

Рамково споразумение за обмен на данни от преводачески памети, генерирани при превод на документи за целите на обществената администрация.

1. ВЪВЕДЕНИЕ.

След приемането на първия пакет правила относно повторната употреба на информация в обществения сектор през 2003 година, обемът генерирани данни се увеличава неимоверно, като се генерират и събират нови видове данни.

В същото време сме свидетели на едно постоянно развитие на технологиите за работа с езиков материал, целящи анализа, използването и обработката на получените данни.

Така описаното бързо развитие на технологиите позволява създаването на нови услуги и приложения на основата на данни и това как тези данни биват ползвани, събирани или комбинирани.

Големите потоци от данни (т.нар. „големи масиви от данни“), събирани чрез различни технологии или извлечени от големи количества информация в множество формати, могат да послужат за формирането на нови познания в редица сектори, но в същото време стават източници на проблеми, свързани със собствеността върху данните и последващата им употреба.

Обществените администрации, както и органите и учрежденията, които формират обществения сектор, генерират значително количество данни, които са от голяма полза за сектора на обработката на (естествен) език. Обществените администрации генерират неимоверно количество данни при изпълнение на задълженията си. Употребата на тази данни и последващото им ползване могат да

Рамково споразумение за обмен на данни от преводачески памети (FATDS) - данни, генерирани при превод на документи за целите на обществената администрация. ПРОЕКТ NEC TM

се окажат значими както за дадена обществена администрация, така и за горепосочения сектор.

Ако се съчетае с интелигентен анализ на данни, техники за обработка на големи масиви от данни и технологии за обработка на естествен език, употребата на тези данни от страна на обществените администрации може да улесни взимането на решения и да подобри ефективността на обществените политики.

Настоящото Рамково споразумение цели да послужи като препоръка към държавите членки относно прилагането на безплатни лицензи, които да премахнат правните пречки, и така посочените видове данни да се включат в процеса на формиране на нови познания.

Понастоящем тези лицензи (по-конкретно на PPDL, GPL и лицензите Creative Commons (CC), при които авторът се отказва от всички права) се ползват от държавни структури, корпорации и международни организации, които генерират обществени данни, като механизъм, благоприятстващ ползването и повторната употреба на големи количества данни.

2. Защо е необходимо рамково споразумение за управление на обмена на данни на европейско равнище?

Причини, пораждащи необходимост от Рамковото споразумение за обмен на данни от преводачески памети.

2.1. Общ подход.

Рамковото споразумение за обмен на данни от преводачески памети (FATDS) не е случайно, с чисто казионен или чрезвичаетен характер. Напротив, то представлява отговор на редица потребности. Обществените администрации в Европа изискват хиляди преводи ежедневно, а тези преводи генерират бази от данни, чието разпространение вече е регламентирано (в Директивата относно отворените данни).

Два са наистина важните моменти, които отразяват липсата на общ консенсус за адекватно регламентиране и управление на проблемите, свързани с обмена на данните, генерирани при преводи на текстове за целите на обществената администрация. Това е **един от секторите, които генерират най-голям обем данни** от подобен тип в хода на ежедневната си работа. В световен мащаб именно този сектор инвестира най-много в инфраструктура за съхранение на данни. В настоящия текст поставяме акцент върху данните, генерирани на база на преводаческите памети от преводи на документи - собственост на обществените администрации предвид нуждите на последните и множеството езици не само в рамките на дадена територия или страна, но и във всички държави - членки на ЕС.

Много важно е да има съответствие с препоръките за повторната употреба на данни и оперативната съвместимост в сектора на обществената администрация. Също толкова важно е да се изложат условията за повторната употреба и оперативната съвместимост.

Без съмнение, оптимално би било данните да са напълно достъпни и повторната им употреба да не изисква нарочни разрешения, каквито предвижда Директивата за отворените данни.

За да има действително достъпност на данните и повторна употреба обаче, следва да са изпълнени определени условия и поради това са необходими лицензи. Затова преводаческите памети следва да са с максимално отворен достъп. Наша приоритетна цел е да установим рамка за действие, която благоприятства по-нататъшното разпространение [на данните от памети] и капацитета за повторната им употреба.

Силно препоръчително е да не се създават допълнителни видове лицензи, а напротив - съществуващите видове да се опростят, както се посочва в *Програмата в областта на цифровите технологии* за Европа (Европейска комисия, 2010).

Понастоящем в световен мащаб съществуват две основни алтернативи в сферата на данните и информацията, които представляват най-често ползваните лицензи.

Рамково споразумение за обмен на данни от преводачески памет (FATDS) - данни, генерирани при превод на документи за целите на обществената администрация. ПРОЕКТ NEC TM

Чрез тях се регламентира и насърчава безплатния достъп до информация и употребата ѝ. Въпросните лицензи са Creative Commons (CC) от една страна и Apache 2, GPL и ODC от друга. Creative Commons се използва както за документи, съдържащи данни, така и за други документи, докато последните се ползват само в сферата на данните. Международните тенденции сочат, че при регламентиране на условията за достъп до данни и тяхната употреба се ползват отворени лицензи, като тези най-често са базирани на CC с известни различия в зависимост от характеристиките на портала за данни и страната, в която се прилагат.

Всичко това е в полза на администрацията, която все повече разчита на цифрови данни и е в позиция **да оцени въздействието на дейностите си, като взема предвид социалната възвращаемост на инвестициите**. Всъщност именно такава възвращаемост искат да постигнат инициативи като платформата „Национална и европейска централизирана преводаческа памет“ (NEC TM) посредством настоящото Рамково споразумение за обмен на данни от преводачески памет, генерирани при превод на документи за целите на обществената администрация.

Тъй като към момента не е постигнат консенсус относно регламентирането на обработката, употребата и споделянето на данни от преводачески памет, генерирани при превод на документи за целите на обществената администрация, е възможно някои да предложат въздържане от общи действия и позоваване на по-общо действащо законодателство във всяка държава членка вместо това. Една по-добра алтернатива обаче е задаването на конкретни правила под формата на регулаторния модел, предложен тук и основаващ се на „доброволни стандарти“, които са едностранни и практически насочени и не са законообвързващи, т.е. настоящото РАМКОВО СПОРАЗУМЕНИЕ.

FATDS взема предвид позициите и на двете страни в така описания процес: обществените администрации и изпълнителите. На първо място в Споразумението се пояснява, че проблемът при обмена на данни от преводачески памет действително съществува, че той стои пред заинтересованите страни и че е

необходима намеса, без да се стига до ексцесии и злоупотреби. На второ място Споразумението предлага една допълваща регулаторна рамка насочена към целевия тип данни, както и модел за систематично управление, който също е гъвкав и се адаптира според нуждите на всички администрации, частноправни субекти и държави членки.

С оглед горното е необходимо и горещо се препоръчва данните да се предоставят на определен за целта изпълнител или централизирано звено във всяка държава членка. Това е в интерес на обществото, което все повече поставя акцент върху данните, които произвежда. С изпълнението на всеки договор за предоставяне на преводачески услуги не само текста и документацията около превода му, но и преводаческата памет, генерирана в хода на превода, следва да бъдат предоставяни.

Освен това, тъй като базата данни NEC TM вече е пусната, са налице условия, които улесняват повторната употреба на подобни преводачески памет, като ги правят леснодостъпни и преди всичко препятстват налагането на технически и юридически ограничения, които биха направили повторната употреба трудна или невъзможна.

3. Многото аспекти на проблема при обмена на данни от преводачески памет, генерирани при превод за целите на обществената администрация: защитата, която предлага FATDS.

В строго юридически план, изпълнението на настоящото Рамково споразумение следва да се осъществи в съответствие с практиките и процедурите във всяка държава членка, като се отчитат както вече повсеместно разпространените практики, така и тези, непознати на изпълнителите в сектора на преводаческите услуги.

С оглед гореизложените съображения става ясно, че не е възможно настоящото Рамково споразумение да бъде пряко приложимо (сякаш е придобило статут на съответния вид законодателен акт на Общността, а източникът му не е Общността), но това не означава, че следва да се подценява ефективността му. Настоящото Споразумение следва да се спазва на чисто доброволен принцип и строго зависи от способността на подписалите го страни да обвържат съответните организации.

Следователно FATDS представлява автономно по своя характер споразумение и цели да се превърне в инструмент с практическа насоченост, т.е. в набор от насоки, в които се посочват действащите правила за управлението, централизирането и обмена на данни. Самото Рамково споразумение единствено следва да допълва вече съществуващото европейско и национално законодателство.

В този смисъл едно споразумение като FATDS, което не е законообвързващо, (може да) представлява полезен инструмент при конкретизирането и поясняването на понятието за обмен на данни от преводачески памети, генерирани в хода на работата на оферентите/изпълнителите.

От горното става ясно, че функцията на FATDS е да разясни и изрично да разпише правото на обществените администрации да изискват данните, генерирани при изпълнението на договори за преводачески услуги и собствеността на администрациите върху оригиналите и преводите (с оглед статута на последните - услуга, изпълнена по договор на основата на оригинал), като отрази консенсуса на участниците в процеса. За участници се считат дружествата оференти (изпълнители), административните органи в държавите членки и централното звено на европейско равнище, което следва да се определи за целта.

В този смисъл следва да се припомни, че получената най-много подкрепа версия на този модел (т.нар. „парадигма за правната стойност“) подчертава необходимостта от инструменти, които макар и да не са представители на традиционните видове законодателни актове, все пак не могат да бъдат изключени от правния мир и имат задължителен характер. Като подписват такива

инструменти, страните по тях строго се ангажират с изпълнението на своите договори за услуги в рамките на своята компетентност по регламентирана за целта процедура, като действително прилагат инструмента, който вече има минимална или относителна правна стойност.

В основата си FATDS поставя акцент най-вече върху задаването на модели и насоки и така се причислява към подобни инструменти.

В крайна сметка моделът на регулация на FATDS е съобразен с новите подходи и методи за регулация и управление, които се предлагат напоследък (в това число от ЕС, който все повече разчита на актове, които осигуряват преминаването на правната норма в практика, вместо да установява още правила). В Споразумението се дават определения на достъп до информация, съдържаща се в базите данни, чието управление е регламентирано в законодателството, както и посредством договори и/или лицензионни споразумения, нейната употреба и възможната ѝ повторна употреба.

Публикуването на Съобщението относно отворените данни и предложените от Комисията изменения са част от една политика, целяща повече отворени данни и подкрепа за пазара на информационни услуги. В основата на тази политика стои приемането на Директива 2003/98/ЕО на Европейския парламент и на Съвета от 17 ноември 2003 година относно повторната употреба на информацията в общественения сектор.

Тук трябва да се спомене и преработеният текст на Директивата относно повторната употреба на информацията в обществения сектор от края на 2011 година, който е част от Програмата в областта на цифровите технологии за Европа - инициатива, предприета от Комисията (2010). Програмата цели да насърчи интернет услугите в ЕС, като нейни приоритети са повечето отворени данни, опростяването на системата за получаване на лицензи за обмен на съдържание и изпълнението на стандартите за оперативна съвместимост.

5. Възможността за регулация, която предоставя FATDS

Обществените данни, съхранявани от обществени органи в ЕС попадат под конкретен режим съгласно директивите относно повторната употреба на информацията в обществения сектор от 2003 и 2013 година. Директивите регламентират предоставянето на документите на обществените администрации за повторна употреба: „всяко съдържание, независимо от носителя му (хартиен или съхраняван в електронна форма, или като звукозапис, видеозапис, или аудиовизуален запис) [...], притежание на институциите в обществения сектор, за търговски или нетърговски цели“. Следователно общият принцип изисква свободен достъп до подобно съдържание, а ако се налагат такси, „тези такси са ограничени до пределните разходи, извършени за тяхното възпроизвеждане, предоставяне и разпространение“ (член 6, параграф 1).

С други думи FATDS има характер на препоръка, която има минимална или относителна правна стойност и се основава на съществуващи договори за предоставяне на услуги и нарочно генерирани данни (преводачески памети). Тази препоръка представя практическите измерения на задължението за обмен на данни под формата на преводачески памети, генерирани при изпълнение на преводачески услуги за целите на обществените администрации, независимо от средата и формата на данните.

6. Съдържание на Рамковото споразумение - системата за управление на процеса на пренос на данни от преводачески памети, която предвижда Споразумението.

Преводаческите памети представляват бази данни, съдържащи езикова информация и постоянно осъвременявани с преводи от специалисти, за да могат да се ползват в бъдеще с оглед осигуряване на по-добра терминологична и стилова последователност, както и с цел намаляване на разходите в резултат от пълни или

частични съвпадения между нови текстове и съществуващи, вече завършени преводи.

Преводаческите памети са цифрови съвкупности от данни, съставени от сегменти текст на езика на оригинала, паралелно с които вървят преводи на тези сегменти на други езици. Оригиналът и превода (преводите) могат да вървят паралелно и когато са разделени на преводни единици. Преводните единици, съхранявани паралелно с еквивалентите си, могат да бъдат различни (фрази, абзаци, група/групи от думи и т.н.). При преводаческия софтуер най-често по подразбиране сегментирането следва препинателен знак, с който се бележи края на изречение, или знака за край на абзац.

Основната функция на преводаческите памети (по-долу „ПП“) се състои в извличането на пълни или частични съвпадения, които се предлагат при превода на изречения, и терминологични съответствия. В хода на даден превод сегментите на езика на оригинала подлежат на търсене в базата данни на ПП. Ако в ПП фигурира сегмент на езика на оригинала, който представлява пълно или частично съвпадение, този сегмент се предлага на преводача, като паралелно с него върви и превод, извлечен от базата данни, и всякаква допълнителна информация, съхранена заедно с конкретния сегмент в базата данни. Преводаческият софтуер (т.нар. „CAT tools“, от английското „computer-assisted translation tools“) посочва и степента на съвпадение („непълни съвпадения“, както са познати в преводаческите среди). При превод такива съвпадения са полезни за специалистите по език само ако стойността им е 50 - 65 % или по-висока, като се залагат конкретни тарифи на база труда, необходим, за да се преведат напълно нови фрази и такива, за които паметта предлага поне частични съвпадения (в зависимост от съответния процент) в същия контекст. Съвпадение от 100 % се счита за пълно съвпадение между изречение в преведения текст и изречение в базата данни.

Общодостъпният формат за обмен на преводачески памети е TMX (Translation Memory eXchange), като обикновено се ползва версия 1.4b. Това е стандарт на XML от типа DTD (*document type definition*). Създаден е от комисията OSCAR (Open Standards for Container/Content Allowing Re-use).

Като използват формата TMX, физическите и юридическите лица са в по-добра възможност да си сътрудничат в хода на дадени преводачески проекти. Освен това форматът TMX улеснява преминаването от един преводачески софтуер към друг, което благоприятства конкуренцията между предлаганите технологии и тяхното постоянно развитие с цел да предлагат нещо повече от конкурентите си. Подобно на други обществено достъпни стандарти този формат е разработен с цел намаляване на проблемите при съвместимостта, насърчаване на повторната употреба на езиковите ресурси, опростяване обмена на данни и стимулиране на технологичните иновации.

Настоящото Рамково споразумение цели да установи една „рамка от добри практики“ (както я окачествихме по-рано) за събирането на данни, генерирани при изпълнение на договори за предоставяне на преводачески услуги на обществените администрации, организирането на тази данни чрез използване на софтуерната система NEC TM и инициативите за централизиране на данни, определени от Европейската комисия, с оглед приемането на Протокол за централизирането на двуезични данни от обществените администрации, използване на данните в полза на широката общественост и на самите държави членки, подпомагане работата по създаването на корпус с големи масиви данни във всяка държава членка и обмен на елементите от него, които всяка държавна администрация счита за относими на европейско равнище.

За да се постигне горното, централизирането на данните, съхранявани посредством държавна компютърна инфраструктура, следва да се управлява от компетентен орган в съответната държава членка, например конкретното държавно звено за цифрови иновации или компетентното министерство. При подписване на настоящото Рамково споразумение и по преценка на компетентния орган, тези данни могат да се съхраняват и на по-високо равнище, т.е. от съответното звено в Европейската комисия (например ELRC-Share).

Предвид горното, първата стъпка следва да бъде централизирането на данните от преводаческите памети, генерирани при изпълнение на преводачески услуги за

обществените администрации на национално ниво, след приемането на настоящото Споразумение. След това, в съответствие със Споразумението, данните следва да се обменят посредством общеевропейски орган, за да бъдат от полза за обществените администрации и другите държави членки. Ползването на софтуера NEC TM (National and European Central Translation Memory) следва да осигурява връзка за получаване и предоставяне на данни на администрациите, които ще имат възможност да съхраняват своите преводачески памет поотделно на централен сървър и да ги обменят с тези, които желаят да повишат ефективността и да намалят разходите в хода на работата си (обикновено това са щатни преводачи и външни изпълнители), на национално равнище, да създават големи масиви от данни на същото равнище и да подбират данните, които им трябва, сред тези, налични на централния сървър, на който се съхранява европейската централизирана преводаческа памет.

Отношения с частноправни субекти и предоставяне на преводачески памет

Преводаческата дейност, извършвана от дружества оференти (изпълнители), следва да включва и представяне на преводаческите памет, генерирани от последните в хода на предоставяне на услугата им. За целта в обявление, публикувано в Държавен вестник на съответната държава членка или в официалния вестник на областта или общината, както и в съответния профил на купувача, следва да се бъде включена клауза за FATDS за NEC TM, в която да се посочва CPV кодът и която да гласи, че ДАННИТЕ ОТ ПРЕВОДАЧЕСКАТА ПАМЕТ СЛЕДВА ДА БЪДАТ КАЧВАНИ НА СЪРВЪРА НА КОМПЕТЕНТНИЯ НАЦИОНАЛЕН ОРГАН, за да е възможно в бъдеще, ако последният прецени, тези данни да бъдат обменяни с определеното за целта звено в Европейската комисия (например ELRC-Share или подобна инициатива).

С оглед горното, дружествата оференти (изпълнители) следва да представят преводаческите памет, генерирани в хода на работата си, като настоящото Рамково споразумение не важи със задна дата и задължителният му характер е в сила от момента на присъединяването към него и подписването под горепосочената клауза, с което изпълнителите се задължават да представят

паралелните данни, генерирани в хода на превода (във формат TMX или друг съвместим с него формат), наред със самия превод.

7. ИНТЕЛЕКТУАЛНА СОБСТВЕНОСТ И УСЛОВИЯ ЗА УПОТРЕБА НА БАЗИТЕ ДАННИ

По отношение на интелектуалната собственост няма съмнение, че обществената администрация има пълни и изключителни права на собственост върху данните, генерирани в хода на превода и в изпълнение на договора за предоставяне на преводачески услуги. Администрацията е и автор на преведения текст и следи за спазването на правата, свързани с него.

Не може да се пренебрегне фактът, че доколкото в повечето случаи информацията води началото си от обществените администрации, следва да се обръща внимание и на спазването на изискванията за повторна употреба на информацията от общественения сектор и превръщането ѝ в отворени данни.

Следователно за да му бъде възложена поръчка, оферентът (изпълнителят) следва да предостави за неограничен срок и без териториални ограничения изключителните права върху всички документи и генерирани данни (независимо от софтуера, необходим за тях, или формата им). В определението на прехвърляните права за възпроизвеждане попадат и разпространението и преформатирането.

Това означава, че данните, получени от такива преводачески памети, не могат да се използват от изпълнителя с користна цел, а само за да се получи информация, ако това е необходимо за конкретната преводаческа задача.

Интелектуална собственост и условия за употреба на базите данни.

Базата данни с преводачески памети на ГД „Писмени преводи“ остава изключителна собственост на Европейската комисия. За периода, в който правата на Комисията за защитени, тя предоставя безплатно и без териториални ограничения на лица, извършващи повторната употреба на данни, неизключителното правото да употребяват данните по какъвто и да е начин, съобразен с условията, заложиени в Решение на Комисията от 12 декември 2011 година относно повторната употреба на документи на Комисията (стр. 39 до 42 от ОВ L 330, 14.12.2011).

Всяка повторна употреба на базата данни или структурираните елементи, съдържащи се в нея, следва да се посочи от лицето, извършващо повторна употреба, като се цитира източникът на използвания документ (адрес на уебсайта, дата на последното изменение и фактът, че Европейската комисия запазва правата си на собственост върху данните).

С оглед горното, тази база данни е оптимална за първоначалното попълване на NEC TM на всяка държава членка.

8. ЗАЩИТА НА ЛИЧНИТЕ ДАННИ. ПРИЛОЖЕНИЕ НА РЕГЛАМЕНТ (ЕС) 2016/679 И РЕГЛАМЕНТ (ЕС) 2018/1725

По отношение на личните данни, които се обработват от страна на изпълнителя и обществената администрация в изпълнение на дадена поръчка, двете страни се задължават да спазват разпоредбите на общите регламенти за защита на личните данни, цитирани по-долу.

[Изисква се спазване на разпоредбите на] Регламент (ЕС) 2016/679 на Европейския парламент и на Съвета от 27 април 2016 година, както и на действащото национално законодателство в областта на защитата на личните данни във всяка от държавите, подписали Рамковото споразумение, и на Регламент [(ЕС) 2018/1725] на Европейския парламент и на Съвета от 23 октомври 2018 година относно защитата на физическите лица във връзка с обработването на лични данни от институциите, органите, службите и агенциите на Съюза и относно свободното движение на такива данни.

ЗАКЛЮЧЕНИЕ.

Обществените администрации създават общи платформи за обработка на естествен език и автоматизиран превод и развиват ресурси за повторната употреба на информация за целите на обществения сектор. Те са сред двигателите на сектора за езикови технологии и като такива са длъжни да разработват политики по обмен на данните и да поставят основите за ефективен обмен, който да се реализира на практика. В допълнение те следва да задължат другите участници в процеса да се ангажират с подписване на клаузата от договора за възлагане на поръчка, според която от тях се изисква да предоставят данните, генерирани при превода (преводаческите памети), наред със самия превод, като по този начин създават набор от безценни езикови ресурси.

ИЗПОЛЗВАНА ЛИТЕРАТУРА: Ферер-Сапена и други (2011) в статията им относно достъпа до обществени данни; Ферер-Сапена и Песет (2012) относно повторната употреба на културни данни или Рамос Симон и съавт. (2012) в проучването им на европейските портали за данни.

Предоставяне на лиценз за отворени данни: Практически наръчник на Съвета за финансиране на висшето образование в Англия (Корн и Опенхайм, 2011). Следва да се отбележат и Наръчник относно предоставянето на лицензи за отворени данни (Open Knowledge Foundation) и насоките, водещи началото си от Програмата в областта на цифровите технологии за Европа (Европейска комисия, 2010).