ГОЛЕМИТЕ ДАННИ И ОБЩЕСТВЕНИЯТ ДОГОВОР: АНАЛИЗ НА КАЗУСА С ИЗТИЧАНЕТО НА ДАННИ ОТ НАЦИОНАЛНАТА АГЕНЦИЯ ЗА ПРИХОДИТЕ

АНТОАНЕТА ГЕТОВА

Софийски университет „Св. Климент Охридски”

tony22a@gbg.bg

BIG DATA AND THE SOCIAL CONTRACT: ANALYSIS OF THE CASE OF DATA BREACH FROM THE NATIONAL REVENUE AGENCY

ANTOANETA GETOVA

Sofia University “St. Kliment Ohridski”

Abstract

It is a well known fact that big data plays significant role in the contemporary information society and even could change it. Is the society ready for such change and to what degree? What are the eventual consequences of the accumulation of big data bases on the relationship between the citizens and the state? The analysis is searching for an answer of these questions, regarding to the case of the data breach from the National Revenue Agency of Bulgaria that happened in 2019.

Keywords: big data, social contract, information control, cyber security

През лятото на 2019 г. беше пробита сигурността на Националната агенция за приходите (НАП) в България . По оценки на самата агенция, данните на 5 млн. души са били свалени чрез хакерски пробив и съответно разпространени до големи български медии и други сайтове/мрежи за свободно сваляне [1]. Това изтичане на данни повдигна много въпроси, като например доколко личната информация на гражданите е защитена от държавата, но също и не повдигна някои въпроси, свързани с въздействието на големите данни върху отношенията между гражданите и държавата, на който проблем е посветено и настоящото изложение.

Въздействието на големите данни върху обществените отношения не се изчерпва единствено с последствията, които би имал подобен пробив. Изключително важно е също да се изясни как събирането и респективно възможността за обобщение на такива

данни от страна на институциите е в състояние да трансформира отношенията между тях и гражданите [2].

Контролът на големите данни

За да се обясни въздействието на големите данни трябва да се изясни какво всъщност представляват те. Събирането на данни, поне от страна на държавната администрация не е ново явление. Наличието на „големи данни" обаче означава нещо различно: накратко казано, това е информация която може да се обработи накуп по различни критерии в големи обеми [**Singh, Firdaus, Sharma**, 2015]. Това става възможно именно благодарение на новите технологии, позволяващи дигитализация на информацията. Тези нови технологии позволиха не просто систематичното и относително улеснено събиране на масиви от данни, но и по-лесното им и бързо обработване. Именно това е причината големите данни да могат да бъдат използвани за клъстеризиране (създаване на сегменти или групи) от единици (хора, компании и др.) със сходни характеристики и респективно за откриването на такива ключови характеристики, по които да се извърши тази сегментация. Също така, с помощта на различни статистически методи е възможно от големите бази данни да се открият и модели на поведение чрез които се разпознава принадлежността на единиците към съответния сегмент/група или в по-широк смисъл да се прави прогнозиране на бъдещо поведение. Тези модели стоят в основата и на т.нар. машинно учене при алгоритмите, които някои софтуери използват за съответното разпознаване на единиците при да речем, сервиране на определен тип реклама в различни интернет – страници. [**Perlich, C., Dalessandro, B., Raeder, T.** , 2014].

Как обаче подобни възможности за използване на големите данни са в състояние да влияят върху обществените отношения и по-конкретно в аспекта на взаимодействията между гражданите и държавните структури?

Информацията като такава е в основата на тези социални взаимодействия. Подобно твърдение изглежда саморазбиращо се: именно благодарение на обмена на информация става възможно общуването между които и двама социални дейци (хора, компании, институции) което иначе не би се осъществило. Респективно достъпът до повече или по-малко информация позволява на едната страна да опознае отсрещната,

което дава възможност както за подобряване на общуването, но и за контрол. Механизмът на това влияние може да бъде илюстриран чрез класическия модел на идеалния затвор на Джереми Бентам за паноптикума или „всевиждащото око", доразвит в последствие от Мишел Фуко в книгата му „Надзор и наказание", третираща именно властта на държавата над индивида като паноптикум [**Фуко**, 1998]. В случая с данните контролът става не толкова чрез директно наблюдение, а чрез придобиване на все повече информация и съответно опознаване (т.е. наблюдаване) на отсрещния чрез тази информация, както е в примера със сервираната чрез софтуер реклама.

Когато става въпрос за отношения между гражданите и държавата в лицето на нейните институции, отказът за предоставяне на информация на отсрещната страна (т.е. на държавата) е ограничен. Държавата има право – в определени граници – да изисква такава информация с цел продължаващото й функциониране, което следва да обезпечава общото съществуване на гражданите в тази държава. Това въплъщава идеята за т.нар. обществен договор, който съществува между държавата и гражданите и който е описан още в едноименната книга на Ж. Ж. Русо [**Русо**, 2018]. Респективно споделянето на тази информация от страна на гражданите към държавата би следвало да е ясно определено: каква част от информацията и коя информация гражданите имат право да откажат да споделят, както и обратното: каква информация държавата може да изисква и за какво тя ще бъде използвана. Не на последно място стои и опазването на неприкосновеността на тази информация: информацията следва да се предоставя от страна на гражданите за използването й за определени цели, което изисква и гарантиране от страна на държавата че няма да бъде използвана по друг начин или че няма да бъде предоставена на трети страни за използването й по друг начин (т.нар. злоупотреби с информация).

Тези условия важат за събирането на всякакъв тип информация, независимо от нивото й на обобщение, трансформация и анализ. Когато става въпрос за събиране на големи данни, трябва да се отчете влиянието и на потенциалната „добавена стойност" на информацията: набирането на база от данни по различни характеристики, което дава възможност за допълнително класифициране/сегментиране на единиците, респективно дава възможност за допълнителен контрол. Казано с други думи, при наличие на възможности за събиране на големи данни би следвало в обществения договор да се

предвидят последствията от работата с тях. В идеалния случай следва в законодателната макрорамка да е предвиден и изчерпателно описан въпросът за събирането, съхраняването и анализирането на големите данни.

Въпросът с обмена на информация между гражданите и държавата е разгледан в законодателството ни. Още през 2002 година у нас е приет закон за защита на личните данни. Той не третира големите данни специфично, а е ориентиран към инструменталното използване на данните (за рекламни, журналистически и др. цели). Регламентирани са основните случаи при които лични данни могат да се събират, обработват и съхраняват, както и задълженията на тези, които ги събират и съхраняват във връзка с опазването и контрола върху тях. Съществува и специален орган – Комисия за защита на личните данни, която следи за нарушения на този закон. От месец май 2018 г. влезе в сила и т.нар. Регламент на ЕС за защита на личните данни (GDPR), който третира в детайли възможността на лицата за избор/отказ на предоставяне на лична информация за по-нататъшна обработка, което в частност касае и големите данни[3]. Доколко тези регламенти обаче работят на практика при случаи като този с изтичането на данни от НАП?

Кратък анализ на случая с изтеклите данни от НАП

В контекста на описаното по-горе, случаят с изтичането на данни от НАП следва да се анализа в няколко основни направления.

Първото е свързано със самия пробив на системата, което поставя под въпрос възможността за сигурно съхранение и неприкосновеност на тези данни изобщо. Това би трябвало да бъде неизменна част от обществения договор и респективно е засегнато както в българското законодателство по въпроса, така и в европейския регламент. Редно е да се каже, че не би могло да се гарантира, че изобщо съществува непробиваема система за информационна сигурност, но е под въпрос доколко изобщо данните са били защитени по отношение на съществуващи вече рискове и какви са механизмите за минимизиране на последствията от изтичането на данни. Това са ключови фактори, свързани с обществения договор, доколкото всяка държава е призвана да обезпечава сигурността на гражданите си, включително и информационната такава. В случая с НАП и двата посочени аспекта на информационната сигурност са под въпрос. От една

страна, в публичното пространство дълго беше обсъждано възможно ли е било технически да бъде предотвратен пробивът, извършен по този начин (дистанционно, а не от мрежата на агенцията). Въпреки че няма налични публични данни/информация, чрез която да бъде доказана степента на техническа обезпеченост на НАП, признанието от страна на ресорния министър и министър-председателя за кадрови проблеми в областта на информационната сигурност на държавните институции е симптом на проблем в това направление [4]. Този проблем не е нов за държавната администрация. Такъв имаше и по-рано в друга държавна институция, Агенцията по вписванията, която отговаря за Търговския регистър. Именно поради техническа необезпеченост, през лятото на 2018 г. сървърите на регистъра останаха блокирани в продължение на дни [5]. В този смисъл казусът с пробива на данните от НАП не е изключение, а продължение на вече съществуващ проблем в държавните институции като цяло.

Що се отнася до минимизирането на последствията от изтичането, единственото предприето действие от страна на агенцията е пускане на софтуерно приложение, което потвърждава дали има изтичане на лични данни, което става след идентификация на търсещия справката [https://check.nra.bg/.].

Самото приложение е уязвимо, тъй идентификацията в началото се осъществяваше чрез въвеждане на ЕГН[6], което би могло да бъде въведено от лице, сдобило се с изтекли лични данни. В момента удостоверяването се осъществява чрез специален личен идентификационен номер (т.нар. ПИК) , но това също е проблематично, тъй като в изтеклите данни се съдържа и такава информация[7].

Полезността на приложението също е под въпрос, тъй като при стартирането си то подаваше единствено информация че има изтекли лични данни за съответното лице, но не и какъв вид са те[8]. При пускането му от агенцията не подадоха информация кога се очаква да проработи пълната му функционалност, също липсва и публична статистика за това колко хора са поискали достъп до приложението, докато все още работеше с ограничена функционалност, но е твърде вероятно една голяма част от ползвалите приложението да са го направили именно в дните на пускането му, тъй като това беше съпроводено и от множество медийни публикации и репортажи. В този случай гражданите бяха поставени в двойно неизгодна позиция: от една страна, от риска от изтичане на личната им информация в интернет, а от друга това че те самите не

знаят каква точно информация е изтекла. Парадоксално, но в този период гражданите имат възможност да научат кои техни данни са откраднати, ако потърсят в споменатия масив, достъпът до който е свободен, но не и от страна на агенцията, едно от задълженията на която е да съхранява тези данни по безопасен начин. Към момента, НАП предоставя информация и за съдържанието на изтеклите данни (включително и на живо в офисите си), но това става от сравнително скоро: надграденото приложение е пуснато на 5 септември, т.е. почти два месеца след пробива. НАП закъснява дори и с първия вариант на приложението, тъй като първата „търсачка" за изтекли данни беше пусната от разследващия сайт Биволъ (www.bivol.bg), който също беше получил копие на данните[9].

Друго, което агенцията реализира със закъснение, беше предоставянето на информация за възможно противодействие на пробива от страна на засегнатите лица. Към момента, на страницата на приложението има такава информация, включително линк към клип в мрежата за споделяне на видео, Youtube, който третира най-често задаваните въпроси като: необходима ли е смяна на личните документи при условие че гражданинът научи, че данните му са изтекли и пр. Този клип е качен обаче едва на 10 септември, а пробивът стана през юли т.г. Първата официална информация за това дали гражданите трябва да предприемат насрещни действия и какви да бъдат те, беше разпространена на 26 юли, т.е. чак 10 дни след пробива. В първите дни след пробива ресорният министър дори защити тезата, че този теч не може да навреди на гражданите, което противоречи на смисъла на съществуването на закон за защита на личните данни изобщо[10] . *Това твърдение на финансовия министър поставя под въпрос самите отношения между гражданите и институциите, доколкото едната страна счита за маловажно нарушението на правата на другата, която е реалният суверен в държавата по смисъла на Конституцията*[11].

Противоречивата информация, идваща от различни официални източници в първите дни около пробива може да се приеме като симптом на криза в обществените отношения. Макар и тя да не произтича от връзката с големите данни, това че проблемите, свързани с тях, очевидно не са обхванати подробно в законодателната макрорамка, също въздейства върху тази криза.

Най-сериозния проблем, възпрепятстващ минимизирането на негативните последствия от пробива в данните е свързан с правосъдната система (която би трябвало да е основният механизъм който да работи за това минимизиране). Тук могат да се изтъкнат много аргументи, но ключовият е съдържанието на обвинението, което по думите на главния прокурор ще бъде отправено към потенциалните извършители на това нарушение и то е „кибертероризъм" [12]. Макар че кибер-тероризмът съществува като концепция в изследванията на тероризма изобщо [**Archer, E.M.**,2014], е важно да се отбележи, че в българския Наказателен кодекс понастоящем не съществува такъв състав на престъпление, което пък поставя под въпрос доколко е възможно изобщо осъждане на потенциалите извършители. Но по-важният проблем в случая е „непригодността" на законодателната макрорамка относно негативните ефекти от манипулации с големи данни изобщо: независимо че съществува потенциална опасност от извършване на престъпление което би имало силен негативен ефект върху обществото, то изобщо не е разглеждано като възможност в законодателството ни[13]. *Това още веднъж показва, че настоящата законова макрорамка не е готова да се изправи срещу последствията от наличие на големи данни.*

Не на последно място следва да бъдат анализирани последствията от този теч, свързани със съдържанието на самите данни. Тук трудността произтича от обстоятелството, че директният анализ на информацията в изтеклите файлове данни на практика представлява нарушение на закона за личните данни (тъй като тези данни са изтекли неправомерно в публичното пространство) . Затова по-надолу ще бъде коментирано само известното от официалните източници и медийни публикации относно съдържанието на изтеклите данни.

Първият проблем е мащабът на изтеклите данни. Те засягат над 5 млн. граждани, което е повече от обема на работната сила у нас, чийто брой е под 4 млн. Приходната агенция има отношения и с граждани извън работната сила, но според официално разпространената информация, в масива се намират данни от база на НЗОК и Агенцията по заетостта например, което поставя под въпрос доколко НАП има правомощия да съхранява данни на други ведомства, включително и работещите в съвсем различни ресори (като гореспоменатите). Събирането на такава огромна база от данни с очевидно множество ключови характеристики дава потенциална възможност на

агенцията за класифициране на единици/субекти, за което стана въпрос в началото на този текст. Подобно класифициране също може да бъде направено в ущърб на гражданите, тъй като е възможно да бъде създаден модел на например, данъчно неблагонадеждни лица, които да търпят по-чести данъчни проверки, независимо че подобни модели все пак са базирани на екстраполации (и в този смисъл са възможни грешки и отклонения). Въпреки че понастоящем това е само потенциална възможност за използването на такива данни, ограничения спрямо подобно използване не стоят, *което би поставило гражданите в неизгодна позиция спрямо агенцията, а нея в положение в което тя притежава свръхконтрол над тях*.

Не по-малко обезпокоителен е и факта, че макар и данните да се отнасят до около 5 млн. лица, те надхвърлят 5 млн. реда, казано с други думи, за споменатите лица е събирано огромно количество информация. Според медийни публикации, в най-големия от всички разпространени файлове се съдържа информация с над 1,5 млн. реда, а файловете са общо над хиляда[14]. Отново тук е важно да се подчертае, че проблемът със съхраняването и възможностите за обработка на тази информация в логиката на големите данни е доста по-сериозен от наличието на самия теч. В системата на НАП съществува много повече информация (която е останала незасегната от хакерската атака) , която очевидно е събирана от най-различни ведомства и източници и която *отново може да бъде използвана за информационно следене на гражданите. Въпросът за наличието и съхранението на тази информация също е неясен и неуреден*.

Най-тревожният факт по отношение на изтеклите данни е наличието на информация за починали граждани в регистрите на НАП[15]. Само в изтеклите данни има такава за над 1 млн. починали. Това поставя изключително остри въпроси доколко подобно съхранение е допустимо по правилата на обществения договор изобщо. Контролирането на информация за покойници поставя под въпрос доколко агенцията изобщо зачита правото на прекъсване на отношенията с държавните институции, което би трябвало да последва акта на смъртта (респективно всички права и задължения на субекта се прехвърлят върху неговите наследници). В случая обаче съхраняването на тези данни в активно действащата система (а не в отделен архив) поставя покойниците в ролята им на реално взаимодействащи субекти с НАП, тъй като информационното общуване с тях (макар и едностранно) не е прекъснато. От една друга гледна точка това

означава тотален контрол върху субектите с помощта на информацията, който поне в модела на Бентам, цитиран по-горе прекъсва с излизането им извън паноптикума, но с помощта на достъпа до бази данни, с които агенцията разполага на практика този контрол се оказва непрестанен във времето и сам по себе си представлява най-големия потенциал за трансформация на обществените отношения с помощта на големите данни, макар и към един много неблагоприятен за гражданите развой.

Възможности за противодействие: вместо заключение

Направеният по-горе кратък анализ на казуса с изтеклите данни от НАП очертава няколко възможности за трансформация на обществените отношения, свързани с големите данни, които накратко казано, не носят повече информационна свобода, а повече възможности за тотален контрол на държавата върху гражданите чрез информацията която тя събира за тях. Противодействието на този тотален контрол е възможно чрез ясно определяне на границите, в които държавата има право да събира, но и съхранява и анализира големи данни, а също и времевия интервал, в който тя може да борави с тях. Не на последно място, е необходимо да се оцени и потенциала за съхраняване на неприкосновеността на тези данни възможностите за минимизиране на рисковете за гражданите при евентуално изтичане на такава информация.

БЕЛЕЖКИ

[1] Като Zamunda.net, най-големият български торент сървър.

[2] Въздействието на големите данни върху обществените отношения е комплексно, дори и само заради факта, че се използват не само в държавната администрация, но и за определяне на потребителскотоповедение, например, но тези въздействи излизат извън рамките на настоящия анализ.

[3] Анализът в детайли на законодателството за защита на личните данни излиза извън рамките на този текст, тук е важно да се подчертае само, че въпросът за съхраняване на лична информация на гражданите, включително в големи обеми е третирана в законодателната макрорамка.

[4] Вж. например тук при **Маринова, Е**. (2019) Борисов: Таванът на заплатите пречи да привлечем киберспециалисти в е-управлението, https://www.investor.bg/ikonomika-i-politika/332/a/borisov-tavanyt-na-zaplatite-prechi-da-privlechem-kiberspecialisti-v-e-upravlenieto-289485/

[5] Повече за срива Търговския регистър може да се види например при **Пейчева, Р**. (2019), Системата на Търговския регистър остава блокирана до понеделник , https://www.investor.bg/ikonomika-i-politika/332/a/sistemata-na-tyrgovskiia-registyr-ostava-blokirana-do-ponedelnik--266331/

[6] Вж.например при **Йорданова, В**. (2019), Изтекли ли са данните ми? НАП пусна приложението за проверка, https://www.dnes.bg/obshtestvo/2019/07/25/iztekli-li-sa-dannite-mi-nap-pusna-prilojenieto-za-proverka.417596

[7] Тази информация е станала достъпна през данните за данъчните декларации, някои от които се попълват и с помощта на този идентификационен код. **Кисьова, М** (2019), НАП най-сетне информира какви данни са изтекли, https://www.segabg.com/node/91329

[8] Пак там.

[9] Официална реакция относно съществуването на тази търсачка нямаше, което е обезпокоително, предвид че въпреки благородните намерения на това начинание, оперирането с данни, извлечени от агенцията по престъпен начин, макар и от трети лица, на практика е закононарушение.

[10] **Чобалигова, Б.** (2019), Горанов: Изтеклите данни от НАП не могат да навредят на гражданите, https://www.investor.bg/ikonomika-i-politika/332/a/goranov-izteklite-danni-ot-nap-ne-mogat-da-navrediat-na-grajdanite-286077/

[11] Това е директно упоменато в чл.1, ал.2. на основния закон на страната.

[12] **Шерифова, Р.** (2019), Спецпрокуратурата: Има достатъчно доказателства за кибертероризъм, извършван от "ТАД Груп" : https://dariknews.bg/novini/bylgariia/specprokuraturata-ima-dostatychno-dokazatelstva-za-kiberterorizym-izvyrshvan-ot-tad-grup-2179560

[13] Тук не се коментира доколко настоящия теч на данни изобщо може да бъде класифициран като кибер-тероризъм, а възможността за извършване на подобно престъпление по принцип.

[14] **Стоянов, Н.** (2019), От НАП са изтекли лични данни на милиони български граждани и фирми, https://www.capital.bg/politika_i_ikonomika/bulgaria/2019/07/15/3938624_ot_nap_sa_iztekli_lichni_danni_na_milioni_bulgarski/

[15] В различни медийни публикации това число варира между 1 и 2 млн., в официалния сайт с приложението на НАП е посочено че са малко над 1 млн.

ЛИТЕРАТУРА

Фуко, М. (1998). *Надзор и Наказание, Раждането на затвора.* С., УИ „Св. Кл. Охридски".

Конституция на Република България. https://www.parliament.bg/bg/const

Русо, Ж. Ж. (2018). За обществения договор, С., Лист.

Archer, E. M. (2014). Crossing the Rubicon: Understanding CyberTerrorism in the European Context, The European Legacy, Vol. 19, No. 5, 606–621.

Perlich, C., Dalessandro, B., Raeder, T. et al. (2014). Machine learning for targeted display advertising: transfer learning in action. *Machine learning.* Volume 95, Issue 1, pp 103–127.

Singh, S., Firdaus T., Sharma. **A. K.** (2015). Survey on Big Data Using Data Mining, *International Journal of Engineering, Development and Research*. Volume 3, Issue 4.