

Rámcová smlouva o sdílení dat z překladové paměti generovaných v rámci překladů dokumentů veřejné správy.

1. ÚVOD.

Od přijetí prvního souboru pravidel pro opakované použití informací z veřejného sektoru v roce 2003 celosvětově objem dat exponenciálně vzrostl a dochází k vytváření a shromažďování nových typů dat.

Současně jsme svědky neustálého vývoje technologií zpracování jazyka pro analýzu, využití a zpracování dat.

Rychlý technologický vývoj umožňuje vytvoření nových služeb a aplikací založených na využití, agregaci a kombinaci dat.

Velká množství dat nebo tzv. „big data“ shromažďovaná pomocí různých technologických nástrojů nebo získávaná z rozsáhlého sběru informací v různých formátech mohou přinášet nové znalosti v nejrůznějších odvětvích, ale také problémy jak z hlediska vlastnictví údajů, tak z hlediska jejich následného využití.

Orgány veřejné správy a orgány a subjekty, které tvoří veřejný sektor, jsou hlavními původci dat, která nacházejí velké využití v odvětví zpracování (přirozeného) jazyka. Orgány veřejné správy generují při plnění svých povinností obrovské množství dat. Využití a vytěžení těchto dat může být jak pro orgány veřejné správy, tak pro relevantní odvětví velmi zajímavé.

Využití těchto dat ze strany veřejné správy v kombinaci s těžbou dat (tzv. „data mining“) a technologiemi velkých dat (big data) a zpracování přirozeného jazyka může pomoci při veřejném rozhodování a při zefektivnění veřejného života.

Tato rámcová smlouva by měla plnit roli doporučení členským státům ohledně implementace otevřených licencí, které odstraňují možné právní překážky s cílem začlenit tyto typy dat do vytváření nových poznatků.

V současné době je pro různé vlády, korporace a mezinárodní organizace, které jsou původci veřejných dat, běžné, že se spoléhají na tyto licence, zejména na PDDL, GPL a Creative Commons (CC), v nichž subjekty nemají práva k databázím, což představuje vhodný mechanismus, který podporuje užití a opětovné využití velkých objemů dat.

2. Proč Evropská rámcová smlouva o řízení sdílení dat.

Důvody rámcové smlouvy o sdílení dat z překladových pamětí.

2.1. Obecný přístup.

Smlouva FATDS (Rámcová smlouva o sdílení TMX dat) není náhodná ani pouze jednorázová či vztažená k jednomu bodu. Naopak reaguje na potřeby, které plynou z různých motivací. Evropská veřejná správa generuje každý den tisícovky překladů a v rámci těchto překladů vznikají databáze, jejichž vydání jako dat je již v současné době zákonnou povinností (Evropská směrnice o otevřených datech).

Vysoce relevantní jsou zejména dva body, které odrážejí nedostatečný konsenzus ohledně adekvátní regulace a zvládnání problémů týkajících se sdílení dat generovaných překlady textů z veřejné správy. Jde o **jeden ze sektorů, který generuje nejvíce dat** tohoto druh v každodenním provozu. Jde také o sektor, který nejvíce investuje do infrastruktury na uchování dat. Soustředíme se zejména na data generovaná překladovými pamětmi dokumentů, jejichž vlastníkem je veřejná správa, s přihlédnutím k potřebám spravovaných osob a k množství jazyků nejen na určitém území nebo v konkrétní zemi, ale ve všech zemích EU.

Je velice důležité řídit se doporučeními ohledně opětovného využití dat a interoperability sektoru veřejné správy. Stejně důležité je stanovení podmínek pro opětovné využití a interoperabilitu.

Bezpochyby by bylo optimální, kdyby byla data plně přístupná a kdyby pro jejich opakované využití nebyla vyžadována zvláštní povolení, jak stanovuje evropská směrnice o otevřených datech.

Aby se však dostupnost a opakované využití mohly stát skutečností, musí být splněny určité podmínky, což je důvod, proč jsou třeba licence – v tomto případě by měly být překladové paměti co nejotevřenější. Naším hlavním cílem je stanovit rámec, který podporuje větší šíření a možnost opětovného použití těchto dat.

Je více než záhodné negenerovat více typů licencí, naopak je třeba co nejvíce jejich typy zjednodušit, jak doporučuje Evropská digitální agenda (Evropská komise, 2010).

V současné době existují dvě hlavní mezinárodní alternativy, které podporují využití licencí v kontextu dat a informací a které regulují a podporují volný přístup a využití informací. Těmito licencemi jsou Creative Commons (CC), Apache 2 a GPL a ODC. Creative Commons se týká jak dat, tak dalších dokumentů, zatímco ostatní výše jmenované licence se týkají jen oblasti dat. Mezinárodní trend naznačuje, že k regulaci podmínek přístupu a využití dat se používají modely otevřených licencí založených zejména na CC, byť s variacemi přizpůsobenými charakteristikám jednotlivých datových portálů a každé země, kde jsou využívány.

Vše výše uvedené je ku prospěchu stále více se digitalizující veřejné správě, která je takto schopná **měřit dopad svých kroků z hlediska sociální návratnosti investic**, což je cíl iniciativ, jako je například Národní a evropská centrála překladových pamětí (NECTM) prostřednictvím této rámcové smlouvy o sdílení dat generovaných překladovými paměťmi nebo textů překládaných pro veřejnou správu.

Jelikož v současné době neexistuje konsenzus o tom, jak regulovat zpracování, využití a sdílení dat z překladových pamětí vytvářených v důsledku publikačních potřeb veřejné správy je jednou možností zdrženlivá varianta s odkazem na stávající obecné zákony v

každém členském státě. Lepší alternativou je samostatný právní předpis pro tuto oblast v podobě zde navrhovaného regulatorního modelu založeného na „měkkých standardech“ jednostranné, nezávazné povahy a praktickém přístupu, jako je RÁMCOVÁ SMLOUVA.

FATDS zohledňuje obě strany zapojené v procesu, veřejnou správu a dodavatele. Nejprve jasně konstatuje, že problém se sdílením dat z překladových pamětí existuje, jde o reálný problém pro zúčastněné instituce a že je nezbytné zasáhnout při současném odmítnutí excesů a zneužití. Zadruhé FATDS navrhuje komplementární a specifický rámec pro regulaci a rovněž systematické řízení, které je zároveň flexibilní, přijatelné pro všechny orgány veřejné správy, soukromé podniky a členské státy.

Je proto silně doporučeníhodné a nezbytné v zájmu společnosti, která se stále více zajímá o hodnotu dat, která generuje, aby data byla zpřístupněna každému odpovědnému dodavateli nebo centrálnímu orgánu v každém členském státě u každé smlouvy na dodávku překladatelských služeb, která bude plněna, přičemž zpřístupněny nebudou jen texty nebo dokumenty daného překladu, ale rovněž překladová paměť vytvořená jako výsledek procesu překladu.

Obdobně jsou tedy díky implementaci databáze NEC TM ustaveny podmínky, které usnadní opětovné využití překladových pamětí díky jejich dostupnosti, a především nebudou podléhat technickým nebo právním omezením, které omezují nebo zabraňují tomuto opětovnému použití.

3. Vícerozměrnost problému sdílení dat z překladových pamětí generovaných veřejnou správou: ochrana poskytovaná FATDS.

V přísně právním smyslu je tato rámcová smlouva zamýšlena jako smlouva, k jejíž realizaci dojde v souladu s praxí a procedurami každého členského státu a se zohledněním postupů, které jsou již v současnosti běžné a s nimiž jsou dodavatelé v oboru překladatelství obeznámeni.

Z výše uvedených úvah vyplývá, že není možné přiznat předmětným rámcovým smlouvám přímou právní účinnost, jako by šlo o další součást legislativy Společenství, v tomto případě pocházející z autonomního zdroje, ale rovněž není možné jejich účinnost podcenit. Dodržování smlouvy je čistě dobrovolné a zcela závisí na pravomoci signatářů zavazovat dané organizace.

FATDS tedy, pochopená jako autonomní smlouva, usiluje o to stát se spíše praktickým nástrojem, tj. souborem pokynů, které uvádějí účinná pravidla pro řízení, centralizaci a sdílení dat, kdy rámcová smlouva pouze doplňuje Společenství a národní legislativu.

Takto může nenormativní smlouva, jakou je FATDS, představovat a tímto i představuje užitečný nástroj korekce a vyjasnění výkladu sdílení dat generovaných překladovými paměťmi vytvořenými jako výsledek práce uchazečů/dodavatelů.

Funkcí FATDS je tedy vyjasnit a specifikovat právo orgánů veřejné správy požadovat veškerá data generovaná v rámci smluv o dodávkách překladů, vlastnictví originálních textů ze strany těchto orgánů, překlad jako smluvní službu odvozenou od originálu, která zohledňuje dohodu zúčastněných aktérů, jimiž se na jedné straně rozumí účastníci výběrových řízení (dodavatelé) a na druhé straně správní orgán daného členského státu a konečně pak centrální orgán EU, který bude pro tento účel určen.

V tomto smyslu je třeba mít na paměti, že nejzavedenější podoba tohoto modelu či paradigmatu právní závaznosti osvětluje existenci nástrojů, které, aniž by odpovídaly tradičním typům právních norem, nemohou ani být ze světa práva vyloučeny a nemůže jim tedy být upřena závazná povaha, a jejichž smyslem je vyjádřit pevný závazek subjektů, které jsou signatáři a které plní smlouvu o poskytování služeb, vykonávat své kompetence dle právně předvídatelné procedury, čímž uvádějí do praxe nástroj s minimální či relativní právní závazností.

Základní obsah FATDS se soustředí spíše na ustavení modelů a pokynů, které tvoří typický nástroj reflexivního práva.

V konečném důsledku je toto paradigma regulace, které představuje FATDS, obsaženo v nových přístupech a nových metodách regulace a postupech které jsou v současnosti navrhovány, a to rovněž ze strany EU, která se spoléhá spíše na nástroje, které uvádějí právní předpis do praxe než na vytváření dalších pravidel, v našem případě definujících přístup, využití a možné opětovné využití informací obsažených v zákonem regulovaných databázích a informací na základě smluv nebo licencí.

V rámci EHS tvoří politiku otevírání dat a podpory trhu s informacemi publikace informací o otevřených datech a návrh Evropské komise na změnu, které mají svůj centrální bod ve schválení směrnice 2003/98/ES Evropského parlamentu a rady ze dne 17. listopadu 2003 o opakovaném použití informací veřejného sektoru.

K tomuto musíme dodat pozměňovací návrh směrnice o opakovaném použití informací veřejného sektoru z konce roku 2011, který tvoří součást Evropské digitální agendy, což je iniciativa Evropské komise (2010). Cílem této agendy je podpora online služeb v Unii a její prioritou je otevření veřejných dat k opakovanému použití, zjednodušení licenčního systému pro sdílení obsahu a zavedení standardů interoperability.

5. Regulatorní možnost FATDS

Veřejná data, kterými disponují veřejné orgány v Evropské unii, podléhají zvláštnímu zacházení na základě směrnic o opakovaném použití dat veřejného sektoru z roku 2003 a 2013, které orgánům veřejné správy ukládají povinnost zpřístupnit k opakovanému použití „jakýkoliv obsah nehledě na médium (psaný na papíře nebo uchovávaný v elektronické podobě jako zvuková, obrazová, nebo audiovizuální nahrávka) uchovávaný orgány veřejné správy pro komerční nebo nekomerční účely“. Obecným principem je tedy volná dostupnost tohoto obsahu, tj. pokud je uplatněn nějaký poplatek, „měl by se omezit na marginální náklady vzniklé v souvislosti s reprodukcí, zpřístupněním a šířením“ (čl. 6.1).

Ve zkratce se tedy FATDS ukazuje být v zásadě „doporučením“ s minimální nebo relevantní právní závazností, založeným na stávajících smlouvách o poskytování služeb a datech (překladových pamětech) generovaných ex profeso. Toto „doporučení“ představuje praktickou specifikaci povinnosti sdílet data generovaná ve formě překladových pamětí v rámci překladatelských služeb poskytovaných orgánům veřejné správy, ať už je jejich médium či formát jakýkoliv.

6. Obsah rámcové smlouvy: systém řízení pro přenos dat z překladových pamětí podporovaných touto smlouvou.

Překladová paměť je jazyková paměť, která průběžně ukládá paralelní překlady, které vytvářejí profesionální překladatelé, aby tyto překlady mohli v budoucnu používat s cílem dosáhnout větší terminologické a stylistické konzistentnosti a rovněž úspor díky úplným či částečným shodám mezi novými texty a starými již dokončenými překlady.

Překladové paměti jsou digitální repozitáře skládající se z řádků textu z obsahu v originálním jazyce spojených s překlady v jiných jazycích. Tyto texty mohou být rovněž efektivně spojeny pomocí překladových jednotek. Překladové jednotky, které jsou uloženy spolu s jejich ekvivalenty, jsou definovány různými způsoby (podle fráze, odstavce, slova nebo skupiny slov atd.), přičemž nejčastější formou oddělení v počítačem asistovaných překladových systémech je segmentace po interpunkčním znaménku na konci věty nebo odstavce.

Hlavní funkcí překladových pamětí (dále jako PM) je získávat úplné nebo částečné návrhy nebo shody z věty nebo konkordance pojmů. Během překladu jsou prohledávány v databázi PM segmenty zdrojového jazyka. Pokud je v PM segment ve zdrojovém jazyce, který se přesně nebo částečně shoduje, tento segment bude překladateli navržen spolu s překladem získaným z této databáze a s jakýmkoliv dalšími informacemi, které byly spolu se segmentem do databáze uloženy. Nástroje počítačem asistovaného překladu (nástroje CAT) ukazují stupeň podobnosti (kterému se v oboru říká *fuzzy shoda*). V

překladatelském oboru mohou být tyto shody pro překladatele užitečné od 50 až-65 % výše a cenové škály jsou také stanoveny v závislosti na úsilí vynaloženém na překlad zcela nových frází nebo jiných úseků, u nichž existují obdobné návrhy nebo procentuální shody v témže kontextu. 100% shoda je považována za úplnou shodu mezi požadovanou větou a toutéž větou v databázi.

Volně dostupným formátem pro výměnu překladových pamětí je formát TMX (Translation Memory eXchange), obecně ve verzi 1.4b. Jde o XML standard typu DTD (*definice typu dokumentu*). Byl vytvořen výborem OSCAR (Otevřené standardy pro schránky/obsah umožňující opakované požití).

Díky aplikaci formátu TMX je pro osoby nebo společnosti snazší spolupracovat na překladatelských projektech. Formát TMX rovněž usnadňuje migraci z jednoho počítačově asistovaného překladatelského systému na jiný, což podporuje konkurenci mezi nabízenými technologiemi a jejich neustálý rozvoj ve snaze odlišit se od konkurentů. Stejně jako jiné otevřené standardy byl tento formát vyvinut se zohledněním problému kompatibility, což ve výsledku podporuje opakované použití lingvistických zdrojů, usnadňuje výměnu a stimuluje tak technologickou inovaci.

Cílem této rámcové smlouvy je stanovit, jak jsme již důkladně definovali, „rámec dobré praxe“ pro získávání dat na základě smluv o poskytování předkladatelských služeb s evropskými orgány veřejné správy, organizaci těchto dat díky implementaci NEC, softwarového systému PM a iniciativ Evropské komise v oblasti centralizace dat s ohledem na přijetí Protokolu pro centralizaci dvojjazyčných dat orgány veřejné správy, jeho použití ve prospěch společnosti jako celku a samotných členských států, pomoc při vytváření korpusu národních velkých dat (big data) a sdílení částí, které jednotlivé národní orgány veřejné správy považují za relevantní na evropské úrovni.

V tomto směru bude centralizace dat uložených v národní počítačové infrastruktuře řízeno příslušným státním orgánem – například konkrétním státním úřadem pro podporu digitalizace nebo příslušným ministerstvem daného členského státu. Tyto údaje lze sdílet

v případě přistoupení k této rámcové smlouvě, a pokud příslušný orgán takto rozhodne, na vyšší úrovni s příslušným orgánem Evropské komise (např. ELRC-Share).

S ohledem na právě uvedené bude prvním krokem centralizace dat z překladových pamětí generovaných orgány veřejné správy na národní úrovni pomocí přijetí této smlouvy a nadále ve shodě s ní a sdílení těchto dat s celoevropským orgánem ve prospěch orgánů veřejné správy ostatních členských států EU. Využití softwaru „Národní a evropské centrální překladové paměti“ (NEC TM) poskytne vstupní/výstupní (I/O) připojení pro správce, kteří budou moci soukromě ukládat své překladové paměti na centrální server a sdílet je s těmi, kdo se rozhodnou zefektivnit svou práci a ušetřit náklady na celostátní úrovni (typicky interní překladatelé nebo externí dodavatelé), vytvářet národní velká data a vybírat data, která chtějí sdílet s centrálním serverem, kde je uložena evropská centrální překladová paměť.

Vztah se soukromými subjekty a zadání TMX

Překladatelská práce vykonávaná uchazeči ve výběrových řízeních (dodavatelé) bude zahrnovat poskytnutí překladových pamětí, které vygenerují v rámci poskytovaných překladatelských služeb. Za tímto účelem by měl být v oznámení publikovaném v oficiálním věstníku členského státu, regionu nebo obce, tj. na portále zadavatele, obsažen článek odkazující na NEC, TM, FATDS obsahující KÓD CPV, který určí, že DATA Z PŘEKLADOVÉ PAMĚTI BUDOU NAHRÁNA NA SERVER PŘÍSLUŠNÉHO VEŘEJNÉHO ORGÁNU za účelem jejich budoucího sdílení (pokud to příslušný národní orgán uzná za vhodné) s orgánem Evropské komise, který bude teprve určen (např. ELRC-Share nebo obdobná iniciativa).

Za tímto účelem dodají uchazeči ve výběrových řízeních (dodavatelé) překladové paměti generované spolu se svou prací, přičemž uplatní princip ne-retroaktivity, což znamená, že tato rámcová smlouva nesmí mít zpětnou účinnost a že její závazná povaha bude působit pouze od okamžiku připojení a podpisu článku, kterým se zavazují dodat, spolu s provedenou prací, paralelní data generovaná svými překlady (v TMX nebo obdobném kompatibilním formátu).

7. DUŠEVNÍ VLASTNICTVÍ A PODMÍNKY UŽITÍ DATABÁZE

Co se týká duševního vlastnictví, nemůže být pochyb, že data generovaná při výkonu práce v souladu se smlouvou, na jejímž základě byl daný překlad zadán, náleží cele a výhradně danému orgánu veřejné správy, který je autorem a správcem původního textu, jenž je předmětem překladu.

Tuto skutečnost nelze ignorovat, nakolik většinou dané informace pocházejí od orgánů veřejné správy; je třeba dbát také na opakované použití veřejných informací a otevírání veřejných dat.

Úspěšný uchazeč (dodavatel) tedy postoupí výhradní a územně neomezená práva k jakémukoliv typu dokumentace nebo generovaných dat neohledně na jejich podporu nebo formát, s porozuměním distribuci a transformaci v tomto převodu práv na reprodukci.

V důsledku toho data generovaná těmito překladovými paměťmi nesmí být vítězným uchazečem užívána k žádnému výdělečnému účelu a lze je v tomto případě použít jen v rámci vykonávané práce.

Duševní vlastnictví a podmínky užití databáze.

Databáze DGT-TM je výhradním vlastnictvím Evropské komise. Komise poskytuje zdarma a celosvětově po dobu ochrany těchto práv svá nevýhradní práva opětovným uživatelům ke všem druhům použití, které splňují podmínky uvedené v rozhodnutí Komise z 12. prosince 2011 o opětovném použití dokumentů Komise, zveřejněném v oficiálním věstníku Evropské unie L 330 z 14. prosince 2011, strany 39 až 42.

Jakékoliv nové použití databáze nebo strukturních prvků v ní obsažených musí být opětovným uživatelem identifikováno a tento uživatel je povinen uvést zdroj použitých dokumentů: internetové stránky, datum a poslední aktualizaci a skutečnost, že Evropská komise zůstává vlastníkem dat.

Tato databáze je tedy optimální pro úvodní zaplnění verze NEC TM členského státu.

8. OCHRANA ÚDAJŮ. UPLATNĚNÍ NAŘÍZENÍ EU 2016/679 A NAŘÍZENÍ 2018/1725

V souvislosti s osobními údaji, které mají být zpracovány na základě plnění smlouvy orgánem veřejné správy a úspěšným uchazečem, jsou obě strany povinny dodržovat níže uvedené obecné nařízení o ochraně osobních údajů.

Nařízení (EU) 2016/679 Evropského parlamentu a Rady z 27. dubna 2016 a rovněž dodržování stávající národní legislativy o ochraně osobních údajů v každém členském státě, který přijme tuto rámcovou smlouvu a rovněž nařízení Evropského parlamentu a Rady z 23. října 2018 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů orgány, institucemi a jinými subjekty Unie a o volném pohybu těchto údajů.

ZÁVĚRY.

Státní správa jako podporovatel odvětví jazykových technologií je s vytvořením společných platforem pro zpracování přirozeného jazyka a automatický překlad a rozvojem zdrojů pro opakované použití informací veřejného sektoru (RISP) povinna rozvíjet zásady sdílení dat a stanovit základy pro toto sdílení tak, aby byly reálné a účinné pro všechny subjekty, které se procesu účastní, a rovněž zajistit, že účastníci tohoto procesu mimo veřejnou správu budou zavázáni podepsat doložku k přijetí svých nabídek, v níž se zavážou spolu s výsledky své práce dodat data vygenerovaná v rámci překladů (překladové paměti), které dohromady tvoří lingvistický zdroj nedocenitelné hodnoty.

REFERENCE. Farrer-Sapena a další (2011) ve svém článku o přístupu k veřejným datům, Ferrer-Sapena a Peset (2012) o opakovaném použití kulturních dat nebo Ramos Simón a další (2012) ve své studii o evropských datových portálech.

Licencování otevřených dat: A Practical Guide for the Higher Education Funding Council for England (Korn and Oppenheim, 2011). Za pozornost stojí také publikace *Guide to Open Data Licensing* (Open Knowledge Foundation, n.d.) a pokyny vycházející z *Evropské digitální agendy* (Evropská komise, 2010).