania itp.

Rozwiązanie 4: Zgoda użytkownika na przetwarzanie jego danych osobowych uzyskiwana jest podczas jego rejestracji w systemie. System wyświetla użytkownikowi zapytanie o zgodę na wykorzystanie jego danych osobowych. Zapytanie to spełnia wymienione powyżej warunki. Rezygnacja ze zgodny przebiega w prosty sposób i jest możliwa poprzez jej odznaczenie w profilu użytkownika po zalogowaniu.

Wymaganie 5: Prawo użytkownika do przeniesienia danych

RODO wprowadza prawo osoby, której dane dotyczą do przeniesienia jej danych osobowych do innego administratora. Osoba, której dane są przetwarzane może poprosić o przekazanie danych wskazanemu przez nią podmiotowi. Dane należy przekazać w ustrukturyzowanym, powszechnie używanym formacie nadającym się do odczytu maszynowego (np. XML czy CSV).

Rozwiązanie 5: W aplikacji webowej podane są dane kontaktowe do administracji serwisu. Prośbę o przeniesienie danych do innego administratora należy wysłać na adres mailowy wskazany w danych kontaktowych. W treści wiadomości należy umieści token (kod), który uwierzytelni użytkownika. Token ten można wygenerować w profilu użytkownika po zalogowaniu. Mechanizm ten służy zwiększeniu bezpieczeństwa.

Wymaganie 6: Prawo użytkownika do bycia zapomnianym

Prawo do bycia zapomnianym to prawo do usunięcia danych. Osoba, której dotyczą dane ma prawo zażądać od administratora danych usunięcia dotyczących jej danych, a żądanie to musi być spełnione jeśli nie występują przesłanki, które to uniemożliwiają.

Bez wątpienia należy usunąć dane m.in. jeśli:

- nie są już one niezbędne do celów, w których zostały zebrane
- zgoda na podstawie, której przetwarzałeś dane została cofnięta
- dane były przetwarzane niezgodnie z prawem
- dane osobowe zostały zebrane w związku z oferowaniem usług społeczeństwa informacyjnego (np. w ramach portalu społecznościowego)

Są jednak sytuacje, w których możliwa jest odmowa spełnienia żądania usunięcia danych. Ma to miejsce m.in. jeśli przetwarzanie tych danych jest niezbędne:

- do wywiązania się z prawnego obowiązku
- do ustalenia, dochodzenia lub obrony roszczeń
- do korzystania z prawa do wolności wypowiedzi i informacji

Rozwiązanie 6: W aplikacji webowej podane są dane kontaktowe do administracji serwisu. Prośbę o usunięcie danych należy wysłać na adres mailowy wskazany w danych kontaktowych. W treści wiadomości należy umieści token (kod), który uwierzytelni użytkownika. Token ten można wygenerować w profilu użytkownika po zalogowaniu. Mechanizm ten służy zwiększeniu bezpieczeństwa.

Wymaganie 7: Prawo użytkownika do żądania ograniczenia przetwarzania danych

Wymaganie 8: Prawo użytkownika do otrzymania kopii danych

Rozwiązanie 8: Wszelkie dane o użytkowniku przechowywane przez aplikacje dostępne są z poziomu interfejsów aplikacji (wyłącznie dla użytkownika). Ponadto istnieje możliwość złożenia żądania wydania kopii danych przez formularz kontaktu z administracją serwisu. Kopia danych przesyłana jest na adres mailowy użytkownika.

Wymaganie 9: Prawo użytkownika do sprzeciwu wobec profilowania

Zgodnie z art. 4 pkt 4 RODO, profilowanie to:

"dowolna forma zautomatyzowanego przetwarzania danych osobowych, które polega na wykorzystaniu danych osobowych do oceny niektórych czynników osobowych osoby fizycznej, w szczególności do analizy lub prognozy aspektów dotyczących efektów pracy tej osoby fizycznej, jej sytuacji ekonomicznej, zdrowia, osobistych preferencji, zainteresowań, wiarygodności, zachowania, lokalizacji lub przemieszczania się".

Rozwiązanie 9: Aplikacja nie wyświetla treści personalizowanych w oparciu o dane osobowe (nie uwzględnia miasta pochodzenia). Aplikacja w wersji obecnej nie wyświetla reklam, ani sugestii. Profilowanie użytkownika nie zachodzi.

Wymaganie 10: Przejrzystość języka

Zabronione jest używanie trudnych, specjalistycznych sformułowań, tak samo jak utrudnianie zapoznania się z ważnymi informacjami, np. stosowania w komunikatach tzw. metody „drobnego druczku".

Rozwiązanie 10: Tekst zgód napisany jest w przystępnej i zrozumiałej formie. Przejrzystość działania serwisu jest częścią jednego z wymagań niefunkcjonalnych serwisu = „zapewnienie możliwie najlepszego User Experience.

2. Zagadnienia bezpieczeństwa

2.1. Dostęp do danych

Dostęp do danych znajdujących się w bazach danych systemu powinni otrzymać użytkownicy do nich uprawnieni, w związku z czym, w systemie jest stosowana identyfikacja użytkowników. Użytkownicy mogą uzyskać dostęp do swoich danych osobowych tylko i wyłącznie poprzez uwierzytelnienie w serwisie (login i hasło).

Aby zminimalizować ryzyko naruszenia ochrony danych osobowych dostęp do nich zostanie przydzielony jak najmniejszej liczbie członków zespołu. Wszystkie próby odczytu/modyfikacji danych wrażliwych muszą być jasno udokumentowane. Bazy danych zostaną umieszczone w kontenerach, więc port usługi bazy danych nie jest dostępny z sieci publicznej, co zapewnia nam większe bezpieczeństwo.

Usługi internetowe RESTful muszą być zabezpieczone przed wyciekami danych uwierzytelniających. Hasła, tokeny bezpieczeństwa i klucze API nie powinny pojawiać się w adresie URL, ponieważ może to prowadzić do wycieku poufnych informacji.

2.2. JWT

JWT – Json Web Token jest otwartym standardem (RFC 7519) tworzenia tokenów dostępowych, które zawierają w sobie pewną porcję danych (payload). Siłą tokenów jest to, że mogą przenosić dowolną porcję danych i informacji o użytkowniku. Token składa się z trzech części:

Nagłówka – określa jaki algorytm szyfrowania został wykorzystany do utworzenia podpisu tokena.

Payloadu – tu znajdują się dekodowane w base64 uprawnienia/informacje przenoszone w tokenie.

Podpisu cyfrowego – podpis payloadu token-a. Podpisany kluczem prywatnym serwera (dostawcy token-a). Zabezpiecza token przed modyfikacją.

Należy zaznaczyć iż domyślnie tokeny JWT nie posiadają funkcji pozwalających na wykluczenie przechwycenia tokena. Często w tokenach umieszcza się takie informacje jak ip, rodzaj przeglądarki (user-agent). Tokeny mają różne czasy życia, w przypadku aplikacji o wysokich wymaganiach bezpieczeństwa tokeny z pozwoleniami np. przelewu mogą być jednorazowe.

Uwaga! Użycie tokenów nie eliminuje konieczności użycia rozwiązań pozwalających zachować poufność przesyłanych informacji. Nie należy rozumieć tokenów JWT jako rozwiązań tej klasy co np. SSL.

Token przesyłany jest w nagłówku „*Authorization*” zapytania HTTP.

Rolą JWT w aplikacji Pubweb jest wykluczenie konieczności każdorazowego logowania loginem i hasłem oraz możliwość przekazywania informacji pomiędzy serwisami takich jak role i przywileje w spójnej formie. Użycie tokenów JWT pozwala również na ograniczenie konieczności odpytywania punktu autoryzacyjnego o dane użytkownika za każdym razem.

Standardowy schemat autoryzacji przy pomocy tokena JWT wygląda następująco:

1. Użytkownik wysyła hasło i login do serwera autoryzacyjnego
2. Serwer autoryzacyjny zwraca token JWT wraz z informacjami o tym jakie uprawnienia posiada użytkownik w aplikacji, do podpisu JWT użyty jest klucz prywatny serwera autoryzacyjnego
3. Użytkownik używa tokenu wygenerowanego przez centrum autoryzacji do wykonania operacji, w którymś z mikroserwisów.
4. Mikroserwis udziela lub nie udziela zgody na wykonanie operacji na podstawie:
    - 4.1. Sprawdzenia za pomocą klucza publicznego serwera autoryzacji autentyczności JWT (porównanie podpisów).
    - 4.2. W przypadku, gdy weryfikacja z 4.1. przebiegła pomyślnie serwis sprawdza czy w JWT zawarte są uprawnienia (np. Role/przywileje) wymagane do wykonania danej operacji
    - 4.3. Mikroserwis wykonuje inne walidacje biznesowe charakterystyczne dla danej operacji

Powyższy schemat ma poniższe implikacje:

- Jeżeli dojdzie do rozszerzenia uprawnień użytkownika to po stronie użytkownika leży wygenerowanie i używanie nowego tokena o adekwatnych uprawnieniach

- Jeżeli dojdzie do ograniczenia uprawnień użytkownika to z punktu widzenia autoryzacji za pomocą JWT wejdzie ono w życie dopiero po wygaśnięciu poprzedniego tokena. Dla operacji o wysokim stopniu bezpieczeństwa potrzebny jest dodatkowy mechanizm autoryzacji (np. każdorazowe odpytywanie serwera autoryzacyjnego o uprawnienia użytkownika, którego dotyczy JWT).

2.3. Przesyłanie danych

Komunikacja z serwerem odbywa się poprzez zastosowanie protokołu HTTPS. HTTPS zapewnia szyfrowanie asymetrycznie między klientem a serwerem, co pozwoli to na uzyskanie żądanej poufności oraz integralności transmisji danych.

Bezpieczne usługi REST powinny jedynie udostępniać punkty końcowe HTTPS. Pozwala to na ochronę poświadczeń uwierzytelniających podczas przesyłania, takich

jak na przykład hasła, klucze API lub tokeny sieciowe JSON. Umożliwia to również klientom uwierzytelnianie usług oraz gwarantuje integralność przesyłanych danych.

Aby chronić przesyłane dane, należy stosować tzw. “dobre praktyki” TLS/SSL, takie jak zweryfikowane certyfikaty, odpowiednio chronione klucze prywatne, stosowanie wyłącznie bezpiecznych szyfrów. Dane prywatne muszą być szyfrowane w pamięci przy użyciu kluczy o odpowiedniej długości i w ściśle określonych warunkach dostępu, zarówno technicznych, jak i proceduralnych. Dane uwierzytelniające użytkownika muszą być hashowane niezależnie od tego, czy są one szyfrowane czy nie.

2.4. Dostęp do serwerów

Dostęp do serwerów jest możliwy poprzez protokół komunikacyjny SSH. Zalogowanie do SSH jest możliwe tylko poprzez zastosowanie SSH Key. Logowanie przy pomocy hasła może zostać złamane atakiem brute force, a rozszyfrowanie SSH Key jest aktualnie niemal niemożliwe.

Całe oprogramowanie serwerów jest aktualizowane na bieżąco w celu zmniejszenia liczby podatności.

Serwery są odpowiednio skonfigurowane pod względem bezpieczeństwa poprzez zastosowanie zapór sieciowych (firewalls).

2.5. Przechowywanie haseł

Hasła w systemie przechowywane są tylko i wyłącznie w postaci ich skrótów wykorzystując bezpieczną funkcję skrótu.

Informacje niejawne dotyczące użytkowników, takie jak hasła, muszą być również chronione przy użyciu silnych, odpornych na kolizje funkcji skrótów, w celu znacznego ograniczenia ryzyka ujawnienia danych uwierzytelniających, jak również zapewnienia właściwej kontroli integralności.

2.6. Kodowanie wyjść

Należy zapewnić, poprzez usługi internetowe, takie kodowanie informacji wysyłanych do klientów aby mogły być one wykorzystywane jako dane, a nie jako skrypty. Jest to szczególnie istotne, gdy klienci usług internetowych wykorzystują dane wyjściowe do renderowania stron HTML bezpośrednio lub pośrednio przy użyciu obiektów AJAX.

2.7. Rozmiar oraz przepustowość komunikatów

Aplikacje internetowe, mogą być obiektem ataków DOS wykonywanych poprzez automatyczne wysyłanie do serwisów internetowych tysięcy obszernych komunikatów

z protokołu SOAP. Może to doprowadzić do sparaliżowania atakowanej aplikacji, uniemożliwiając jej odpowiadanie na prawdziwe wiadomości. Aby zapobiec takiej sytuacji, rozmiar komunikatów SOAP powinien być ograniczony. Większy rozmiar (lub brak limitu w ogóle) zwiększa szanse na udany atak DoS.

Przepustowość komunikatów oznacza liczbę zapytań serwisów internetowych obsługiwanych w zadanym czasie. Należy dokonać konfiguracji zoptymalizowanej pod kątem maksymalnej przepustowości wiadomości, co pozwoli na uniknięcie sytuacji podobnych do DoS.

2.8. Prepared query

W celu zapobiegania atakowi SQL injection zostanie wykorzystany mechanizm prepared query opierający się na wysyłaniu zapytania oraz danych do serwera bazy danych oddzielnie.

3. Zagadnienia niezawodności

3.1. Odporność na awarie

Systemy rozproszone charakteryzują się wyższą odpornością na uszkodzenia, ponieważ zasoby mogą być zwielokrotnione oraz awarie zazwyczaj są częściowe. Awarie częściowe nie powodują zatrzymania całego systemu tylko pojedynczej składowej. Odporność na awarie zostanie zapewniona przez m.in. automatyczne restartowanie kontenerów Docker. Funkcja autorestart zapewnia automatyczne resetowanie kontenerów w przypadku ich zatrzymania bądź awarii. Zapewni nam to przywrócenie stanu części systemu sprzed awarii.

3.2. Replikacja danych

Awaria niektórych systemów może wpływać na dostępność danych. Oprócz zastosowania auto restartowania kontenerów zastosowana również zostanie replikacja danych, czyli przechowywanie danych w kilku kopiach. Replikacja danych jest realizowana poprzez kopiowanie i przesyłanie danych między serwerami, i ich synchronizacji w celu zapewnienia spójności. Pozwala na skrócenie czasu dostępu do danych, oraz uniezależnienie się od czasowej niedostępności serwerów. Wadą jest konieczność aktualizowania repliki w przypadku zmiany danych źródłowych.

W systemie zastosowana jest baza danych PostgreSQL, która oferuję łatwą konfigurację replikacji danych. PostgreSQL zapewnia mechanizmy niezawodności opisane w *https://www.postgresql.org/docs/9.3/wal-reliability.html*.

3.3. Usterki

Niezawodny system powinien posiadać mechanizm diagnozowania usterek i podejmowania odpowiednich działań. Działania te będą obejmować maskowanie usterek za pomocą całkowitego ukrycia i kontynuowania dalszej pracy; chwilowe lub całkowite przerwanie funkcjonowania w przypadku, gdy maskowanie jest niemożliwe do wykonania. Całkowite przerwanie jest wykonywane w ściśle określony sposób pozwalający na uniknięcie utraty lub spójności danych.