

Acord-cadru privind partajarea datelor din memoriile de traducere generate ca urmare a traducerii documentelor aparținând administrației publice.

1. INTRODUCERE.

De la adoptarea, în anul 2003, a primului set de reguli privind reutilizarea informațiilor din sectorul public, volumul datelor generate a crescut exponențial la nivel mondial, concomitent cu generarea și colectarea unor noi tipuri de date.

În același timp, suntem martorii unei evoluții permanente a tehnologiilor de procesare a limbajului pentru analizarea, exploatarea și prelucrarea datelor.

Această evoluție tehnologică rapidă permite crearea de noi servicii și aplicații, bazate pe utilizarea, agregarea sau combinarea datelor.

Cantitățile masive de date sau volumele mari de date colectate prin diferite instrumente tehnologice sau extrase din colecții extinse de informații în formate diferite au capacitatea de a genera noi cunoștințe în cele mai diverse domenii, însă ridică, în același timp, probleme legate de proprietatea asupra datelor și de utilizarea ulterioară a acestora.

Administrațiile publice și organismele și organele care alcătuiesc sectorul public sunt generatori majori de date, care sunt de mare folos pentru industriile de procesare a limbajului (natural). Administrațiile publice generează cantități enorme de date în cadrul desfășurării activităților lor. Utilizarea și exploatarea acestor date poate fi de mare interes atât pentru administrația publică în sine, cât și pentru industrie.

Utilizarea acestor date de către administrațiile publice, împreună cu extragerea datelor, cu tehnicile privind volumele mari de date și cu tehnologiile de procesare a limbajului natural, poate facilita luarea deciziilor publice și eficiența politicilor publice.

Acord-cadru privind partajarea datelor din memoriile de traducere (FATDS) – date generate prin traducerea documentelor aparținând administrației publice. PROIECT NEC TM.

Acest acord-cadru se dorește a fi o recomandare adresată statelor membre pentru implementarea licențelor deschise, care să înlăture eventualele obstacole legale, în scopul de a integra aceste tipuri de date în crearea de noi cunoștințe.

Practica curentă a diferitelor guverne, corporații și organizații internaționale care generează date publice este de a se baza pe aceste licențe – în special în modalitățile PDDL, GPL și Creative Commons (CC) pentru care s-a renunțat la orice drept asupra bazelor de date – ca pe un mecanism adecvat pentru favorizarea utilizării și reutilizării acestor volume mari de date.

2. De ce un acord-cadru european pentru managementul partajării datelor.

Motivele pentru un acord-cadru pentru partajarea datelor din memoriile de traducere.

2.1. Abordarea generală.

FATDS (Acordul-cadru privind **partajarea datelor din memoriile de traducere**) nu este accidental, nu are caracter episodic și nici nu este punctual. Dimpotrivă, acesta răspunde unei nevoi căreia îi corespund diferite motivații. Administrațiile publice europene generează zilnic mii de traduceri, iar aceste traduceri generează baze de date a căror eliberare sub formă de date reprezintă deja o obligație legală (Directiva europeană privind datele deschise).

Acestea sunt două aspecte cu o relevanță majoră, care reflectă lipsa unui acord fundamental pentru o reglementare adecvată și pentru gestionarea problemelor legate de partajarea datelor generate prin traducerea textelor provenite de la administrațiile publice. Acesta este **unul dintre sectoarele care generează cea mai mare cantitate de date** de acest tip în cadrul activității lor cotidiene. Este de asemenea sectorul care investește cel mai mult în infrastructurile de stocare a datelor la nivel mondial. Noi ne concentrăm în special pe datele generate de memoriile de traducere, rezultate în urma traducerii documentelor deținute de administrația publică, ținând seama de nevoile celor

Acord-cadru privind partajarea datelor din memoriile de traducere (FATDS) – date generate prin traducerea documentelor aparținând administrației publice. PROIECT NEC TM.

administrații și de numeroasele limbi, nu doar dintr-o anumită țară sau teritoriu, dar din toate țările care alcătuiesc Uniunea Europeană.

Este foarte importantă respectarea recomandărilor privind reutilizarea datelor și interoperabilitatea din sectorul administrației publice. La fel de importantă este și stabilirea condițiilor pentru reutilizare și interoperabilitate.

Fără nicio îndoială, starea optimă ar fi aceea în care datele ar fi accesibile integral și nu ar fi necesare permise speciale pentru reutilizarea acestora, astfel cum s-a stabilit în cadrul Directivei europene privind datele deschise.

Cu toate acestea, pentru a putea fi accesibile și reutilizate, sunt necesare anumite condiții – și anume, în acest caz, ca memoriile de traducere să fie cât mai deschise posibil. Obiectivul nostru prioritar este stabilirea unui cadru de acțiune care să favorizeze într-o mai mare măsură diseminarea și capacitatea de reutilizare a acestora.

Este mai mult decât recomandabil să nu se genereze mai multe tipuri de licențe, ci dimpotrivă, simplificarea tipurilor potrivit recomandărilor din *Agenda digitală* europeană (Comisia Europeană, 2010).

În prezent, există două alternative internaționale principale care promovează utilizarea licențelor în contextul datelor și al informațiilor și care reglementează și promovează accesul liber la informații și utilizarea liberă a acestora. Aceste licențe sunt Creative Commons (CC), Apache 2, GPL și ODC. Creative Commons se aplică datelor și altor documente, iar celelalte sunt utilizate numai în domeniul datelor. Tendința internațională indică faptul că, pentru a reglementa condițiile pentru accesul la date și pentru utilizarea acestora, sunt utilizate modele de licențe deschise, bazate în principal pe CC, însă cu variații adaptate la caracteristicile fiecărui portal de date și ale fiecărei țări unde sunt aplicate.

Toate acestea reprezintă un beneficiu pentru o administrație din ce în ce mai digitalizată, care are capacitatea de a **cuantifica impactul acțiunilor sale în privința randamentului social al investițiilor**, ceea ce reprezintă scopul unor acțiuni precum Memoria de Traducere Națională și Central Europeană (NECTM), acțiuni vizate de

prezentul acord-cadru pentru partajarea datelor generate de memoriile de traducere din textele traduse pentru administrațiile publice.

Prin urmare, din moment ce nu există în prezent niciun consens referitor la modul de reglementare a prelucrării, utilizării și partajării datelor din memoriile de traducere generate ca urmare a nevoilor de publicare ale administrației publice, poate fi propusă abținerea ca opțiune, cu referire la legislația specifică existentă de natură generală din fiecare stat. Cea mai bună alternativă este o reglementare specială pentru această problemă, sub forma unui model de reglementare propus prin prezenta, bazat pe „standarde ușoare”, unilaterale și neobligatorii, și o abordare practică, precum un ACORD-CADRU.

FATDS este de partea ambelor părți implicate în proces, administrația publică și furnizorii. În primul rând, clarifică faptul că există problema privind partajarea datelor din memoriile de traducere, că aceasta este o problemă reală pentru organizațiile implicate și că este necesar să se intervină, respingând totodată excesele și abuzurile. În al doilea rând, FATDS propune un cadru de reglementare complementar și specific, precum și un management sistematic, de asemenea flexibil, adaptabil pentru toate administrațiile, organismele private și statele membre.

În acest scop, este foarte recomandat și necesar, în interesul unei societăți tot mai preocupate de valoarea datelor pe care le generează, ca datele să fie puse la dispoziția contractorului responsabil sau a unui organism central din fiecare stat membru, odată cu executarea fiecărui contract de prestări de servicii de traducere, prin punerea la dispoziție nu doar a textului sau a documentelor respective traduse, ci și a memoriei de traducere generate în procesul de traducere.

De asemenea, grație implementării bazei de date NEC TM, sunt stabilite astfel condiții pentru a facilita reutilizarea acestor memorii de traducere, datorită caracterului lor accesibil și, mai înainte de toate, datorită faptului că nu vor fi supuse niciunui fel de restricții tehnice sau legale care ar putea limita sau restrânge o asemenea reutilizare.

3. Caracterul multidimensional al problemei privind partajarea datelor din memoriile de traducere generate de administrațiile publice: protecție oferită de FATDS.

În sensul strict al naturii sale legale, prezentul contract-cadru se dorește a fi un acord a cărui implementare va avea loc în conformitate cu practicile și procedurile din fiecare stat membru, ținând seama de practicile devenite deja uzuale și cunoscute furnizorilor din industria traducerilor.

Având în vedere considerațiile de mai sus, este evident că acordul-cadru nu poate fi recunoscut ca fiind de aplicare obligatorie din punct de vedere legal, ca și cum ar fi orice altă reglementare a Uniunii, în acest caz provenind dintr-o sursă autonomă, însă nu este posibilă nici subestimarea eficienței sale. Acordul se aplică numai în mod voluntar, în funcție de puterea părților semnatare de a crea obligații în sarcina organizațiilor.

Astfel, FATDS, înțeles ca un acord autonom, se vrea a fi mai mult un instrument practic, cu alte cuvinte, un set de linii directoare care conțin reguli eficiente pentru managementul, centralizarea și partajarea datelor, acordul-cadru fiind strict complementar legislației Uniunii și legislațiilor naționale.

În acest fel, un acord non-normativ, așa cum este FATDS, poate constitui, și constituie prin prezenta, un instrument util pentru concentrarea și clarificarea interpretării partajării datelor generate de memoriile de traducere ca urmare a activității desfășurate de ofertanți/furnizori.

Prin urmare, funcția FATDS este de a clarifica și preciza dreptul administrațiilor publice de a solicita toate datele generate în executarea contractelor de traducere, dreptul lor de proprietate asupra textelor originale și a traducerii, ca serviciu contractat și derivat dintr-un original, reflectând consensul dintre actorii implicați, prin cei implicați înțelegându-se, pe de o parte, companiile ofertante (furnizorii) și, pe de altă parte, organismul administrativ din fiecare stat membru și, în cele din urmă, organismul central al Uniunii Europene care va fi stabilit în acest scop.

În acest sens, nu ar trebui să se uite faptul că configurația cea mai acreditată a acestui model sau paradigme juridice ilustrează existența unor instrumente care, fără a se conforma tipurilor tradiționale de norme juridice, nu pot fi excluse din lumea juridică, fiind astfel obligatorii, pentru a exprima voința fermă a părților semnatare care execută contractul de servicii în executarea competențelor lor în cadrul unei proceduri legale previzibile de a pune în practică un instrument cu o valoare juridică minimă sau relativă.

Conținutul fundamental al FATDS se concentrează mai mult pe stabilirea de modele și linii directoare, formând un instrument tipic de legislație de reflexie.

În cele din urmă, această paradigmă de reglementare materializată prin FATDS este încadrată strict în noile abordări și noile metode de reglementare și de acțiune propuse în prezent, atât de Uniunea Europeană, care se bazează mai mult pe instrumente care transformă dispozițiile legale în practici uzuale decât pe adoptarea a tot mai multor norme juridice, care definesc în cazul nostru utilizarea și posibila reutilizare a informațiilor conținute în bazele de date care fac obiectul de reglementare al legii, cât și prin intermediul contractelor și/sau licențelor.

CEE, publicarea comunicării privind datele deschise și propunerea Comisiei Europene de modificare formează o politică de deschidere a datelor și de promovare a unei piețe informaționale care are ca punct central aprobarea Directivei 2003/98/CE a Parlamentului European și a Consiliului din 17 noiembrie 2003 privind reutilizarea informațiilor din sectorul public.

Mai trebuie adăugată la acestea revizuirea Directivei privind reutilizarea informațiilor din sectorul public, introdusă la sfârșitul anului 2011, care face parte din agenda digitală europeană, care este o inițiativă a Comisiei (2010); această agendă are scopul de a favoriza serviciile online în cadrul Uniunii și a cărei prioritate este deschiderea în vederea reutilizării a datelor publice, simplificarea sistemului de licențiere pentru schimbul de conținuturi și implementarea standardelor de interoperabilitate.

5. Opțiunea de reglementare a FATDS

Datele publice deținute de organismele publice din Uniunea Europeană fac obiectul unui tratament specific în conformitate cu directivele din 2003 și 2013 privind reutilizarea informațiilor din sectorul public, care prevăd că documentele administrației publice trebuie să fie puse la dispoziția publicului pentru reutilizare, și anume „orice conținut, indiferent de mediul de stocare (scris pe hârtie sau stocat în format electronic sau sub formă de înregistrare audio, video sau audiovizuală) deținut de organismele din sectorul public în scopuri comerciale sau necomerciale”. Prin urmare, principiul general este disponibilitatea liberă a acestor conținuturi, iar dacă se aplică vreo taxă, aceasta „se limitează la costurile marginale suportate în legătură cu reproducerea, punerea la dispoziție și difuzarea” [art. 6 alineatul (1)].

Pe scurt, FATDS se prezintă în esență ca o „recomandare” bazată pe un minim de relativitate juridică, întemeiată pe contractele de servicii sau datele existente (memorii de traducere) generate de un profesionist (ex-professo). Această „recomandare” este specificația practică a obligației de a partaja datele generate sub formă de memorii de traducere create în cadrul serviciilor de traducere contractate de administrațiile publice, indiferent de mediul pe care sunt stocate sau de formatul acestora.

6. Conținutul acordului-cadru: sistemul de management pentru transferul datelor din memoriile de traducere promovat prin acord.

O memorie de traducere este o bază de date lingvistică în care se stochează în mod continuu traduceri paralele efectuate de profesioniști, pentru a fi utilizate în viitor, în scopul de a obține o mai mare consecvență terminologică și stilistică și pentru reducerea costurilor de traducere ca urmare a concordanțelor totale sau parțiale ale noilor texte cu traducerile anterioare efectuate deja.

Memoriile de traducere sunt colecții digitale formate din segmente de text în limba originală aliniată cu traducerile acestora în alte limbi. Aceste texte pot fi de asemenea

aliniat în mod eficient pe unități de traducere. Unitățile de traducere care sunt stocate împreună cu echivalentele lor sunt definite în mod diferit (frază, paragraf, cuvânt, grup de cuvinte etc.), segmentarea după fiecare punct marcând sfârșitul propoziției sau al paragrafului, instrucțiunea de revenire (return) fiind cel mai frecvent separator în mediul sistemelor de traducere asistate de calculator.

Funcția principală a memoriilor de traducere (denumite în continuare „TM”) este de a extrage sugestii totale sau parțiale ori concordanțe dintr-o propoziție și corespondențe pentru termeni. În timpul traducerii, segmentele din limba sursă sunt căutate în baza de date a TM. Dacă TM conține un segment în limba sursă care concordă total sau parțial, acest segment va fi sugerat traducătorului, alături de traducerea identificată în baza de date, precum și orice informație suplimentară care a fost salvată împreună cu segmentul în baza de date. Instrumentele de traducere asistată de calculator (denumite „CAT tools”) afișează diferite niveluri de similitudine [cunoscute în domeniu sub denumirea de *fuzzy matches* (concordanțe imprecise)]. În industria traducerilor, aceste concordanțe pot fi utile lingviștilor pornind de la un nivel de precizie de 50%–65%, prețurile fiind de asemenea stabilite în funcție de efortul depus în traducerea frazelor complet noi sau a altor fraze pentru care există sugestii similare și concordanțe procentuale din același context. Un nivel de precizie de 100% este considerat a fi o concordanță perfectă între propoziția de tradus și o propoziție identică din baza de date.

Formatul liber pentru schimbul de memorii de traducere este TMX (Translation Memory eXchange), în general în versiunea 1.4b. Acesta este un standard XML de tipul DTD (*document type definition*). Acesta a fost creat de comitetul OSCAR (Standarde deschise pentru suporturi/conținuturi care pot fi reutilizate).

Datorită aplicării formatului TMX, crește fiabilitatea colaborării în proiectele de traduceri pentru oameni și companii. De asemenea, formatul TMX face mai ușoară migrarea de la un sistem de traducere asistată de calculator la altul, ceea ce favorizează competitivitatea între tehnologiile oferite și dezvoltarea lor continuă, în scopul diferențierii calitative dintre competitorii care dețin aceste tehnologii. La fel ca alte standarde deschise, acest format a fost conceput în scopul de a reduce problemele de compatibilitate, de a încuraja

reutilizarea resurselor lingvistice, de a simplifica schimbul de date și, astfel, de a stimula inovația tehnologică.

Astfel cum am arătat, scopul acestui acord-cadru este de a stabili „un cadru pentru bune practici”, în vederea obținerii datelor generate pe parcursul executării contractelor de traduceri încheiate cu administrațiile publice europene, organizarea acestora prin implementarea sistemului software NEC TM și a inițiativelor de centralizare a datelor ale Comisiei Europene, în scopul adoptării unui protocol pentru centralizarea datelor bilingve ale administrațiilor publice, utilizarea acestora în beneficiul societății în general și al statelor membre însele, asistență la crearea unui corpus de date naționale de mare volum și partajarea acelor părți pe care administrațiile naționale le consideră relevante la nivel european.

În acest scop, centralizarea datelor găzduite de infrastructurile informatice naționale va fi administrată de organismul competent al statului membru respectiv – de exemplu, Secretariatul de Stat pentru Dezvoltare Digitală – sau de către ministerul de resort din statul membru respectiv. Aceste date pot fi partajate în cazul aderării la acest acord-cadru și dacă se stabilește astfel de către autoritatea competentă la un nivel superior cu organismul relevant din cadrul Comisiei Europene (de exemplu, partajare ELRC).

Acestea fiind spuse, primul dintre pași va fi centralizarea datelor din memoriile de traducere generate de administrațiile publice la nivel național prin adoptarea prezentului acord, iar apoi, în conformitate cu acesta, partajarea lor cu un organism paneuropean în beneficiul administrațiilor publice și al celorlalte state membre. Utilizarea programului software „Memoria de Traducere Națională și Central Europeană” (NEC TM) va furniza conexiunea intrare/ieșire (I/O) pentru administrații, care își vor putea stoca în mod privat memoriile de traducere găzduite pe un server central și să le partajeze cu aceia care decid să își eficientizeze munca și să reducă costurile la nivel național (în mod normal, traducătorii interni sau furnizorii externi), prin crearea de date naționale de volum mare și selectarea datelor dorite de pe serverul central care găzduiește memoria de traducere central europeană.

Relația cu entitățile private și alocarea TMX-ului

Munca de traducere executată de companiile ofertante (furnizorii) va cuprinde furnizarea memoriilor de traducere pe care le generează în cursul prestării serviciului. În acest scop, în anunțul care apare în publicația oficială a statului membru, a regiunii, municipalității sau județului, și anume pe portalul autorității contractante, ar trebui să fie inclusă o clauză referitoare la NEC TM FATDS, inclusiv CODUL CPV, care va stabili CĂ DATELE DIN MEMORIA DE TRADUCERE VOR FI FURNIZATE PE SERVERUL AUTORITĂȚII PUBLICE COMPETENTE în scopul de a fi partajate în viitor, dacă autoritatea națională va decide astfel, cu organismul din cadrul Comisiei Europene care urmează să fie stabilit ulterior (partajare ELRC sau altă inițiativă similară, de exemplu).

În acest scop, companiile ofertante (furnizorii) vor livra memoriile de traducere generate împreună cu traducerea efectuată, aplicând principiul neretroactivității, ceea ce înseamnă că acest acord-cadru nu trebuie să producă efecte pentru trecut și că acesta va avea efecte obligatorii numai din momentul aderării și semnării clauzei prin care aceștia se obligă să livreze împreună cu traducerea efectuată datele paralele generate în procesul de traducere (în format TMX sau alt format similar compatibil).

7. DREPTUL DE PROPRIETATE INTELECTUALĂ ȘI CONDIȚIILE PENTRU UTILIZAREA BAZEI DE DATE

În ceea ce privește dreptul de proprietate intelectuală, nu există nicio îndoială că datele generate în executarea prestației în conformitate cu contractul, în virtutea căruia a fost acordată traducerea în discuție, aparțin în întregime și exclusiv administrației publice, care este autorul sau managerul textului original care face obiectul traducerii.

Nu poate fi ignorat faptul că în măsura în care, în majoritatea cazurilor, informația provine de la administrațiile publice, trebuie avută în vedere, de asemenea, reglementarea privind reutilizarea informațiilor publice și deschiderea datelor publice.

Astfel, ofertantul căruia i-a fost atribuită lucrarea (furnizorul) va atribui în mod exclusiv și fără nicio limită temporală sau spațială drepturile asupra oricărui tip de documentație sau

Acord-cadru privind partajarea datelor din memoriile de traducere (FATDS) – date generate prin traducerea documentelor aparținând administrației publice. PROIECT NEC TM.

asupra datelor generate, indiferent de suportul pe care acestea se află sau de formatul lor, distribuția și transformarea lor fiind incluse în transferul drepturilor de reproducere.

Prin urmare, datele generate de astfel de memorii de traducere nu vor putea fi folosite de contractantul căruia i-a fost atribuit contractul în scopuri lucrative, putând fi folosite numai pentru furnizarea de informații, dacă furnizarea este de natura prestației executate.

Dreptul de proprietate intelectuală și condițiile pentru utilizarea bazelor de date.

Baza de date DGT-TM este proprietatea exclusivă a Comisiei Europene. Comisia conferă, în mod gratuit și la nivel mondial, pe întreaga perioadă în care aceste drepturi sunt protejate, reutilizatorilor de date drepturile sale neexclusive pentru orice tip de utilizări pentru care sunt îndeplinite condițiile prevăzute de Decizia Comisiei din 12 decembrie 2011 privind reutilizarea documentelor Comisiei, publicată în Jurnalul Oficial al Uniunii Europene, L 330 din 14 decembrie 2011, p. 39-42.

Orice reutilizare a bazei de date sau a elementelor structurate pe care le conține trebuie să fie identificată de reutilizator, care este obligat să indice sursa documentelor utilizate: adresa paginilor de internet, data ultimei actualizări și faptul că Comisia Europeană este titulara dreptului de proprietate asupra datelor.

Această bază de date este, așadar, optimă pentru popularea versiunii NEC TM a unui stat membru.

8. PROTECȚIA DATELOR. APLICAREA REGULAMENTULUI UE NR. 679/2016 ȘI A REGULAMENTULUI NR. 1725/2018

În ceea ce privește datele cu caracter personal care trebuie prelucrate în contextul executării contractului de către administrația publică și de către ofertantul căruia i s-a atribuit contractul, ambele părți sunt obligate să respecte Regulamentul general privind protecția datelor de mai jos.

Acord-cadru privind partajarea datelor din memoriile de traducere (FATDS) – date generate prin traducerea documentelor aparținând administrației publice. PROIECT NEC TM.

Regulamentul (UE) 2016/679 al Parlamentului European și al Consiliului din 27 aprilie 2016, precum și respectarea legislației naționale în vigoare privind protecția datelor din fiecare stat membru care adoptă prezentul acord-cadru, precum și Regulamentul Parlamentului European și al Consiliului din 23 octombrie 2018 privind protecția persoanelor fizice cu privire la prelucrarea datelor cu caracter personal de către instituțiile și organele Uniunii și privind libera circulație a acestor date.

CONCLUZII FINALE.

Administrația publică, în calitate de promotor al industriei tehnologiilor lingvistice, este obligată, odată cu crearea platformelor comune pentru prelucrarea limbilor naturale și traducerea automată, precum și dezvoltarea resurselor pentru reutilizarea informațiilor din sectorul public (RISP), să dezvolte politici privind partajarea și să stabilească bazele pentru ca această partajare să aibă loc în mod real și efectiv între toate entitățile care participă în proces, asigurând totodată că participanții din afara administrației vor obține promisiunea de a semna clauza de acceptare a ofertelor lor și de a se obliga să livreze împreună cu traducerile efectuate datele generate prin traducerile efectuate (memoriile de traducere), care formează împreună un set de resurse lingvistice foarte valoroase.

LISTA DE REFERINȚE. Ferrer-Sapena și alții (2011) în articolul privind accesul la datele publice; Ferrer-Sapena și Peset (2012) privind reutilizarea datelor culturale sau Ramos Simón și alții (2012) în studiul privind portalurile de date europene.

Licențierea datelor deschise: A Practical Guide for the Higher Education Funding Council for England (Korn și Oppenheim,2011). Trebuie de asemenea amintite *Guide to Open Data Licensing* (Open Knowledge Foundation, n.d.) și liniile directoare din *Agenda digital europeană* (Comisia Europeană, 2010).