

Ochrona danych osobowych w dobie Big Data

raport z konferencji 28.01.2016

Cyberlaw.pl



POMOCNIK
RODO



Big data to najprościej mówiąc pobieranie danych z różnych źródeł, porównywanie ich, analiza i wyciąganie wniosków. To proces przetwarzania, który pomaga budować profile - nie tylko o osobach. Big data wykorzystywane jest w inteligentnych miastach czy halach produkcyjnych gdzie badane są pewne tendencje albo zachowania osób, które nie są identyfikowane. Jednak nie zawsze. Niejednokrotnie ten rodzaj przetwarzania wiąże się z profilowaniem osób i podejmowaniem decyzji mających wpływ na konkretne osoby fizyczne. Dlatego big data wiąże się ze szczególnymi ryzykami.

Sam proces przetwarzania dużej ilości danych wykorzystywany jest zarówno w sektorze prywatnym jak i publicznym. O tym na co należy zwrócić uwagę przy okazji korzystania z big data rozmawiano w trakcie konferencji „Ochrona danych osobowych w dobie Big Data”, która odbyła się w ramach obchodów X dnia ochrony danych osobowych organizowanych przez Biuro Generalnego Inspektora Ochrony Danych Osobowych. Serdecznie zapraszam do zapoznania się z raportem z konferencji.

Jeśli interesuje Ciebie temat reformy przepisów ochrony danych zapraszam na stronę pomocnikrodo.pl

Beata Marek
cyberlaw.pl

Wydawca, grafika, skład i łamanie:

Kompendium informacji o reformie ochrony danych

Cyberlaw.pl



Z okazji Dnia Ochrony Danych Osobowych 28 stycznia 2016 r. w budynku Biblioteki Uniwersytetu Warszawskiego odbyła się konferencja organizowana przez Generalnego Inspektora Ochrony Danych Osobowych (GIODO) oraz Wydział Prawa i Administracji Uniwersytetu Warszawskiego dotycząca ochrony danych osobowych w dobie Big Data.

Dr Edyta Bielak-Jomaa, Generalny Inspektor Ochrony Danych Osobowych, otwierając konferencję podkreśliła, że w dobie big data, w związku z przetwarzaniem - przy użyciu nowoczesnych technologii - coraz większej ilości danych ważne jest, by przeanalizować wszystkie aspekty tego zjawiska, zarówno związane z nim korzyści społeczne gospodarcze, jak i zagrożenia dla prywatności. Często bowiem dotyczące nas dane pozyskiwane są z wielu źródeł, bez naszej wiedzy i zgody. Następnie dane te są zestawiane i analizowane w celu stworzenia naszego profilu. Niejednokrotnie wyciągnięte wnioski są fałszywe. Dlatego ważne jest przedyskutowanie, z jakimi zagrożeniami dla ochrony danych osobowych i prywatności wiąże się big data, oraz wypracowanie standardów postępowania w tym obszarze.

Dr Edyta Bielak-Jomaa przekazała również informacje o wynegocjowaniu w grudniu 2015 r. ostatecznego tekstu rozporządzenia unijnego o ochronie danych. Zapewniła o trwających w Urzędzie, prowadzonych pracach, mających na celu odpowiednie przygotowanie się do jego wejścia w życie.

Witold Kołodziejski, sekretarz stanu w Ministerstwie Cyfryzacji dodał, że rozmawianie o big data w kontekście ochrony danych, to temat na czasie choćby dlatego, że finalizują się duże inwestycje infrastrukturalne z poprzedniej perspektywy unijnej związane z centrami przetwarzania danych w Polsce. Na przykład Interdyscyplinarne Centrum Modelowania Matematycznego ma dzisiaj superkomputery ze ścisłej światowej czołówki. Także inne centra dysponują ogromnym potencjałem do gromadzenia i przetwarzania danych. Wszystko jest przygotowane do big data, a przetwarzanie danych ma ogromne znaczenie dla rozwoju nowoczesnej gospodarki i sprawnego funkcjonowania państwa. Łączenie i analizowanie dużej ilości danych czyli big data jest krwiobiegiem nowych usług i produktów. Ten model przetwarzania ma jedną cechę. Musi być zasilony informacją. Temu zasilaniu służą nowe technologie. W kontekście big data mówimy o terabajtach, docierających w ciągu sekundy do centrów obliczeniowych. Efektem zwiększonej zdolności do przetwarzania danych powinna być poprawa jakości naszego życia.

Zdaniem Ministra, już dziś trudno sobie wyobrazić funkcjonowanie gospodarki bez big data. Jednocześnie jednak im większa jest partycypacja społeczeństwa w świecie wirtualnym, tym na większe zagrożenie narażona jest jednostka. Dlatego tak ważna jest dzisiejsza dyskusja.

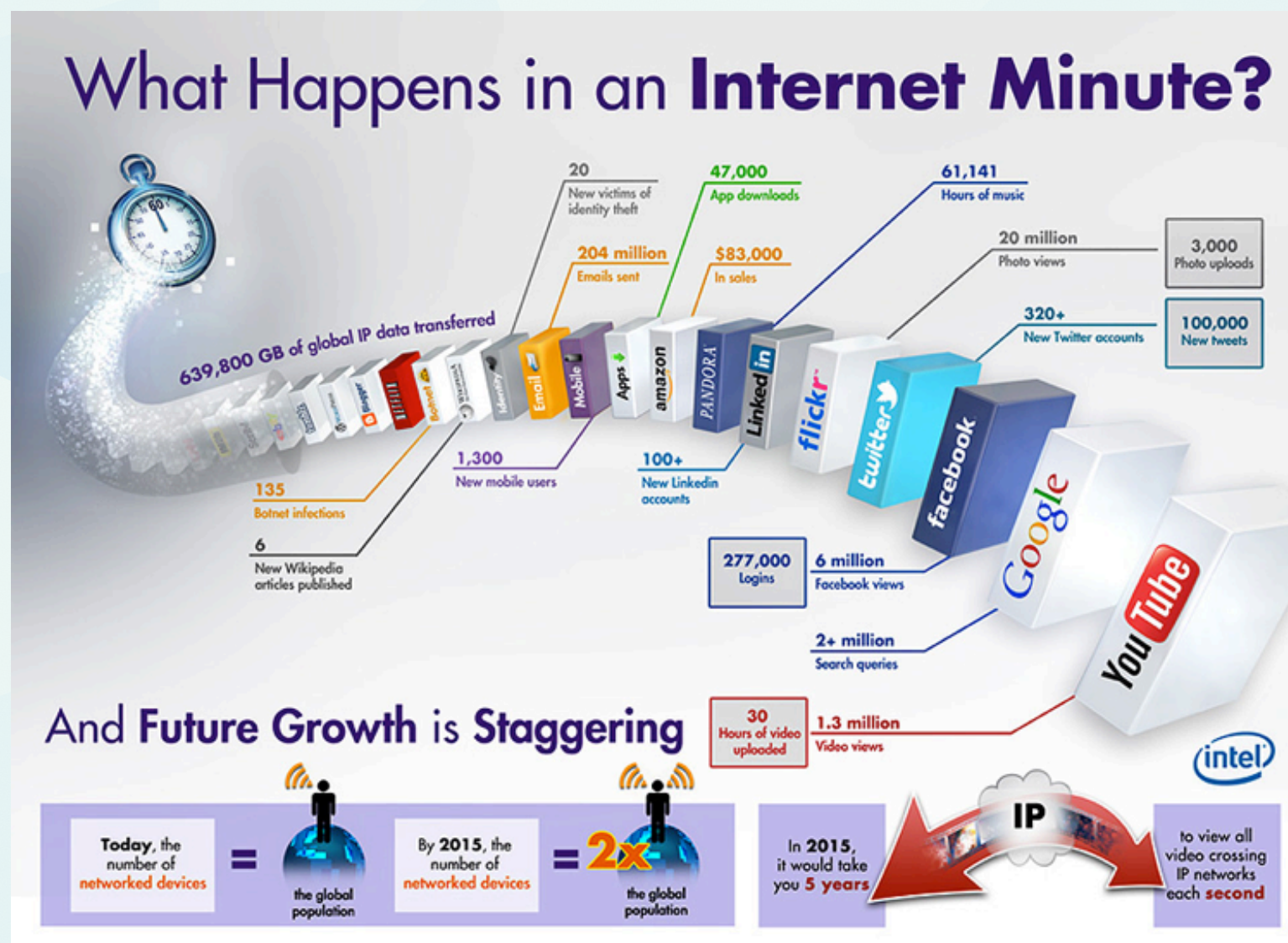
- Jestem przekonany, że świat technologii ICT niesie ogromne korzyści zarówno dla obywateli, jak i przedsiębiorców. Podstawą ich współpracy musi być jednak wzajemne zaufanie. Jako obywatele chcemy mieć pełną świadomość na co się zgadzamy, udostępniając informacje o sobie. O to chodzi w implementacji nowych przepisów. Likwidując bariery rozwojowe musimy jednocześnie dbać o zwiększenie zaufania do usług i usługodawców.

Gospodarzem procesu wdrożenia rozporządzenia jest Ministerstwo Cyfryzacji. Nowe przepisy będą obowiązywać za dwa lata, dlatego Ministerstwo pracuje nad tym, by do tego czasu dostosować polskie prawo. Nowe przepisy zakładają pewną dowolność dla państw członkowskich w zakresie określonych regulacji, np. dolnej granicy wieku dostępu do portali społecznościowych. W tych obszarach na pewno prowadzone będą szerokie konsultacje społeczne.

Sekretarz stanu dodał, że Ministerstwo Cyfryzacji jest gospodarzem systemu rejestrów publicznych. Koordynuje także procesy informatyczne, które toczą się w innych resortach – np. finansów czy zdrowia. Nawiązał także do problematyki przetwarzania wielkiej ilości danych sektora publicznego, włączając w to zasoby muzeów, bibliotek i archiwów udostępnionych powszechnie w wyniku wprowadzanej właśnie ustawy.

Po zakończonej prezentacji, w imieniu Krajowego Stowarzyszenia Ochrony Informacji Niejawnych oraz Stowarzyszenia Wspierania Bezpieczeństwa Narodowego i radia Top Secret, ich Prezes płk. Mieczysław Koczkowski wręczył GODO pamiątkową laudację z okazji X obchodów Dnia Ochrony Danych Osobowych oraz złożył podziękowania dla wszystkich pracowników GODO za wieloletnią współpracę na rzecz ochrony danych osobowych.

Sesję wprowadzającą rozpoczął wykład „Analiza Big Data: nowe możliwości oraz zagrożenia“, który poprowadził prof. Jerzy Stefanowski z Instytutu Informatyki na Politechnice Poznańskiej. Jak podkreślił, od kilkudziesięciu lat nastąpił gwałtowny przyrost gromadzonych elektronicznie danych. Jest to efektem zarówno rozwoju technologii informatycznych, w tym coraz powszechniejszego stosowania relacyjnych baz lub hurtowni danych, nowych metod rejestracji pomiarów, jak i pojawienia się nowych źródeł danych. Przyrost gromadzonych wolumenów danych jest tak ogromny, że obecnie mówimy o takim zjawisku, jak „powódź informacyjna“. W celu zilustrowania nowych źródeł generowania takich danych prelegent przedstawił infografikę pokazującą, co dzieje się w ciągu minuty w Internecie jako przejaw aktywności ludzkiej.



Zródło: Intel

Prelegent zwrócił szczególną uwagę na media społecznościowe, gdyż użytkownicy sami są źródłem milionów wpisów czy rejestracji swoich aktywności. Wiele z danych obecnych w Internecie jest publicznie dostępne albo jest relatywnie łatwe do automatycznego pozyskania poprzez różnego rodzaju oprogramowanie. W dużym stopniu to właśnie rozwój różnego rodzaju usług internetowych przyczynił się do pojawienia się zjawiska nazywanego z j. ang. Big Data. Ponadto można zauważyć rosnącą rolę tzw. sieci sensorów (różnego rodzaju systemów pomiarowych), urządzeń mobilnych lub tzw. Internetu rzeczy (z ang. Internet of Things) – zdaniem prelegenta w tym przypadku dane gromadzone automatycznie są już drugim poważnym źródłem powstawania nowego rodzaju zasobów informacyjnych, które mogą być przedmiotem eksploracji z wykorzystaniem algorytmów. Na koniec należy wspomnieć o zwiększających się rozmiarach danych przechowywanych w bardziej tradycyjnych ośrodkach obliczeniowych, np., w administracji, bankowości lub ubezpieczeniach.

- Zmiana technologii dotyczy także sposobu rejestracji danych. Jeszcze kilkadziesiąt lat temu dane wprowadzano głównie z wykorzystaniem klawiatury. Teraz już wprowadza się dane za pomocą kodów, pasków magnetycznych, sieci sensorów. Użytkownicy Internetu są sami źródłem danych. Ponadto w wielu przypadkach takie urządzenie jak smartphon'y lub inne czujniki monotowane standardowo w sprzęcie, który codziennie używamy, są źródłami bardzo precyzyjnych pomiarów, które są śladem Naszych codziennych aktywności.

Wszystkie te obserwacje prowadzą do stwierdzenia, że z punktu widzenia informatyków wyzwaniem nie jest zatem rejestrowanie i gromadzenie danych. Prawdziwym wyzwaniem jest jak analizować te dane, aby w ich powodzi znaleźć interesujące, nowe i użyteczne praktycznie wzorce – tzn. odkrywać reprezentacje wiedzy z danych.

Prelegent zdefiniował termin Big Data jako dane, których skala, zróżnicowanie i złożoność wymaga użycia nowych technologii i algorytmów w celu odkrycia wartościowej wiedzy. Równocześnie podkreślił, że wbrew popularnym opiniom Big Data nie powinno być wyłącznie rozumiane jako tzw. duże rozmiary danych oraz że tłumaczenie polskie tego terminu odnoszące się wyłącznie do wielkości nie jest właściwe. Problem przetwarzania tzw. masywnych danych (lub z j. ang. large scale analysis) istniał już wcześniej w informatyce. Profesor Stefanowski zwrócił uwagę, że natura nowych rodzajów danych wiąże się z ich innymi charakterystykami, które świadczą o ich złożoności i trudności do analizy. Jego zdaniem najlepiej specyfikę Big Data wyrażają definicje, wg. których Big Data są scharakteryzowane za pomocą wielu V: „High Volume (wielkość przetwarzanych danych), Variety (złożone, niejednorodne reprezentacje źródeł danych) Velocity (aspekt czasu przetwarzania, oraz problem zmienności danych wraz z upływem czasu), Veracity (gorsza jakość danych i pewność co do ich wartości niż w przypadku klasycznej analizy statystycznej).

Zdaniem profesora obecnie jesteśmy w takim miejscu rozwoju informatyki, że mamy nowe wyzwania dla algorytmów eksploracji danych. Potrzebne są takie rozwiązania, które spełniają rygorystyczne wymagania wydajnościowe, np. w odniesieniu do zarządzania pamięcią oraz przetwarzania danych i znajdowania odpowiedzi w ograniczonym czasie. Problemem jest także to, że reprezentacje danych są zróżnicowane i przez co trudne w analizie oraz pochodzą z wielu źródeł jednocześnie. Wiele standardowych systemów informatycznych jest oparte na relacyjnym modelu danych, który jest dobrze zdefiniowany i posiada precyzyjną strukturę logiczną. Lecz w zastosowaniach Big Data analizowane dane są dostępne w różnych formatach. Większość z nich to dane tzw. semi-strukturalne (np. jak strone internetowe) albo zupełnie niestukturalne - takie jak dokumenty, teksty, wpisy na mediach społecznościowych lub obrazy. Wiele z takich danych, jak posty – teksty użytkowników internetu, są na ogół dość kiepskiej jakości do automatycznego przetwarzania. Wg. Wiele raportów dobrze zdefiniowanych (w pełni strukturalnych) danych jest tylko 10%.

Silnie zróżnicowane dane występują także w wielu innych zastosowaniach - prelegent omówił przykłady z konstruowania medycznych systemów informatycznych, które zawierają zarówno standardowe bazy danych, jak i surowe zdjęcia, zapisy sygnałów i opisy tekstowe, co powoduje, że algorytmy muszą analizować różne równocześnie dane w różnych formatach i bardzo szybko je przetwarzać - co jest źródłem trudności. Algorytmy muszą być także przygotowane na to, że mają sobie radzić z niepewnością związaną z ew. nieprecyzyjnością zapisu niektórych pomiarów.

Następnie prelegent zwrócił uwagę, że wielu zastosowaniach, np. systemach finansowych, działaniach reklamowych, równocześnie konieczne osiągnięcie rezultatu działania

algorytmów w bardzo krótkim czasie, nawet kosztem jakości wyniku:

- Jeżeli odpowiedź nie nadejdzie w odpowiedniej chwili, to będzie strata. Jeżeli myślimy o personalizowaniu reklam, decyzji finansowej etc., to chcemy otrzymać jak najszybciej, dobrze sformatowaną wiadomość. Może być przybliżona, ale powinna zostać przekazana w odpowiednim czasie.

Jako sztandarowy przykład wykorzystania algorytmów Big Data, prelegent podał Google Translate. Rozwiązania oparte są o algorytmy statystycznej analizy języka naturalnego, które badają różne korpusy tekstów i zdygitalizowane przykłady gotowych tłumaczeń z dwóch języków. Zasada działania algorytmów opiera się na szukaniu statycznych zależności między pojedynczymi słowami, frazami, zdaniami, itd.. I jego siła wynika przede wszystkim z estymacji parametrów modelu statystycznego w oparciu o olbrzymie repozytoria danych o równoległych tłumaczeniach tekstów w różnych językach, które nie były dotychczas dostępne.

Inną ilustracją złożonych projektów Big Data było krótkie omówienie wyników analizy mobilności mieszkańców Toskanii oraz studium socjologicznego tworzących się mini społeczności realizowane przez wiele lat przez laboratorium KDD Uniwersytetu w Pizie.

Jako źródła referencyjne opisu wielu projektów Big Data profesor Stefanowski podał świeżo przetłumaczoną na język polski książkę Viktora Mayera Schonberger'a i Kennetha Cukiera „Big Data: a revolution that will transform how we live, work and think” oraz wspólną (zredagowaną wraz z Nathalie Japkowicz) nową monografię wieloautorową „Big Data Analysis: New algorithms for a new society” wydaną w tym roku.

Prelegent przypomniał, że badania i zastosowania Big Data nie pojawiły się nagle. O ile termin w literaturze anglosaskiej pojawił się pod koniec ubiegłego wieku, wcześniej prowadzono intensywne badania w takich dziedzinach jak statystyczna analiza danych, sztuczna inteligencja oraz maszynowe uczenie się. W szczególności pod koniec lat osiemdziesiątych pojawiło się duże zainteresowanie tzw. odkrywaniem wiedzy z danych (z ang. Knowledge Discovery in Data) oraz eksploracją danych – data mining. Przy czym rozwiązania stosowane w data mining a Big Data są różne. Data mining jest najczęściej ukierunkowany na przetwarzanie jednej bazy lub hurtowni danych na ograniczonej liczbie serwerów obliczeniowych, a w Big Data najczęściej prowadzi się przetwarzanie rozproszone na relatywnie dużym gronie komputerów, a dane spływają równolegle z różnych źródeł, przez co mamy odmienny model przetwarzania.

Ponadto pojawiły się nowe rodzaje problemów i algorytmów, które nie były obecne albo nie były tak rozwinięte w dawnym świecie data mining. Przykładami są analizy sieci społecznych i szerzej danych reprezentowanych jako grafy, przetwarzanie strumieni danych, obliczenia na urządzeniach mobilnych, zarządzanie pochodzeniem i jakością danych w dużych powiązanych repozytoriach, wizualizacja olbrzymich danych oraz zagadnienia prywatności, ochrony danych oraz aspektów etyczno-moralnych stosowania algorytmów predykcyjnych wobec ludzi.

Rozważając ostatnie zagadnienia profesor Stefanowski stwierdził, że w klasycznej eksploracji danych lub zastosowaniach data mining na typowych bazach lub hurtowniach danych łatwiej można było anonimizować dane i zapewnić ich ochronę. Jednak w przypadku Big Data nie jest to już takie proste. Klasyczne metody anonimizacji są po prostu nieskuteczne. Przykładowo jeśli gdzieś zostawimy ślad aktywności, np. jako użytkownik lokalizatora GPS, i nasza tożsamość zostanie zatarta, to i tak możemy zostać rozpoznani po charakterystycznych sposobach poruszania się jeśli podczas analizy różnych wolumenów danych odnajdzie się korelacje między elementami danych, z których jedna będzie miała aspekty personalne. Przykładowo konkurs badania rekomendacji firmy NetFlix lub analiza tzw. otwartych anonimowych danych z zapytań w serwisie AOL pokazała, że łącząc różne, dodatkowe źródła danych, zidentyfikowano poprawnie tożsamość kilku osób pomimo wysiłków w zapewnienie ich anonimowości.

Podsumowując analiza Big Data prowadzi do rewolucyjnych i fundamentalnych zmian, gdyż gromadzimy znacznie więcej danych niż poprzednio, co prowadzi do

jakościowych zmian. Pozwala bowiem łączyć bardzo wiele źródeł danych i odkrywać korelacje, które mogą być coraz bardziej precyzyjne. Z drugiej strony dane są często współdzielone i udostępnione do dalszej analizy, co rodzi wyzwania dla ochrony danych osobowych oraz wymaga rozpoczęcia badań nad nowymi zagadnieniami etycznymi dla prowadzenia analizy Big Data oraz wykorzystywania wyników. Nie należy jednak pójść zbyt rygorystycznie w zakresie blokowania możliwości. Zdaniem profesora zbyt skrajne zabezpieczenia będą uniemożliwiać interesującą analizę oraz uniemożliwiać zastosowania dobre dla społeczeństwa. Należy prowadzić dyskusje, jak znaleźć rozsądny kompromis między badaczami, firmami a ochroną praw człowieka. Jego zdaniem, rozwiązaniem może być zachęcanie przedstawicieli różnych zainteresowanych środowisk do tworzenia grup eksperckich i prób wypracowania tzw. dobrych praktyk i zasad etycznych korzystania z wyników analizy Big Data. Jesteśmy jednak dopiero na początku takiej drogi.

Dr Edyta Bielak-Jomaa, GODO podkreślała, jak ważna, w dobie big data, jest ochrona podstawowych praw człowieka, do których należy prywatność:

- Dzięki wielkim zasobom danych następuje rozwój wszystkich dziedzin gospodarki, zatem big data to nie samo zło, ale również korzyści. Oczekujemy, że wykorzystanie analityki biznesowej danych przyczyni się do rozwoju innowacyjności w gospodarce i przyniesie postęp w badaniach. Pamięając więc o tym, nie należy zakazywać korzystania z big data, lecz wziąć pod uwagę zagrożenia dla prywatności wynikające z tego zjawiska. Big data może też wkraczać w sferę wolności i intymności, stąd konieczne jest stworzenie nowych ram prawnych, które zapewnią nienaruszalność praw jednostki, w tym prawa do prywatności, które jest jednym z podstawowych praw człowieka. Tym bardziej, że to dane osobowe powodują prawdziwą rewolucję, a nie maszyny i urządzenia.

Minister zaznaczała, że wynikające z big data korzyści są odczuwalne w codziennym życiu każdego z nas. Ułatwiają bowiem i upraszczają załatwianie wielu spraw, przez co poprawiają jakość naszego życia. Jednak korzystając z tych udogodnień musimy sami zadbać o prawo do prywatności i ochrony naszych danych osobowych, a przetwarzające je podmioty uczulać na konieczność przestrzegania podstawowych, wynikających z przepisów prawa, zasad. Administratorom danych należy przypomnieć, że danych osobowych nie można swobodnie łączyć i udostępniać innym podmiotom. Przetwarzanie danych w wielkich bazach musi odbywać się na jednej ze wskazanych w ustawie podstaw prawnych, którą może być np. zgoda osoby, której dane dotyczą. Istotne znaczenie ma także to, by dane nie były wykorzystywane w innym celu niż zostało to pierwotnie określone.

Dr Edyta Bielak-Jomaa zasygnalizowała, że big data rodzi pokusę, by pozyskiwać dane w nadmiarze i przetwarzać je ponad okres niezbędny do osiągnięcia określonego celu. Bywa też tak, że dane gromadzone są na zapas, co prowadzi do naruszenia zasady adekwatności przetwarzania danych. Dlatego tak ważne jest, by podmioty, które przetwarzają dane, dopełniły w stosunku do każdej osoby, której dane wykorzystują, obowiązku informacyjnego, tj. wskazały, kto, jakie dane na jej temat i w jakim celu przetwarza. Administratorzy danych powinni też zapewnić każdemu prawo dostępu do danych oraz prawo do ich poprawiania lub usuwania. Ponieważ jednak realizacja obowiązków informacyjnych w big data wiąże się z dużymi kosztami, niektóre podmioty obecnie całkowicie odstępują od ich realizacji, co jest niezgodne z prawem.

GODO mocno niepokoi też tworzenie profili predykcyjnych. Wydobywanie informacji w celu przewidywania przyszłych trendów czy zachowań jest fascynujące, ale te wnioski mogą nie być prawdziwe. W tym miejscu rodzi się niebezpieczeństwo, że zestawiając dane, będziemy oceniać np. prawdopodobieństwo długości życia i na tej podstawie podejmować decyzję o wysokości ubezpieczenia albo szacować, czy dana osoba będzie dobrym klientem czy też nie i oferować jej określone rabaty. To może prowadzić do dyskryminacji.

Dr Edyta Bielak-Jomaa wskazała, że na poziomie unijnego rozporządzenia operacje, które mogą nieść zagrożenie dla praw i wolności osób, będą musiały być poprzedzone formalną oceną ryzyka i skutków ich wpływu na prywatność oraz poziom ochrony danych (Data Protection Impact Assessment, DPIA). Rozwiązanie to ma na celu

wymuszenie na posiadaczach dużych zbiorów danych dokonania uprzedniej analizy wszystkich możliwych skutków przetwarzania danych na ochronę prywatności. Tym samym zwiększą się wymagania w stosunku do administratorów danych, którzy chcąc zastosować określone rozwiązania, będą musieli zaprojektować je tak, by zapewniały ochronę prywatności.

- Żyjemy w ciekawych czasach, w epoce cyfryzacji i wkraczamy w zupełnie nowy świat – Internet Rzeczy. Wszystkie używane przez nas przedmioty będą komputerami i będą między sobą wymieniać się informacjami. Będą za nas myślały, będą wiedziały, czego nam potrzeba i czego oczekujemy. W związku z tym pozyskiwanych będzie więcej niż do tej pory informacji na nasz temat i na ich podstawie będziemy profilowani. Mając to na uwadze, trzeba stworzyć odpowiednie podstawy prawne takiego działania.

Janusz Dygaszewicz, dyrektor Departamentu Programowania i Koordynacji Badań w Głównym Urzędzie Statystycznym zoobrazował, że GUS to „oczy i uszy” państwa w dobrej sprawie. Jednocześnie firmy prywatne zbierają więcej danych niż podmioty publiczne. Łatwiej jest także regulować jednostki publiczne niż prywatne. To wszystko sprawia, że przewaga w przetwarzaniu danych obecnie znajduje się w rękach prywatnych. Jego zdaniem potrzebny jest balans między danymi, które mogą zbierać zarówno podmioty z sektora publicznego, jak i prywatnego dla dobra nas wszystkich.

- W ciągu ostatnich dwóch lat ludzkość wyprodukowała więcej danych niż dotychczas. Znaczenie ma, jak je wykorzystamy. W GUS możemy zbierać tylko te dane, które są niezbędne do potrzeb funkcjonowania państwa.

GUS współpracuje z innymi podmiotami w zakresie big data – np. operatorami ruchu drogowego czy granicznego. Prelegent wyjaśnił, że dane jednostkowe, które przekazywane są do GUS, nie są przekazywane innym podmiotom. Dane te jedynie są przekazywane do GUS, a następnie po połączeniu niezbędnych danych anonimizuje się i wyciągane są dane statyczne. Te dane są objęte tajemnicą. W Urzędzie przykładą się dużą wagę do jej zachowania. Dodał, że są prowadzone inicjatywy na poziomie Komitetu Gospodarczego ONZ czy Eurostatu, gdzie w kontekście big data i ochrony danych dostrzeżono duży problem w obszarze zachowania prywatności.

Prelegent podkreślił, że GUS jest urzędem badawczym, któremu można zaufać, bo przetwarza dane w bezpieczny sposób. Z tego względu jednostce tej zależy na tym, by mieć dostęp do dużej grupy informacji i łączyć je z różnych źródeł. Firmy już teraz robią to bez żadnych ograniczeń. Jego zdaniem, poważnym problemem jest to, że GUS nie może przetwarzać danych z różnych źródeł i tym samym skutecznie za pomocą big data dla potrzeb statystyki.

Agnieszka Komar-Morawska, dyrektor w Gabinecie Prezesa Głównego Urzędu Statystycznego omówiła prawne dylematy statystyki publicznej. Wskazała, że statystyka publiczna podlega rygorom konstytucyjnym (w szczególności art. 7 Konstytucji) i przetwarzanie danych może odbywać się tylko w granicach prawa i na podstawie prawa:

- Jeśli myślimy o big data to powinniśmy i chcielibyśmy, jako Urząd, mieć jasną i niepodważalną podstawę prawną. Proces przetwarzania danych statystycznych jest ograniczony różnymi przepisami prawa. Nie wolno nam pozyskiwać żadnych danych nadmiarowo i nie w powiązaniu z prowadzonymi badaniami.

Dodała, że w art. 10 ustawy o o statystyce publicznej uregulowana jest tajemnica statyczna. Na jej podstawie dane jednostkowe, które trafiają do Urzędu nie są przekazywane innym podmiotom w tej samej postaci. Tajemnica statystyczna gwarantuje, że dane na temat obywateli nie są w żaden sposób ujawniane. Tajemnica ta jest źródłem zaufania do tego co się dzieje z danymi, czyli w jaki sposób są przetwarzane. Podobnie w prawodawstwie europejskim dane jednostkowe są chronione. Zawód statystyka jest obwarowany szeregiem norm deontologicznych i standardów.

Prelegentka zaznaczyła, że potrzebne są nowe przepisy dla tego by móc przetwarzać dane w big data:

- *Statystyka publiczna bardzo chce sięgnąć po big data. Jednak do tego potrzebna jest właściwa legalizacja i uzyskanie przyzwolenia społecznego.*

Podobnie jak jej przedmówca jest zdania, że potrzebne jest wyrównanie szans pomiędzy sferą publiczną a prywatną. Przy czym w jej ocenie warto pamiętać, że statystyka działa dla dobra publicznego więc big data przez Urząd będzie wykorzystany tylko dla dobra publicznego.

Dr Arwid Mednis, nawiązując do tych wystąpień, wskazał, że przy ostatnich zmianach ustawy o GUS właśnie katalog przetwarzanych danych był przedmiotem kontrowersji. Jak podkreślił, jest to niewątpliwie temat ważny do omówienia szerzej w kolejnej dyskusji. Jednocześnie moderator płynnie nawiązał do następnego wystąpienia. Przypomniął, że dawniej Policja zbierała dane operacyjne „z ulicy”, a dzisiaj zatrudnia analityków, których głównym zadaniem jest analiza zgromadzonych przez Policję danych. Dodał, że policja niemiecka korzysta już z analiz przewidywalności, a Amerykanie testują rozwiązanie, któreznajac skład sądu z 70% skutecznością szacuje, jaki będzie wyrok.

Insp. Andrzej Naperty odpowiedzialny za analizę kryminalną i systemy informacyjne w Komendzie Głównej Policji wskazał, że organy ścigania chciałyby przetwarzać jak najwięcej danych, ale nie mogą przekraczać norm europejskich czy krajowych, a w szczególności konstytucji. W ocenie prelegenta, mówiąc o big data i policji zawsze zastanawiamy się o inwigilacji. Tymczasem big data to pobieranie, przetwarzanie i wizualizacja danych. Sama w sobie nie wiąże się z inwigilacją. Polega na szybszym zbieraniu danych i wyciąganiu wniosków. Efektem końcowym jest co prawda profil zachowań człowieka czy profil sytuacyjny, ale nie ma to przełożenia na inwigilację. Policja nie prowadzi indywidualnych profili osobowych, a wirtualny zamiar nie jest karalny.

- *Organy państwowe muszą ściśle przestrzegać reguł prawa. Dane pochodzą z różnych źródeł zewnętrznych, np. z Internetu. Policja także prowadzi działania poprzez obserwację profili w social media. W świecie analityki pojawił się też nowy kierunek big data jakim jest wykorzystanie wartości drzemiących w takich źródłach danych, które były do tej pory marginalizowane.*

Zastępca dyrektora Biura Służby Kryminalnej dodał, że w przeszłości wystarczała ogólna wiedza o organizacji przestępczej, a teraz policja w zakresie prewencji zajmuje się także szukaniem trendów i tworzeniem wzorców. Podejście proaktywne oparte na analizie danych i big data właśnie jest bardzo ważne. Przykładowo burmistrz Nowego Jorku, Bloomberg, ma specjalną jednostkę analizującą w big data. W Polsce powstały dedykowane wydziały przy komendach wojewódzkich, które zajmują się gromadzeniem i analizą informacji. Są dwa rodzaje analiz. Analiza strategiczna, która pozwala określić rozkład zagrożeń czy przestępczości na danym terytorium gdzie bazą są zarówno dane statyczne, ekonomiczne, dobrobki naukowe (np. kryminalistyka, kryminologia) czy wręcz cały szereg dziedzin, który pozwala przewidywać zagrożenia. Drugi rodzaj to analiza operacyjna polegająca na analizie profilu np. cech ofiar.

Policja dokonuje analiz m.in. korzystając z Krajowego Systemu Informacji Policji i Informacji Kryminalnej, zbiorów AFIS (gdzie gromadzone są linie papilarne oraz ślady NN - do których nie można przypisać żadnej osoby). Prelegent dodał, że ślady NN nie są gromadzone na zapas gdyż mają związek z prowadzonym postępowaniem. Przy okazji zabezpieczenia śladów zatrzymanego przy innym przestępstwie może okazać się, że popełnił on lub uczestniczył w innych przestępstwach, w których sprawy wcześniej były prowadzone. W kontekście zaś danych telekomunikacyjnych to warto pamiętać, że liczba wniosków, które składa policja jest sumą wszystkich złożonych zapytań do operatorów, a dane uzyskiwane są tylko od jednego operatora, który ma dane określonego abonenta. Z tego względu wnioski o udostępnienie danych trzeba dzielić na co najmniej cztery. Ponadto policja występuje do operatorów telekomunikacyjnych dopiero jak ustalony zostanie związek przyczynowo-skutkowy i dane są faktycznie potrzebne.

Dr Łukasz Bolikowski z Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego zaznaczył, że w medycynie i socjologii stosuje się analizę sieci społecznych z uwagi na bardzo duże ilości danych przetwarzane w nich w postaci cyfrowej. Mogą to być dane tekstowe, graficzne czy nawet dźwiękowe. W medycynie te dane są przetwarzane w celu lepszego zrozumienia procesów, jakie zachodzą w naszych organizmach oraz predykcji, np. pod kątem skłonności do posiadania pewnych chorób.

Prelegent podkreślił, że dla osób, których dane dotyczą, jest różnica między przetwarzaniem danych przez lekarza do celów medycznych a tym, jak dane medyczne przetwarzane są „na własną rękę” przez urządzenie, które nie informuje, co robi z danymi, które pozyskuje na temat tych osób.

W medycynie jest szereg zbiorów danych, które są dzielone i publikowane po wstępnej anonimizacji. Dzieje się tak z uwagi na fakt, że interesujące są, że medycznego punktu widzenia interesują się dane o pewnym zjawisku, chorobie a nie o konkretnej osobie fizycznej. Jednocześnie zbierając dane bardzo wiele metadanych identyfikujących osoby fizyczne może być anonimizowane.

- Badamy, jak rozprzestrzeniają się choroby w społeczeństwie. Nie jest interesujące, kto choruje, ale jak rozprzestrzenia się choroba. Takie badania pozwalają lepiej przewidywać pewne działania naprawcze bez uszczerbku dla prywatności osób, których dane w jakimś stopniu przyczyniły się do określonych wniosków.

Doktor podkreślił jednak, że bardzo ważne jest, by badacze mogli wymieniać się danymi, by mogli wzajemnie zweryfikować uzyskane wyniki. W tym kontekście jest problem otwartości publikowania danych, tzw. otwartej nauki.

Agnieszka Grzelak reprezentująca na konferencji Rzecznika Praw Obywatelskich wskazała, że big data zmienia paradygmat do ochrony prywatności. Narzędzia pozwalają na przewidywanie zachowań jednostki. Dodała, że RPO przygląda się, jakie dane przetwarzane są w różnego rodzaju danych w rejestrach. Przykładowo w 2013 r. Rzecznik Praw Obywatelskich skierował do Ministerstwa Zdrowia obawy tworzenia rejestrów na podstawie rozporządzenia a nie ustawy. Sprawa znalazła swój finał w Trybunale Konstytucyjnym, który podzielił część zarzutów Rzecznika w wyroku z 18 grudnia 2014 r. W efekcie odbyło się kilka spotkań, by uregulować kwestie sporne ponownie i uchwalono nową ustawę.

Prelegentka opisała, że Rzecznik miał również uwagi do systemu informacji oświatowej. Początkowo system ten gromadził dane w sposób zbiorczy, w 2011 r. wprowadzono reformę tego systemu i oparto system o zbieranie danych jednostkowych. W opinii Urzędu, ustawa regulująca system informacji oświatowej budzi wątpliwości. Katalog danych, które mogą być przetwarzane, jest zbyt szeroki. Jej zdaniem należy dokonać oceny zgodności w zakresie autonomii przetwarzanych danych o obywatelach. W odniesieniu do danych o stanie zdrowia dla sektora ubezpieczeniowego i bankowego gdzie big data jest wykorzystywane Rzecznik stoi na stanowisku, że kluczowe jest jakie dane i w jakich okolicznościach będą udostępniane za zgodą. W tym kontekście RPO szczególnie przyglądał się działaniom prowadzonym w ramach umowy o rezydentach podatkowych zawartej pomiędzy Polską a Stanami Zjednoczonymi Ameryki Północnej. Zdaniem Agnieszki Grzelak umowa budzi uzasadnioną wątpliwość co do zgodności transferu danych z konstytucyjnym standardem ochrony prywatności obowiązującym w Polsce.

Prelegentka zasygnalizowała także, że RPO przygląda się najnowszym zmianom w ustawie o policji w szczególności w zakresie udostępniania danych internetowych, pocztowych czy telekomunikacyjnych.

- Rzecznik Praw Obywatelskich wyraził głębokie zaniepokojenie nową ustawą o policji i wystosował zapytania. Podzielił pogląd, że specyfika nowych technologii uzasadnia powierzenie adekwatnych uprawnień, dzięki którym organy będą w stanie zapobiegać przestępstwom i je wykrywać. Działania te jednak nie mogą naruszać przepisów prawa.

Obywatele też powinni być uprzedzeni o stosowaniu tych środków.

Sesję drugą o big data w sektorze prywatnym moderował **dr Grzegorz Sibiga** z Instytutu Nauk Prawnych Polskiej Akademii Nauk, a pierwszą prezentację w tej części przedstawił **prof. Wojciech Bijak**, dyrektor Ośrodka Informacji Ubezpieczeniowego Funduszu Gwarancyjnego. Podkreślił, że w ujęciu sektora, który on reprezentuje, badanie statyczne nie daje pełnej odpowiedzi na stawiane pytania i hipotezy, stąd istnieje potrzeba przetwarzania na poziomie danych jednostkowych. To zaś wiąże się z potrzebą anonimizacji danych, które będą przekazywane do badacza, który będzie pracował na zaanonimizowanych danych jednostkowych. Dodał, że takie badania mogłyby być prowadzone również policją, np. korelacji zachowaniu kierowców na drogach i posiadanych przez nich punktów karnych a wpływem na szkodowość, na liczbę wypadków.

Prelegent nadmienił, że UFG chce udostępniać oraz pozyskiwać nowe dane dla prawidłowego funkcjonowania i współpracy z innymi organami:

- Niedaleka przyszłość wiąże się ze zmianą roli, jaką będzie pełnił Ubezpieczeniowy Fundusz Gwarancyjny. Chcemy pełnić rolę brokera informacji chociażby między policją czy ZUS. Chcemy wystąpić w roli procesora. To na nas będzie ciążyła odpowiedzialność, aby pytania były składane zgodnie z celami przez uprawnione podmioty, osoby.

Nawiązując do big data, profesor opisał podstawy działania systemu do wykrywania nieubezpieczonych pojazdów mechanicznych. W założeniu system ma informować, na kogo nakładać opłatę karną poprzez wyszukiwanie negatywne – kto nie dokonuje opłat ubezpieczenia OC.

- UFG może dostawać informacje o braku OC od różnych podmiotów, ale też mamy swój własny system. Działa on tak by „odsiać” jak największą liczbę przypadków gdzie można byłoby kogoś ukarać. Chcemy stworzyć platformę (organy ścigania, zakłady ubezpieczeń, osoby), która byłaby zintegrowaną platformą identyfikacji i weryfikacji zajwiska przestępczości ubezpieczeniowej. Z dniem wejścia w życie ustawy o działalności ubezpieczeniowej UFG powstanie Baza Danych Ubezpieczonych. Analizujemy ją właśnie pod kątem privacy by design and by default. Chcemy przeprowadzać akcje społeczne, by wskazywać pewne schematy działań przestępczych. Jak np. takie, że przy zdarzeniach komunikacyjnych ofiara może być jednocześnie sprawcą – gdy osoba cały dzień jeździ po parkingu szukając kogoś, kto uderzy w jej samochód.

Kamil Zegier z Orange Polska przedstawił, jak dane dotyczące lokalizacji użytkowników sieci komórkowej są przetwarzane przez operatora telekomunikacyjnego. Podkreślił, że w ciągu 1 dnia przetwarzane są setki tera bajtów ruchu internetowego, dziesiątki megabajtów smsów dziennie. W jego ocenie pomimo tego, że metadane w sieci są różnorodne to te szczególne związane z geolokalizacją są interesujące np. dla firm technologicznych. Opisał, że dane o lokalizacji są przez operatorów anonimizowane, a dopiero później są agregowane.

- Dzięki big data możemy przeprowadzić badania ilościowe. Przykładowo jakie jest natężenie ruchu w określonych godzinach. Drugie zastosowanie, które jest większym wyzwaniem jeżeli chodzi o anonimowość to badania bardziej szczegółowe np. skąd przyjechali kierowcy. Skupiamy się na agregatach i danych zanonimizowanych. Nie możemy analizować danych jednostkowych.

Paweł Gora, doktorant na Uniwersytecie Warszawskim, który w swej pracy zawodowej i badawczej zajmuje się modelowaniem i optymalizacją ruchu drogowego opowiedział o zastosowaniach big data.

- Mamy dwa typy danych, które mogą znaleźć wiele zastosowań i mogą być bardzo przydatne: dane zagregowane – macierze podróży, natężenie ruchu oraz dane indywidualne – położenie, czasy jazdy, źródła i cele podróży, preferencje kierowców. Takie dane zbierane mogą być wyłącznie za zgodą.

W ocenie prelegenta badania big data dotyczące natężenia ruchu mogą być pomocne w smart city do wyznaczania lokalizacji parkingów, stacji benzynowe, stacji ładowania pojazdów elektrycznych, siedzib firm logistycznych, szkół, szpitali pod kątem dobrego i sprawnego dojazdu czy wypożyczalni pojazdów. Projektowanie sieci drogowej też może uwzględniać wykorzystanie big data. Dla przykładu systemy informowania podróżnych nie tylko zbierają, ale też udostępniają określone dane, które mogą mieć lub mają wpływ na podejmowane decyzji przez podróżnych. Podobnie big data może pomóc w ustalaniu składki za ubezpieczenie. Przykładowo aplikacja może wykrywać, jak dana osoba korzysta z samochodu – np. czy nie jest uczestnikiem wypadków, czy rzadko korzysta z pojazdu i dobierać np. mniejszą składkę.

Dr Tadeusz Białek, dyrektor zespołu prawno-legislacyjnego Związku Banków Polskich wskazał, że big data w bankach nie jest czymś nowym. Banki korzystają z systemów bazodanowych zewnętrznych i wewnętrznych.

Nawiązując do przykładów big data prelegent podał System Dokumenty Zastrzeżone, który gromadzi informacje o dokumentach (skradzionych, zagubionych). Dzięki temu systemowi udało się zapobiec wyłudzeniom kredytów w latach 2008-2015 na łączoną kwotę ponad 3 mln złotych, a łącznie zastrzeżono ponad 1 mln dokumentów. Warto dodać, że znajdują się w nim nie tylko dowody tożsamości, ale w ogóle dokumenty ze zdjęciem oraz, że te dane udostępniane są też organom telekomunikacyjnym.

Inny projekt mieszczący się w big data to System Wymiany Informacji o Zagrożeniach. Mieszą się tam informacje dot. wyłudzeń kredytów oraz powiązane z tymi zdarzeniami dokumenty, które są fałszowane lub kradzione.

- Od 1 lipca 2016 projekt nowelizacji ustawy prawo bankowe wprowadzi system, na który czekaliśmy wiele lat. Mowa o Centralnej Informacji o rachunkach (rejestr rachunków bankowych). Krajowa Izba Rozliczeniowa będzie prowadzić ten rejestr. Nie będzie to system bazodanowy. Będzie to system działający na zasadzie platformy odpytującej. Kowalski będzie mógł w banku zapytać gdzie posiada rachunki bankowe albo zmarły posiadał rachunki bankowe. Będzie miał informację o swoich bieżących rachunkach, a także sądy i prokuratura będą mogły kierować pytanie w jedno miejsce. Zmiana dotyczy też wygaśnięcia rachunku, który będzie wygasał z mocy prawa jeżeli przez 10 lat nie będzie żadnych dyspozycji na rachunku (aktualnie w ciągu 2 lat jak nie ma minimalnego obrotu).

Banki rozszerzają swoją współpracę i zawierają ją np. z operatorami telekomunikacyjnymi. W ramach grupy kapitałowej dane udostępniane są za zgodą odpowiednio abonenta lub klienta i wiążą się z przekazaniem bonusów za wyrażenie zgody.

Na zakończenie, dyrektor Tadeusz Białek podkreślił, że dane bankowe są w restrykcyjnym nadzorze prawnym i Komisja Nadzoru Finansowego również bada obieg informacji oraz banki współpracują z GIODO. Jednocześnie trzeba pamiętać także o tym, że jeśli klient spłaca zobowiązanie nieterminowo, a później żąda natychmiastowego usunięcia danych na jego temat, to nie jest to realizowane, bo prawnie 5 lat taka informacja jest przechowywana w systemie dla oceny zdolności kredytowej. Leży to w interesie podmiotu, który udziela kredytu, który ze swej istoty oznacza, że usługa ta obarczona jest wysokim ryzykiem.

Dr Grzegorz Sibiga podsumowując wystąpienia dodał, że warto byłoby rozważyć zakres uzasadnionej ingerencji rozwiązań big data w ochronę prywatności. W jego ocenie zbyt mało mówi się o zagrożeniach, a zbyt dużo o korzyściach. Zachęcił, by kontynuować dyskusję pod tym właśnie kątem. W jego ocenie klient początkowy nie ma świadomości, w jaki sposób będą wykorzystywane jego dane, a klauzule informacyjne powinny być bardziej transparentne.

Trzecią sesję - traktującą o big data w świetle nowych regulacji prawnych o ochronie danych osobowych - moderowała **Monika Krasieńska**, dyrektor Departamentu Orzecznictwa, Legislacji i Skarg w Biurze Generalnego Inspektora Ochrony Danych Osobowych. Podkreślała, że na administratorach danych ciążyą wszystkie – wynikające

z ustawy o ochronie danych osobowych – obowiązki, w tym m. in. obowiązek legalnego przetwarzania danych, z poszanowaniem zasady adekwatności. Każdy z nas ma zaś prawo do prywatności, a także prawo do informacji, kto, w jakim zakresie i celu wykorzystuje nasze dane. Przysługuje nam również prawo do kontroli przetwarzania dotyczących nas danych. Jednocześnie dodała, że big data i profilowanie to zjawiska, które GODO będzie bacznie obserwował pod kątem przestrzegania wskazanych zasad i praw.

Dr Paweł Litwiński z Instytutu Allerhanda razem z Anną Walkowiak z Fundacji Panoptykon wspólnie poprowadzili prezentację poświęconą profilowaniu – w ocenie prelegentów, w kontekście big data należy zwrócić szczególną uwagę właśnie na profilowanie. W rozporządzeniu unijnym podjęto próbę definicji profilowania, co wskazano w motywie 57 preambuły. Dzięki temu można wskazać na cztery elementy profilowania: automatyczne przetwarzanie informacji, ewaluacja (ocena) osoby, analizowanie (dobranie modelu) lub przewidywanie jej zachowań (tzw. profilowanie produkcyjne).

Profilowanie może przybierać różne formy. W kontekście pracowników może być to np. analizowanie zachowań w pracy, dla sklepu internetowego analizowanie zainteresowań użytkowników pod kątem wyświetlania im reklam, dla banku może być to ocena wiarygodności kredytowej osoby, w smart city sposoby przemieszczania ludności. Prelegenci wskazali jednak na istotne zagrożenia związane z profilowaniem.

- Dane, które nas dotyczą się od nas odrywają, tracimy kontrolę nad tym co się z nimi dzieje i jak są wykorzystywane – mówiła Anna Walkowiak.

Problemem jest też przechowywanie danych stworzonych w ramach profilu, a jeszcze nie wykorzystywanych. Możemy bowiem wyobrazić sobie sytuację, w której zostanie stworzony profil, ale na jego podstawie nie zostaną podjęte żadne działania wpływające na sytuację prawną lub faktyczną osoby, której dane dotyczą. Następnie profil ten zostanie wykorzystany w dalszym profilowaniu. Tu pojawia się pytanie, czy ten dane będą podlegać ochronie na podstawie rozporządzenia.

W definicji profilowania w rozporządzeniu występuje element ocenny, jakim jest konieczność ustalenia, czy profilowanie „w istotny sposób wpływ na sytuację faktyczną” osoby profilowanej. Posłużenie się taką klauzulą powoduje, że trzeba będzie ją uszczegółowić w orzecznictwie i dopiero konkretne sprawy pokażą właściwy kierunek. Jednak do tej pory nie doczekaliśmy się przez sądy administracyjne wykładni prawnie usprawiedliwionego celu, co zdaniem dra Pawła Litwińskiego pozwala na odrobinę sceptycyzmu, co do dokonania wykładni pojęcia istotnego wpływu na sytuację faktyczną profilowanego. Jednocześnie dodał, że profile mogą być wykorzystywane do interakcji z innymi osobami i obecne przepisy nie są wystarczająco precyzyjne.

- Zasada jest taka, że profilowanie jest niedopuszczalne. Profilowanie jest dopuszczalne jeżeli jest to niezbędne do zawarcia lub wykonania umowy (np. ocena wiarygodności kredytowej), zostanie wyrażona wyraźna zgoda, przy czym w rozporządzeniu pojawia się wprost, że zgoda wyrażona w ramach stosunku pracy nie jest zgodą dobrowolną albo przepisy prawa krajowego będą na to zezwalać – powiedział dr Paweł Litwiński.

W ocenie prelegentów profilowanie powinno zapewniać realizację praw osób, których dane dotyczą, a przepisy prawa krajowego nie mogą być sprzeczne z przepisami rozporządzenia. Jednocześnie nowe przepisy kładą duży nacisk na dobrowolność wyrażanej zgody oraz transparentność profilowania (obowiązek informowania o stosowanym mechanizmie profilowania i ostrzegania o konsekwencjach).

- Mamy być bardziej świadomi, a informacje mają być nam przekazywane w języku dla nas zrozumiałym podkreśliła Anna Walkowiak. Mamy prawo wiedzieć, jaki jest system profilowania, na czym on polega, jaka jest logika profilowania, jak budowany jest profil - dodał dr Paweł Litwiński.

Prelegenci podkreślili, że administrator danych powinien dopuszczać możliwość sprzeciwu wobec profilowania.

Zdaniem dr Pawła Litwińskiego zasada celowości przetwarzania danych osobowych pozostaje w dalszym ciągu elementem istotnym w praktyce, podobnie jak ocena, czy dane są przetwarzane nadal dla celu, w jakim zostały zebrane, czy może już dla innego celu. Obecnie obowiązujące w Polsce przepisy nie regulują tego problemu w sposób właściwy. To się jednak zmieni. Art. 6 rozporządzenia nie będzie wskazywał w formie katalogu warunków, czy przesłanek, od których warunkuje się legalność zmiany celu. To na administratorze danych będzie spoczywało dokonanie oceny legalności, czy przetwarzanie jest dopuszczalne w innym celu, niż dane zostały zebrane.

Dyrektor **Monika Krasieńska** zasygnalizowała, że obecnie Urząd czeka ogromna praca związana z przeglądem polskiego ustawodawstwa, które trzeba będzie dostosować do unormowań unijnego rozporządzenia. Pewne pojęcia będzie bowiem należało zdefiniować na nowo, a niektóre z rozwiązań ustawowych przebudować – np. profilowanie bezrobotnych.

- Choć trudno mi sobie wyobrazić spełnienie obowiązków informacyjnych w przypadku big data, to jednak nie jest to niemożliwe. Warto też zaznaczyć, że prawo do informacji jest jedną z elementarnych kwestii w rozporządzeniu. Tworzenie pewnych prognoz, na które nie będziemy mieć wpływu, może powodować stygmatyzację, dlatego przepisy dotyczące profilowania będą musiały być przez nas dokładnie przeanalizowane.

Włodzimierz Schmidt ze Związku Pracodawców Branży Internetowej IAB Polska podkreślił, że od samego początku istnienia Internetu głównym źródłem jego finansowania jest reklama. Internet należy traktować jako główne medium. Cyfryzacja jest stymulowana przez rozwój Internetu, Internet jest siłą napędową rozwoju gospodarczego. Profilowanie natomiast jest nierozdzielnie związane z reklamą internetową.

- Legislators stanął na stanowisku, że z jednej strony powinniśmy ograniczyć wykorzystanie profilowania, ale jednocześnie chcemy by gospodarka się rozwijała. Na reklamę w internecie wydaje się w Polsce ponad 3 mld zł co stanowi już blisko 30% wszystkich wydatków na reklamę. W Europie na reklamę internetową wydaje się ponad 30 mld euro. Pod koniec 2015 wydatki na reklamę w Internecie były już większe niż na reklamę w telewizji. Wartość dodana brutto wg badania przeprowadzonego na zlecenie IAB Europe przez Instytut IHS to ok. 0,5 biliona euro. Dzięki reklamie internetowej w Europie ma pracę ponad milion osób. Dodajmy do tego jeszcze firmy, które obsługują branżę to liczba ta wzrośnie do 5,5 mln miejsc pracy. Te osoby pracują dzięki temu, że w reklamie można stosować profilowanie. Bez profilowania reklama i marketing będą mocno zagrożone.

Prezes IAB Polska dodał, że przy targetowaniu nie są potrzebne tak szczegółowe dane jak dane osobowe (np. imię i nazwisko). Jako przykład podał dotarcie z reklamą innowacyjnego kosmetyku dla grupy kobiet 20-25 lat. Wskazał, że bez sprofilowania wg tych dwóch parametrów: kobiety w wieku 20-25 lat przedsiębiorca będzie musiał mieć wielomilionowy budżet na rozproszoną reklamę co zahamuje jego rozwój. Zamiast wydawać na innowacje będzie musiał inwestować w reklamę, by o jego produkcie dowiedzieli się klienci. W jego ocenie dane spseudonimizowane (dane osobowe), są dla marketingu dobre. Te dane pozwalają na profilowanie i definiowanie grup docelowych.

Prelegent dodał, że branża będzie dążyła do samoregulacji poprzez wyznaczanie sobie standardów w zakresie pracy na danych, gdyż reklama internetowa jest oparta o big data. Jednocześnie jednak warto pamiętać, że przewidywania zakładają, iż w ciągu najbliższej dekady reklama cyfrowa będzie stanowiła już ok. 80% rozwiązań reklamowych. Mówimy tu o marketingu napędzanym przez dane (data driven marketing). Już teraz realizowane są kampanie pilotażowe reklamy out-door z wykorzystaniem modeli programatycznych opartych na profilowaniu, gdzie reklamy są wyświetlane w zależności od tego, jaka grupa znajduje się w ich zasięgu. Dodał, że big data w reklamie dynamicznie się rozwija i jeśli w Europie zaczniemy wprowadzać nadmierne ograniczenia możemy przestać napędzać gospodarkę tak dobrze jak byśmy chcieli.

Podsumowując tę część spotkania, **Monika Krasieńska** wskazała, że profilowanie jest dopuszczalne, ale na ściśle określonych warunkach. Jeśli ma być oparte na zgodzie, to zgoda ta musi być dobrowolna, a nie wymuszona czy blankietowa. Zawsze zaś zakres pozyskiwanych danych musi być adekwatny do celu ich przetwarzania.

- Profilowanie nie jest zjawiskiem negatywnym, ale stwarza ogromne zagrożenie dla prywatności. Należy zatem zapewnić dużą transparentność tego procesu. W big data mamy do czynienia ze strumieniem danych, które spływają z różnych źródeł. Nie możemy doprowadzić do powstania dyktatury danych. Musimy pamiętać, że moralności i etyki nie zastąpimy algorytmami oraz że kształtowanie profili na domniemaniach jest niebezpieczne. Z informacji można stworzyć groźne narzędzie do ingerencji w podstawowe prawa człowieka.

Na koniec tej sesji zostały zadane pytania. Szczególnie interesujące okazało się pytanie w kontekście profilowania, które w rozporządzeniu powiązane jest z narzędziem do podejmowania automatycznych decyzji. Szczególnie interesujące zostało zadane doktorowi Pawłowi Litwińskiemu - czy możliwy jest przypadek gdy profilowanie będzie miało miejsce, ale nie będzie wykorzystywane do automatycznych procesów przetwarzania.

Dr Paweł Litwiński stwierdził, że profilowanie jest osadzone w automatycznym przetwarzaniu danych, a nawet wiąże się z podejmowaniem decyzji dotyczących osoby fizycznej w sposób zautomatyzowany. Przykładowo w odniesieniu do oceny zdolności kredytowej na podstawie aktywności w social media podejmowana jest automatyczna decyzja. Jednakże przy profilowaniu bezrobotnych to urzędnik podejmuje decyzję. Nie jest to zatem w pełni automatyczne przetwarzanie, a więc rodzą się wątpliwości co do kwalifikacji tego jako profilowanie w rozumieniu przepisów rozporządzenia.

Równie ważne pytanie wiązało się z rolą administratora bezpieczeństwa informacji (ABI) i skargami kierowanymi do GODO związanymi z niezgodnym z prawem przetwarzaniem danych.

Monika Krasieńska w odpowiedzi wskazała, że po zmianie pozycji i roli ABI, to właśnie ABI jako pierwszy będzie rozpatrywał skargi dotyczące przetwarzania danych.

Na zakończenie spotkania wszyscy jego uczestnicy, zainteresowani tematyką nowej roli i pozycji administratorów bezpieczeństwa informacji (ABI) zostali zaproszeni do wzięcia udziału w organizowanej w Warszawie przez GODO we współpracy z Akademią Leona Koźmińskiego konferencji poświęconej tej tematyce, zaplanowanej na 23 lutego 2016 r.

Ochrona danych osobowych w dobie Big Data

raport z konferencji 28.01.2016

Cyberlaw.pl



POMOCNIK
RODO