

Ramavtal för delning av data från översättningsminnen som genererats från översättningar av dokument från den offentliga förvaltningen.

1. INTRODUKTION.

Sedan den första uppsättningen av regler om återanvändning av information från den offentliga sektorn antogs 2003 har volymen av data som genererats ökat exponentiellt över hela världen, medan nya typer av data genereras och samlas in.

Samtidigt bevitnar vi en permanent utveckling av språkbehandlingstekniker för analys, utnyttjande och behandling av data.

Denna snabba tekniska utveckling gör det möjligt att skapa nya tjänster och applikationer baserade på användning, aggregering eller kombination av data.

De enorma mängderna data, eller big data, som samlas in genom olika tekniska verktyg eller som hämtats från stora samlingar av information i olika format kan generera ny kunskap inom de mest mångfaldiga sektorerna, men innebär samtidigt problem för både dataägande och för dess efterföljande användning.

Offentliga förvaltningar (OF:er) och de organ och enheter som utgör den offentliga sektorn är stora producenter av data som är till stor nytta för (naturlig) språkbehandlingsindustrier. Offentliga förvaltningar genererar enorma mängder data i sitt arbete. Den användning och det utnyttjandet av dessa data kan vara av stort intresse både för både OF och industrin.

Offentliga förvaltningars användning av dessa data i kombination med data mining och big data-tekniker och naturliga språkbehandlingstekniker kan underlätta beslutsfattande och effektiviteten vad gäller policyer i den offentliga sektorn.

Detta ramavtal ämnar vara en rekommendation till medlemsstaterna om genomförande av öppna licenser som avlägsnar eventuella juridiska hinder för att integrera dessa typer av data i skapandet av ny kunskap.

Det har blivit en aktuell praxis för olika regeringar, företag och internationella organisationer som producerar offentliga data att förlita sig på dessa licenser – i synnerhet modaliteterna PDDL, GPL och Creative Commons (CC) där alla rättigheter till databaser är fransägs som en lämplig mekanism till fördel för användning och återanvändning av dessa stora volymer av data.

2. Varför ett europeiskt ramavtal för hantering av data

Skäl till ramavtalet för att dela data från översättningsminnen.

2.1. Allmän riktlinje.

FATDS (Framework **Agreement** on **TMX** Data **Sharing**) är varken en tillfällighet eller rent episodisk eller punktlig. Tvärtom svarar det på ett behov som finner olika motiv. Europeiska offentliga förvaltningar genererar tusentals översättningar varje dag, och dessa översättningar genererar databaser vars utsläpp som data redan är ett lagligt mandat (Open Data Europe-direktivet).

Det finns två punkter av stor betydelse som avspeglar avsaknaden av en grundläggande samsyn kring adekvat reglering och den hantering av problemen i samband med utbyte av de data som genereras av översättningar av texterna från offentliga förvaltningar. Detta är **en av de sektorer som genererar den största mängden data** av detta slag under sin dagliga verksamhet. Det är också den sektor som investerar mest i infrastruktur för datalagring över hela världen . Vi fokuserar särskilt på de data som genereras av översättningsminnen från dokument som ägs av den offentliga förvaltningen, med tanke på behoven hos de administrerade och de många språken, inte bara inom ett territorium eller ett land, utan i alla länder som utgör EU.

Det är mycket viktigt att följa rekommendationerna för återanvändning av data och interoperabilitet i den offentliga förvaltningen. Det är lika viktigt att fastställa villkoren för återanvändning och driftskompatibilitet.

Utan tvekan skulle det optimala tillståndet vara att uppgifterna är fullt tillgängliga och för att deras återanvändning inte kräver specifika tillstånd, vilket fastställs i EU-direktivet om öppna data.

Men för att möjliggöra tillgänglighet och fortsatt användning krävs vissa villkor, och det är därför licenser är nödvändiga – i detta fall bör översättningsminnen vara så öppna som möjligt. Det är vårt prioriterade mål att skapa en ram för åtgärder som gynnar dess större spridning och kapacitet för återanvändning,

Det är mer än lämpligt att inte generera fler typer av licenser, utan tvärtom, vilket förenklar de typer som rekommenderas av *den europeiska digitala agendan* (Europeiska kommissionen, 2010).

För närvarande finns det två huvudsakliga internationella alternativ som främjar användningen av licenser i samband med data och information, och som reglerar och främjar fri tillgång till och användning av information. Dessa licenser är Creative Commons (CC), Apache 2 och GPL och ODC. Creative Commons gäller för både data och andra dokument, medan den senare används endast i dataområdet. Den internationella trenden indikerar att om villkoren för tillgång till och användning av data ska regleras, måste öppna licensmodeller användas, huvudsakligen baserade på CC, men med variationer anpassade till egenskaperna hos varje dataportal och varje land där de tillämpas.

Allt detta är till förmån för en alltmer digital förvaltning som kan **mäta effekterna av sina åtgärder när det gäller social avkastning på investeringar**, vilket är målet för åtgärder som det nationella och europeiska centrala översättningsminnet (NECTM) genom detta Ramavtal för att dela data som genererats av översättningsminnen från texter översatta för offentliga förvaltningar.

Eftersom det för närvarande inte råder enighet om hur man reglerar behandlingen, att använda och att dela data från översättningsminnen som

genereras till följd av offentliga förvaltningars publiceringsbehov, kan antingen ett abstentionistiskt alternativ föreslås, med hänvisning till specifika existerande lagar av allmän karaktär i varje medlemsstat . Det bättre alternativet är en särskild reglering för denna fråga i form av en lagstiftningsmodell som härmed föreslagits och baseras på "mjuka standarder", av ensidig, icke-bindande karaktär, och en praktisk metod som ett RAMAVTAL.

FATDS stöder båda sidor som är involverade i processen, PA och leverantörer. För det första klargör man att det finns problem med att dela data från översättningsminnet, att det är något involverade organisationer arbetar med och att det är nödvändigt att ingripa, samtidigt som man avvisar överdrifter och missbruk. För det andra föreslår FATDS en kompletterande och specifik ram för reglering, samt en systematisk ledning, som också är flexibel, anpassningsbar för alla förvaltningar, privata enheter och medlemsstaterna.

Därför är det starkt rekommenderat och nödvändigt för ett samhälle som i allt högre grad överväger värdet av data det genererar att dessa data görs tillgängliga för den ansvariga entreprenören eller ett centralt organ inom varje medlemsstat, med alla de översättningsavtal som genomförs, vilket inte bara gör texten eller dokumenten för den aktuella översättningen tillgängliga utan även det översättningsminne som genererades under processen.

På samma sätt, tack vare implementeringen av NEC TM-databasen, skapas därför villkor för att underlätta återanvändning av dessa översättningsminnen tack vare deras tillgänglighet, och de kommer framför allt inte att bli föremål för tekniska eller lagliga begränsningar som begränsar eller hindrar sådan återanvändning.

3. Flerdimensionalitet av problemet med datadelning av översättningsminnen som genereras av offentliga administrationer: skydd som erbjuds av FATDS.

I dess rättsliga naturs strikta bemärkelse är detta ramavtal avsett att vara ett avtal vars genomförande sker i enlighet med varje medlemsstats metoder och

procedurer och med beaktande av metoder som redan är vanliga och inte okända för leverantörer i översättningsbranschen.

På grundval av ovanstående överväganden verkar det uppenbart att det inte är möjligt att erkänna ramavtalen med en direkt rättslig effektivitet som om det bara var en annan del av gemenskapslagstiftningen, i detta fall från en autonom källa, men det är inte heller möjligt att underskatta dess effektivitet. Efterlevnad av avtalet är helt frivilligt, och strikt beroende av de undertecknande personernas makt att binda organisationerna.

FATDS som ett autonomt avtal syftar således till att vara ett praktiskt instrument, det vill säga en uppsättning vägledande riktlinjer som anger effektiva regler för hantering, centralisering och delning av data, varvid ramavtalet strikt kompletterar gemenskapens och nationell lagstiftning.

Därför kan ett icke-normativt avtal, såsom FATDS, utgöra, och härmed utgör, ett användbart instrument för att konkretisera och tydliggöra tolkningen av utbyte av de data som genereras av översättningsminnen som skapas som ett resultat av det arbete som utförts av budgivarna/leverantörerna.

FATDS:s funktion är därför att klargöra och specificera rätten för offentliga förvaltningar att begära alla data som genererats i översättningskontrakt, deras äganderätt till originaltexter och översättning som en tjänst som kontrakterats och härrör från en ursprunglig, vilken återspeglar de berörda aktörernas samsyn, förståelse som de involverade å ena sidan budgivningsföretagen (leverantörerna) och å andra sidan det administrativa organet från varje medlemsstat och slutligen EU:s centrala organ som kommer att bestämmas för detta ändamål.

I detta avseende bör man komma ihåg att den mest ackrediterade konfigurationen av denna modell eller juridiska paradigm illustrerar förekomsten av instrument som, utan att överensstämja med traditionella typer av laglig norm, inte heller kan uteslutas från lagens värld och därför från obligatorisk karaktär, för att uttrycka ett fast åtagande från de som undertecknar de som undertecknar tjänsterna i utövandet av sina befogenheter enligt ett

lagligt förutsagt förfarande, genom att använda ett instrument som har en lägsta eller relativ juridisk handling.

Det grundläggande innehållet i FATDS fokuserar mer på upprättandet av modeller och riktlinjer och utgör ett typiskt instrument för reflektionsrätt.

I slutändan utformas detta paradigm av reglering som realiserats av FATDS i de nya strategierna och de nya metoderna för reglering och åtgärder som för närvarande föreslås, liksom av EU, som förlitar sig mer på instrument som omvandlar den rättsliga regeln till praktik snarare än att fastställa fler regler, i vårt fall som definierar tillgång, användning och eventuell återanvändning av informationen i databaserna reglerade genom lag, liksom genom kontrakt och/eller licenser.

Från EEG utgör offentliggörandet av meddelandet om öppna data och Europeiska kommissionens förslag till ändring en policy om att öppna data och främja en informationsmarknad som har sin utgångspunkt i godkännandet av Europaparlamentets direktiv 2003/98/EG och av rådet av den 17 november 2003 om återanvändning av den offentliga sektorns information.

Till detta måste vi lägga till översynen av direktivet om återanvändning av information från den offentliga sektorn, som infördes i slutet av 2011, som utgör en del av den europeiska digitala agendan, ett initiativ som genomförts av Europeiska kommissionen (2010), varvid denna agenda syftar till att gynna onlinetjänster inom unionen och att prioritera öppnande av offentliga data för återanvändning, förenkling av licenssystemet för utbyte av innehåll och implementering av standarder för interoperabilitet.

5. FATDS:s regleringsalternativ

Offentliga data som innehas av offentliga organ i Europeiska unionen är föremål för särskild behandling enligt direktivet om återanvändning av offentliga sektorer 2003 och 2013, som föreskriver att PA- dokument ska tillgängliggöra

"allt innehåll för allmänheten för återanvändning och oavsett medium (skrivet på papper eller om det är lagrat i elektronisk form eller som ljud, visuell eller audiovisuell inspelning) som hålls av offentliga organ för kommersiella eller icke-kommersiella ändamål". Därför är den allmänna principen att fri tillgång till sådant innehåll, om någon avgift tillämpas, ska "begränsas till de marginalkostnader som uppkommer för dess reproduktion, tillgänglighet och spridning" (artikel 6.1).

Kort sagt presenterades FATDS i princip som en "rekommendation" baserad på en minimal eller relativ juridiskhet, baserad på befintliga serviceavtal och data (översättningsminnen) genererade ex profeso. Denna "rekommendation" är den praktiska specifikationen av skyldigheten att dela data som genereras i form av översättningsminnen härrörande från översättningstjänster som kontrakterats av offentliga förvaltningar, oavsett medium eller format.

6. Innehållet i ramavtalet: hanteringssystemet för överföring av data från de översättningsminnen som avtalet främjar.

Ett översättningsminne är en språkdatabas som kontinuerligt lagrar parallella översättningar som genereras av yrkesverksamma i syfte att kunna använda dem i framtiden, i syfte att få större terminologisk och stilmässig konsistens, liksom besparingar på grund av fullständiga eller delvisa matchningar mellan nya texter och gamla översättningar som redan har slutförts.

Översättningsminnen är digitala lager som består av textlinjer från innehållet på originalspråket i linje med dess översättning till andra språk. Dessa texter kan också anpassas effektivt av översättningsenheter. Översättningsenheterna som lagras tillsammans med sina medel definieras på olika sätt (genom fras, stycke, ord eller en grupp av ord, o.s.v.). Segmenteringen efter ett skiljetecken som markerare slutet av meningen eller en manuell radbrytning är oftast den vanligaste avgränsaren inom datorassisterade översättningssystem.

Huvudfunktionen för översättningsminnen (nedan TM) är att extrahera totala eller partiella förslag eller matchningar från en mening och konkordanser för termer. Under översättningen söker man på källspråkets segment i TM-databasen. Om TM har ett segment på källspråket som exakt eller delvis matchar det föreslås detta segment till översättaren, tillsammans med översättningen som hämtats från den databasen, och all ytterligare information som har sparats med segmentet i databasen. Datorassisterade översättningssystem (CAT-verktyg) visar grader av likhet (kallas i industrin för *suddiga* matchningar). I översättningsbranschen kan dessa matchningar vara användbara för lingvister från 50 %–65 % upp och prislistor fastställs även utifrån det arbete som krävs för att översätta helt nya fraser eller andra för vilka det finns liknande förslag och procentmatchningar inom sammakontext. En likhetsgrad på 100 % anses vara en fullständig matchning mellan en begärd mening och en identisk mening i databasen.

Det fria formatet för utbyte av översättningsminnen är TMX (Translation Memory eXchange), vanligtvis i version 1.4b. Detta är en XML-standard för typen DTD (*Document Type definition*). Den skapades av OSCAR-kommittén (Open Standards for Container/Content Allowing Re-use).

Genom att använda TMX-formatet är det mer genomförbart för människor eller företag att samarbeta i översättningsprojekt. TMX-formatet gör det också enklare att migrera från ett datorassisterat översättningssystem till ett annat, vilket gynnar konkurrenskraften mellan de tekniker som erbjuds och deras ständiga utveckling för att göra en skillnad gentemot sina konkurrenter. Liksom andra öppna standarder utvecklades detta format för att minska kompatibilitetsproblem, driva återanvändning av språkliga resurser, förenkla informationsutbytet och därmed stimulera teknisk innovation.

Detta ramavtal syftar till att upprätta, vilket vi har definierat väl, en "ram för god praxis" för att erhålla data genererade i översättningsavtal av europeiska offentliga förvaltningar, dess organisation genom implementeringen av NEC TM-programvarusystemet och datacentraliseringsinitiativ utses av Europeiska kommissionen för att anta ett protokoll för centralisering av tvåspråkiga data av offentliga förvaltningar, deras användning och fördelar för samhället i

allmänhet och medlemsstaterna själva, stöd till skapandet av ett korpus av nationella Big Data och delningen av de delar som varje nationell förvaltning anser vara relevant på europeisk nivå.

För detta ändamål kommer centraliseringen av data som finns på nationell datorinfrastruktur att hanteras av det behöriga organet i den berörda staten – till exempel det specifika statssekretariatet för digitalt främjande – eller av det behöriga ministeriet i den berörda staten. Dessa uppgifter kan delas, i samband med anslutning till detta ramavtal och om så bestäms av den behöriga myndigheten, på en högre nivå med Europeiska kommissionens berörda organ (t.ex. ELRC-Share).

Med detta sagt kommer det första av stegen att vara centraliseringen av uppgifterna från de översättningsminnen som genereras av de offentliga myndigheterna på nationell nivå genom antagandet av detta avtal, och därefter, och i enlighet med det, deras delning med en ett paneuropeiskt organ till förmån för de övriga europeiska medlemsstaterna. Användningen av programvaran "National and European Central Translation Memory" (NEC TM) kommer att tillhandahålla input/output-anslutningen (I/O) för förvaltningar, som kommer att kunna lagra sina översättningsminnen privat, på en central server och dela dem med de som bestämmer sig för att göra sitt arbete mer effektivt och kostnadsbesparande på nationell nivå (vanligtvis interna översättare eller externa leverantörer), skapa nationella Big Data och välja de data de önskar med den centrala servern där det europeiska centrala översättningsminnet är inrymt.

Förhållande till privata enheter och tilldelning av TMX

Översättningsarbetet som utförs av budgivningsföretag (leverantörer) kommer att medföra leveransen av de översättningsminnen som genereras av dem som ett resultat av tillhandahållandet av deras tjänst. För detta ändamål i det meddelande som förekommer i den officiella tidningen i den medlemsstat, region eller kommun, det vill säga i entreprenörens portal, ska en klausul innefatta hänvisar ring till NEC TM FATDS, inklusive CPV Code, som kommer att fastställa ATT DATA FRÅN ÖVERSÄTTNINGSMINNET KOMMER ATT TILLHANDAHÅLLAS TILL SERVERN HOS BEHÖRIG OFFENTLIG FÖRVALTNING för de syften som i framtiden, och om den nationella

myndigheten beslutar om det, skulle kunna delas med Europeiska kommissionens organ som ska fastställas (ELRC-Share eller liknande initiativ, till exempel).

För detta ändamål kommer budgivningsföretag (leverantörer) att leverera översättningsminnen som genererats tillsammans med sitt arbete, tillämpning av principen om förbud mot icke-retroaktivitet, vilket innebär att detta ramavtal inte får ha retroaktiva effekter, och att denna obligatoriska karaktär endast kommer att gälla från undertecknandet av den klausul genom vilken de åtar sig att, tillsammans med det utförda arbetet, leverera de parallella data som genereras av deras översättningar (TMX eller liknande kompatibelt format).

7. IMMATERIALRÄTT OCH DATABASENS ANVÄNDNINGSVILLKOR

När det gäller immaterialrätt måste man veta att de data som genererats vid utförandet av arbetet i enlighet med kontraktet, för vilket den aktuella översättningen har tilldelats, är helt och exklusivt tillhörande den offentliga förvaltningen som är författaren eller förvaltare av den ursprungliga texten som är föremål för översättningen.

När informationen har sitt ursprung hos de offentliga förvaltningarna, måste reglering av återanvändning av offentlig information och öppnande av offentliga uppgifter också beaktas.

Således tilldelar den framgångsrika budgivaren (leverantören) uteslutande och utan några tidsbegränsningar eller territoriella begränsningar rätt till alla typer av dokumentation eller data som genereras oberoende av deras stöd eller format, förstå distribution och transformation i denna överföring av reproduktionsrättigheter.

Följaktligen får data som genereras av sådana översättningsminnen inte användas av den vinnande budgivaren för något lukrativt syfte och får endast användas för att tillhandahålla information om detta är fallet för det utförda arbetet.

Immaterialrätt och databasernas användningsvillkor

DGT-TM-databasen är Europeiska kommissionens exklusiva egendom. Kommissionen beviljar, kostnadsfritt och med global omfattning, under hela skyddsperioden för dessa rättigheter, dess icke-exklusiva rättigheter till återanvändare för alla typer av användningar som uppfyller villkoren i kommissionens beslut från den 12 december 2011 om återanvändning av kommissionens handlingar, publicerade i Europeiska unionens officiella tidning L 330 av den 14 december 2011, sidorna 39 till 42.

Återanvändning av databasen eller de strukturerade elementen i den måste identifieras av återanvändaren, som är skyldig att ange källan till de dokument som används: webbplatsens adress, datumet för den senaste uppdateringen och det faktum att Europeiska kommissionen behåller ägandet av uppgifterna.

Denna databas är därför optimal för att ursprungligen fylla NEC TM-versionen av en medlemsstat.

8. DATASKYDD. TILLÄMPNING AV FÖRORDNING EU 2016/679 OCH FÖRORDNING 2018/1725

Beträffande personuppgifter som måste behandlas enligt avtalet mellan den offentliga förvaltningen och den vinnande budgivaren, ska båda parter vara skyldiga att följa nedanstående allmänna bestämmelser om dataskydd.

Europaparlamentets och rådets förordning (EU) 2016/679 från den 27 april 2016, samt överensstämmelse med gällande statlig lagstiftning om dataskydd i var och en av de medlemsstater som antar detta ramavtal samt regleringen av Europaparlamentet och rådet från den 23 oktober 2018 gällande skydd för enskilda med avseende på behandling av personuppgifter av unionens institutioner, organ, kontor och byråer och spridning av sådana data.

SLUTSATSER.

Den offentliga förvaltningen i egenskap av förespråkare för språkteknikindustrin, med skapandet av gemensamma plattformar för naturlig språkbehandling och automatisk översättning samt utveckling av resurser för återanvändning av information från den offentliga sektorn (RISP), är skyldiga att utveckla policyer för delning av data och lägga ut grunden för att denna delning ska vara implementerad och i kraft bland alla enheter som deltar i processen, och även säkerställa att deltagare i processen utanför förvaltningen undertecknar klausulen om godkännande av sina anbud, att de åtar sig att, tillsammans med sitt arbete, leverera de data som genereras från deras översättningar (översättningsminnen) som tillsammans utgör en ovärderlig uppsättning lingvistiska resurser.

Referenser. Ferrer-Sapena och andra (2011) i sin artikel om tillgång till offentliga uppgifter; Ferrer-Sapena och Peset (2012) om återanvändning av kulturella data eller Ramos Simon och andra (2012) i sin studie om europeiska dataportaler.

Licensing Open Data: A Practical Guide for the Higher Education Funding Council for England (Korn and Oppenheim, 2011). *Guide to Open Data Licensing* (Open Knowledge Foundation, nd) och riktlinjerna från *European Digital Agenda* (Europeiska kommissionen, 2010) är också värda att nämna.