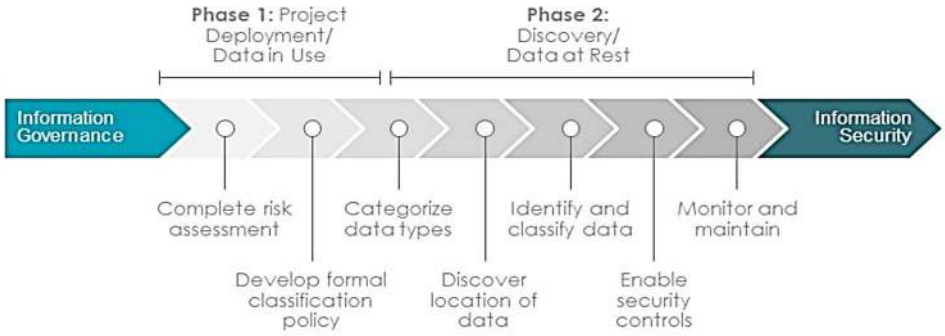
7 kroków do efektywnej klasyfikacji danych

Andrzej *Polubiec*
Dyrektor Techniczny
TUKAN IT

Czy droga do efektywnej klasyfikacji danych musi być procesem nieuporządkowanym i skomplikowanym? Może warto posłużyć się schematem, który podpowie, od czego ten proces zacząć i jakie kolejne kroki należałoby podjąć? Na bardzo ciekawy koncept natrafiłem jakiś czas temu na stronach Forsythe, gdzie w artykule – napisanym w formie poradnika – pod tytułem „7 Steps to Effective Data Classification" zaprezentowano jak krok po kroku wdrożyć klasyfikację danych w organizacji w powiązaniu z innymi systemami bezpieczeństwa informacji.

Propozycja Thomas Eck, Anne Grahn, autorów artykułu, prezentuje się następująco:



Jestem współtwórcą produktu GREENmod – systemu do klasyfikacji dokumentów oraz poczty elektronicznej – i od paru lat uczestniczę w jego wdrożeniach, konfiguracji i personalizacji. Korzystając ze swojego doświadczenia chciałbym przybliżyć opisaną w artykule metodologię i podzielić się swoimi spostrzeżeniami.

Artykuł powołuje się na raport o cyberbezpieczeństwie – „2017 State of Cybersecurity Metrics Annual Report" firmy Thycotic, z którego wynika, że 80% firm nie wykrywa i nie śledzi wrażliwych danych oraz nie jest też w stanie określić, gdzie ich dane krytyczne są duplikowane lub przenoszone w ramach ich własnej sieci.

Podstawowym problem jest ogromna ilość przechowywanych danych, z których część jest zduplikowana, zdezaktualizowana, niepotrzebna, m.in. dlatego, iż żadne regulacje nie wymagają ich przechowywania. Wrażliwe informacje przechowywane cyfrowo, zawierające własność intelektualną lub przemysłową, dane osobowe pracowników lub klientów (takie jak: numery identyfikacyjne, informacje o stanie zdrowia, informacje finansowe, czy też szczegóły kart kredytowych), muszą być odpowiednio zabezpieczone. Dla większości firm identyfikacja oraz wyszukanie tych ważnych danych stanowi nie lada wyzwanie.

W celu dostosowania się do regulacji dotyczących ochrony i prywatności, takich jak **Rozporządzenie o ochronie danych osobowych**, **RODO** (ang. *General Data Protection Regulation*, GDPR), niezbędna jest analiza przechowywanych informacji, polegająca na identyfikacji, gdzie znajdują się wrażliwe dane, na opracowaniu zasad postępowania z nimi oraz wdrożeniu odpowiednich mechanizmów kontrolnych. Dodatkowo z mojego doświadczenia wynika, że kluczową rolę odgrywa edukacja użytkowników – w zakresie zagrożeń dotyczących danych oraz najlepszych praktyk w celu zachowania bezpieczeństwa – mająca na celu zbudowanie świadomości wagi przetwarzanych przez nich informacji.

Ochrona danych wrażliwych nie jest łatwym zadaniem, gdyż każda organizacja ma swoją specyfikę działania i nie ma jednej dopasowanej dla wszystkich strategii. Kluczem do sukcesu jest klasyfikacja danych i wspomniane wcześniej na stronach Forsythe 7 kroków do osiągnięcia tego celu.
Uwidocznione na powyższym schemacie przejście od zarządzania informacją do bezpieczeństwa informacji w organizacji zostało podzielone na dwie główne fazy:

Faza I. Wdrożenie projektu – dotyczy danych w spoczynku, czyli przechowywanych w magazynach elektronicznych.

Faza II. Wykrywanie – dotyczy danych w ruchu lub w użyciu, czyli danych transmitowanych lub przetwarzanych.
Z pewnym uproszeniem można stwierdzić, że klasyfikujemy dane w spoczynku, aby zapewnić im odpowiednią ochronę, zanim zaczną być danymi w ruchu.

Poniżej przedstawię siedem podstawowych kroków, wykonując które – wg autorów artykułu – będziemy mogli skutecznie klasyfikować dane:

1. **Wykonaj analizę ryzyka pracy na danych wrażliwych**.

Cele stosowania klasyfikacji w firmie muszą być jasne i zrozumiałe. Dodatkowo należy zapewnić zgodność z normami wewnętrznymi i ustawowymi, dotyczącymi ochrony prywatności i poufności danych. Można to osiągnąć poprzez wywiady obejmujące kluczowe podmioty, mające wpływ na projekty lub uczestniczące w ich realizacji, np. liderów działów prawnych, biznesowych oraz IT. Uzyskana w ten sposób wiedza umożliwi szacowanie ryzyka związanego z przetwarzaniem danych wrażliwych w istniejącym systemie zarzadzania bezpieczeństwem informacji pod kątem opracowania polityki klasyfikacji.

2. **Opracuj sformalizowaną politykę klasyfikacji.**
Nadmiernie rozbudowane schematy, polegające na formułowaniu licznych poziomów klasyfikacji, powodują zamieszanie. Aby proces wdrażania systemu wspomagającego klasyfikację danych był skuteczny, system powinien być przyjazny dla użytkownika. Zalecane jest, aby ilość kategorii, którą posługują się pracownicy, była ograniczona.
Zalecenia skutecznej klasyfikacji:

PUBLICZNE – dane, które mogą być swobodnie ujawniane publicznie - np. materiały marketingowe, informacje kontaktowe, cenniki itp.
WEWNĘTRZNE – dane wewnętrzne, nieprzeznaczone do ujawnienia publicznego - np. informacje o konkurencyjności rynkowej (analiza SWOT), strategie biznesowe, kampanie sprzedażowe, kampanie marketingowe, schemat organizacji itp.
POUFNE – wrażliwe dane, których ujawnienie może mieć negatywny wpływ na działanie i wizerunek firmy - np. umowy ze sprzedawcami, oceny pracowników, lista płac itp.
ZASTRZEŻONE – wyjątkowo ważne dokumenty, takie jak dane klientów, których ujawnienie może powodować skutki prawne i/lub finansowe - np. własność intelektualna, dane kart płatniczych, dane osobowe.

Opisy kategorii powinny zawierać przykłady, jakiego rodzaju danych dotyczą. Dobrą praktyką jest także opisanie wytycznych dotyczących postępowania z danymi oraz potencjalnego ryzyka związanego, ze świadomym lub nie, udostępnieniem danych.
Z mojego doświadczenia wynika, iż wielu klientów stosuje podkategorie, jednak tego typu zabieg zalecany jest tylko dla kategorii najbardziej restrykcyjnych. Autorzy artykułu jeszcze bardziej zawężają kategorie podrzędne, sugerując, iż powinny one być związane z wymaganiami narzucanymi przez regulacje lub wręcz opisywać kontrolę dostępu.

Przykłady takich podkategorii:
- Dane osobowe (GDPR).
- Dane medyczne.
- Dane finansowe.
- Dostęp tylko IT.

3. **Rozpoznaj wrażliwe dane.**
Należy określić, z jakimi rodzajami wrażliwych danych mamy do czynienia w naszej organizacji. W tym celu należy się skupić na analizie procesów biznesowych, o których najwięcej informacji można uzyskać od ich właścicieli. Rozpoznając poszczególne procesy i śledząc przepływ danych uzyskujemy pogląd na to, jakie dane należy chronić i w jaki sposób to robić.

Pomocnym będzie udzielenie sobie odpowiedzi na kilka pytań:
- Jakie dane o klientach i partnerach zbiera twoja organizacja?
- Jakie dane na ich temat są wytwarzane?
- Jakie zastrzeżone według regulacji dane tworzysz?
- Jak zamierzasz podejść do danych transakcyjnych?
- Które ze zbieranych i wytwarzanych danych są poufne?

4. **Ustal lokalizację danych**.
Po ustaleniu rodzajów wrażliwych danych w organizacji ważne jest, aby skatalogować wszystkie miejsca, w których dane są przechowywane w sposób cyfrowy. Kluczowe znaczenie ma także przepływ danych do i poza organizację. Dlatego należy dokonać szczegółowej analizy tego, w jaki sposób twoja organizacja przechowuje i udostępnia dane wewnętrznie i zewnętrznie oraz czy korzysta z usług w chmurze.

5. **Identyfikuj i klasyfikuj dane.**
Tylko wtedy, gdy mamy wiedzę o miejscu przechowywania wrażliwych danych, możemy je zidentyfikować, a następnie sklasyfikować w celu zapewnienia odpowiedniej ochrony.
Stworzenie podstawowych środków bezpieczeństwa cybernetycznego i zdefiniowanie kontroli opartej na zasadach klasyfikacji danych wymaga zapewnienia odpowiednich rozwiązań.
Automatyczna klasyfikacja z reguły jest częścią systemu DLP (ang. Data Leak Prevention).
Umożliwia ona m.in.:
- Klasyfikację wg lokalizacji (serwer plików, udostępnione dyski itp.).
- Klasyfikację wg zawartości (słowa kluczowe, wyrażenia regularne, wyszukiwanie zapamiętanych wzorców np. pięć numerów PESEL blisko siebie w dokumencie).
- Klasyfikację wg typu plików (użycie możliwe, jeśli posiadamy system, który generuje specyficzny typ danych, który powinien być chroniony).
- Klasyfikację wg unikalnego identyfikatora (podpis cyfrowy PKI, hash, inny znacznik).

Reguły automatycznej klasyfikacji muszą być wcześniej zaprojektowane i zaimplementowane w używanym systemie DLP, ale nadają się głównie do danych ustrukturyzowanych i mieszczą się w schemacie, który jesteśmy w stanie nimi opisać.
Jednak co z dokumentami nieustrukturyzowanymi lub takimi, które nie posiadają pożądanych cech (takich jak PESEL, numery kart kredytowych), a są dla nas ważne i nie powinny opuszczać firmy? W tym momencie nieodzowna okazuje się klasyfikacja danych przez użytkownika, czyli przez autora dokumentu, posiadającego wiedzę o jego treści i potrafiącego przydzielić mu odpowiednią kategorię.
W wyniku analizy, dokumentom w każdej z wymienionych metod nadawana jest określona etykieta, dopisywana do metadanych. Dzięki uzyskanej w ten sposób świadomości wagi przetwarzanych danych, możliwe jest pełne wykorzystanie funkcjonalności systemów informatycznych do zapewnienia odpowiedniego poziomu ochrony.

6. **Włącz sterowanie.**
Środki cyberbezpieczeństwa, poprzez polityki nimi sterujące, powinny być odpowiednio powiązane z każdą etykietą klasyfikacyjną, zapewniając właściwe ich użycie. Dane o wysokim ryzyku wymagają większego poziomu ochrony, podczas gdy dane o niskim ryzyku takiego poziomu zabezpieczeń nie potrzebują. Dzięki zrozumieniu, gdzie znajdują się dane wrażliwe i jaka jest ich wartości dla organizacji, można wdrożyć odpowiednie mechanizmy bezpieczeństwa w oparciu o powiązane ryzyko. Metadane dokumentu zawierające klasyfikację mogą być używane przez systemy – zapobiegające utracie danych (DLP), szyfrujące oraz inne systemy bezpieczeństwa – do podjęcia decyzji, które informacje i w jaki sposób powinny być chronione.

7. **Monitoruj i zarządzaj.**
Ciągłe monitorowanie i przystosowywanie się do zmian warunków, zarówno na rynku jaki w organizacji, pozwoli utrzymać wciąż aktualny system do klasyfikacji i w razie potrzeby dokonać w nim poprawek. Proces przeglądania i aktualizacji polityk klasyfikacji w firmie, który angażuje użytkowników, jest bardzo ważny i pozwoli polityce klasyfikacyjnej nadążać za zmianami biznesowymi jakie zachodzą w organizacji.

Od momentu utworzenia informacji do czasu jej zniszczenia, klasyfikacja danych pomaga organizacji w zapewnieniu skutecznej ochrony, przechowywania i zarządzania. Klasyfikacja danych pozwoli zmniejszyć ryzyko związane z danymi wrażliwymi, usprawnić proces decyzyjny i zwiększyć skuteczność systemów

zapobiegających wyciekowi danych (DLP), szyfrowania i innych zabezpieczeń.

Pamiętajmy, że tworząc jasne zasady klasyfikacji, dokładnie oceniając i lokalizując dane oraz wdrażając odpowiednie rozwiązania przeciwko wyciekom, organizacja może być pewna, że dane wrażliwe są odpowiednio zabezpieczone. Dzięki temu ryzyko utraty ważnych dla nas danych zmniejsza się. Równie ważni są zaangażowani w proces bezpieczeństwa pracownicy, którzy poprzez praktyczne zdobywanie wiedzy, dotyczącej zagadnień bezpieczeństwa oraz ważności przetwarzanych przez nich informacji, stają się bardziej świadomi. W efekcie spada liczba wycieków informacji, które wynikają z czynności wykonywanych bez zastanowienia lub z braku wiedzy na temat skutków udostępnienia chronionych treści osobom nieupoważnionym. Pracownicy, którzy są zaangażowani w proces zapewniania bezpieczeństwa, stają się też bardziej odpowiedzialni za informacje, które tworzą.

Podsumowując, przedstawione powyżej „7 kroków do efektywnej klasyfikacji danych" doskonale opisuje procedurę wdrożenia systemu klasyfikacji danych, dopięcia tego kolejnego ogniwa w łańcuchu systemów bezpieczeństwa informacji w organizacji. Ci, którzy już mają za sobą podobny proces na pewno znajdą analogię do swoich poczynań, natomiast ci, którzy dopiero planują wdrożenie klasyfikacji w swojej firmie, otrzymają gotową receptę na sukces swoich przyszłych poczynań.

Źródło:
https://focus.forsythe.com/articles/619/7-Steps-to-Effective-Data-Classification