

## **Rammeaftale om deling af data fra oversættelseshukommelse genereret som et resultat af oversættelser af Public Administration-dokumenter.**

### **1. INTRODUKTION.**

Siden vedtagelsen i 2003 af det første sæt regler om genbrug af information fra den offentlige sektor er mængden af genererede data steget eksponentielt på verdensplan, mens nye typer data genereres og indsamles.

Samtidig er vi vidne til en permanent udvikling af sprogbehandlingsteknologier til dataanalyse, udnyttelse og behandling.

Denne hurtige teknologiske udvikling muliggør oprettelse af nye tjenester og applikationer baseret på brug, sammenlægning eller kombination af data.

De enorme mængder data eller big data indsamlet gennem forskellige teknologiske værktøjer eller udvundet fra store samlinger af information i forskellige formater er i stand til at generere ny viden i de mest forskellige sektorer, men på samme tid udgør problemer både for dataejerskab og for dens efterfølgende anvendelse.

Offentlig administrationer (PA'er) , og de organer og enheder, der udgør den offentlige sektor er store producenter af data, som er til stor nytte for ( naturlig- ) sprogforbearbejdningsindustrier . Offentlige administrationer genererer enorme mængder data til deres forpligtelser . Den anvendelse og udnyttelse af disse data kan være af stor interesse både for PA'erne selv og til industrien.

Anvendelse af disse data af offentlige administrationer kombineret med dataindvinding og big data-teknikker og naturlige sprogbehandlingsteknologier kan lette offentlig beslutningstagning og effektiviteten af de offentlige politikker.

Denne rammeaftale ønsker at være en henstilling til medlemsstaterne om implementering af åbne licenser, der fjerner mulige juridiske hindringer for at indarbejde disse typer data i oprettelsen af ny viden.

Det er blevet en aktuel praksis for forskellige regeringer, virksomheder og internationale organisationer, der producerer offentlige data til at stole på disse licenser -i særdeleshed PDDL-, GPL- og Creative Commons (CC)-modaliteterne, hvor alle rettigheder til databaser er frafaldt- som et passende mekanisme til at favorisere anvendelse og genbrug af disse store datamængder.

## 2. Hvorfor en europæisk rammeaftale for styring af datadeling .

*Årsager til rammeaftalen for deling af data fra oversættelseshukommelser.*

### 2.1. Generel tilgang.

FATDS (Framework **Agreement** on **TMX** Data **Sharing**) er hverken tilfældig eller rent episodisk eller punktlig. Tværtimod reagerer det på et behov, der finder forskellige motiveringer. Europæiske offentlige administrationer genererer tusinder af oversættelser hver dag, og disse oversættelser genererer databaser, hvis frigivelse som data allerede er et lovligt mandat (Open Data Europe-direktivet).

Der er to punkter af stor relevans, som afspejler manglen på en grundlæggende konsensus for passende regulering og den forvaltning af problemerne i forbindelse med deling af de data, der genereres af oversættelser af tekster , der kommer fra de offentlige myndigheder. Dette er **en af de sektorer, der genererer den største mængde data** af denne art under deres daglige forretning. Det er også den sektor, der investerer mest i datalagringsinfrastrukturer over hele verden . Vi fokuserer især på de data, der genereres af oversættelseshukommelser fra dokumenter, der ejes af den offentlige administration, i betragtning af de administrerede behov og de mange sprog, ikke kun inden for et territorium eller et land, men i alle de lande, der udgør EU.

Det er meget vigtigt at følge anbefalingerne for data genbrug og interoperabilitet i forvaltningslovens sektor. Lige så vigtigt er fastlæggelsen af betingelserne for genbrug og interoperabilitet .

Uden tvivl ville den optimale tilstand være, at dataene er fuldt tilgængelige, og at deres genbrug ikke kræver specifikke tilladelser, som er fastlagt i det europæiske direktiv om åbne data.

Men for at tilgængelighed og genbrug kan ske, er visse betingelser påkrævet , og det er derfor det er nødvendigt med licenser - i dette tilfælde oversættelseshukommelser bør være så åben som muligt. Det er vores prioriterede mål at etablere en ramme for handling, der favoriserer dens større formidling og kapacitet til genbrug,

Det er mere end tilrådeligt ikke at generere flere typer af licenser, men tværtimod, at forenkle de typer som anbefalet af *den europæiske digitale dagsorden* (Europa -Kommissionen, 2010).

I øjeblikket er der to vigtige internationale alternativer, der fremmer brugen af licenser i forbindelse med data og information, og som regulerer og fremmer den frie adgang og brug af information. Disse licenser er Creative Commons (CC), Apache 2 og GPL og ODC. Creative Commons gælder for både data og andre dokumenter, mens sidstnævnte kun anvendes i dataområdet. Den internationale tendens indikerer, at der til regulering af betingelserne for adgang og brug af dataene anvendes åbne licensmodeller, der hovedsageligt er baseret på CC, skønt med variationer tilpasset egenskaberne for hver dataportal og hvert land, hvor de anvendes.

Alt dette er til fordel for en stadig mere digital administration, der er i stand **til at måle virkningen af dens handlinger med hensyn til socialt afkast af investeringer**, hvilket er målet med handlinger som den nationale og europæiske centrale oversættelseshukommelse (NECTM) gennem dette rammeaftale for deling af data genereret af oversættelseshukommelser fra tekster oversat for offentlige myndigheder.

Eftersom der i øjeblikket ikke er enighed om, hvordan man kan regulere behandlingen, den brug og den deling af data fra oversættelseshukommelser

genereret som følge af offentlige myndigheders offentliggørelsesbehov , enten en afståelsesmulighed kunne foreslås med henvisning til specifikke eksisterende love af generel karakter i hver medlemsstat . Det bedre alternativ er en bestemt forskrift for dette problem , i form af en reguleringsmodel hermed foreslået og baseret på "bløde standarder", af en ensidig , ikke-bindende karakter, og en praktisk tilgang såsom en rammeaftale.

FATDS tager parti for begge sider, der er involveret i processen, PA og leverandørerne. For det første, det gør det klart, at datadelingsproblemet med oversættelseshukommelser eksisterer, at det er virkeligt for de organisationer, der er involveret , og at det er nødvendigt at gribe ind, mens udskjelser og misbrug afvises. For det andet, FATDS forslår en supplerende og specifik ramme for regulering, samt en systematisk styring, som også er fleksibel, tilpasses til alle myndigheder, private enheder og medlemsstater.

Til dette formål, er det stærkt anbefalet og nødvendig af hensyn til et samfund, der i stigende grad er bekymret med værdien af de data, det genererer, at de data stilles til rådighed for den ansvarlige entreprenør eller et centralt organ i hver medlemsstat, med hver af de oversættelsestjenestekontrakter, der udføres, hvilket ikke kun gør teksten eller dokumenterne til den pågældende oversættelse tilgængelig, men også den oversættelseshukommelse, der genereres som et resultat af processen .

Ligeledes takket være implementeringen af NEC TM-databasen etableres der derfor betingelser for at lette genbrug af disse oversættelseshukommelser takket være deres tilgængelighed, og frem for alt, at de ikke vil være underlagt tekniske eller juridiske begrænsninger, der begrænser eller hindrer sådan genbrug.

### **3. Multidimensionalitet af problemet med datadeling af oversættelseshukommelser genereret af offentlige administrationer: beskyttelse, der tilbydes af FATDS .**

I snæver forstand af dets juridiske karakter, har denne rammeaftale til formål at være en aftale, hvis implementering vil finde sted i overensstemmelse med den

praksis og procedurer de enkelte medlemsstater oplever og under hensyntagen til praksis, der allerede er fælles og ikke ukendt for leverandører i oversættelsesbranchen.

På baggrund af ovenstående betragtninger synes det klart, at det ikke er muligt at anerkende rammeaftalerne med en direkte retlig effektivitet, som om det blot var et andet stykke af fællesskabslovgivningen, i dette tilfælde fra en autonom kilde, men det er heller ikke muligt at undervurdere dens effektivitet. Aftalen er af rent frivillig overholdelse og er strengt afhængig af underskrivelsespersoners magt til at binde organisationerne.

Således FATDS forstået som en selvstændig aftale har til formål at et være temmelig praktisk redskab, det vil sige, et sæt vejledende retningslinjer, der angiver effektive regler for forvaltning, centralisering og deling af data, rammeaftalen er strengt supplerende til fællesskabslovgivningen og national lovgivning.

På denne måde, kan en ikke-normativ aftale, såsom FATDS, udgøre, og herved udgøre, et nyttigt instrument for præciseret og afklaret fortolkning af deling af data fra de oversættelseshukommelser der er skabt som følge af det udførte arbejde af budgivere / leverandører.

FATDS's funktion er derfor at klarlægge og specificere den offentlige administrations ret til at anmode om alle de data, der er genereret i oversættelseskontrakter, deres ejerskab af de originale tekster og oversættelse som en tjeneste, der er indgået og afledt af en original, hvilket afspejler de involverede aktørers konsensus, forståelse for de involverede på den ene side budvirksomhederne (sælgerne) og på den anden side det administrative organ fra hver medlemsstat og endelig det centrale organ i EU, der vil blive bestemt til dette formål.

I denne forstand skal det huskes, at den mest akkrediterede konfiguration af denne model eller juridiske paradigme illustrerer eksistensen af instrumenter, der uden at overholde traditionelle typer juridiske normer heller ikke kan udelukkes fra lovens verden og derfor fra obligatorisk karakter for at udtrykke

en fast forpligtelse fra de underskrivende fagpersoner, der udfører tjenesteydelseskontrakt i udøvelsen af deres kompetencer efter en lovligt forudset procedure, idet de anvender et instrument udstyret med en minimal eller relativ juridisk karakter.

Det grundlæggende indhold af FATDS fokuserer mere på etableringen af modeller og retningslinjer, der danner et typisk instrument for reflekterende lovgivning.

I sidste ende er dette paradigme af regulering, der er materialiseret af FATDS , indrammet fladt inden for de nye tilgange og nye metoder til regulering og handling, der i øjeblikket er foreslået, samt af EU, der i højere grad er afhængig af instrumenter, der omdanner den juridiske regel til praksis snarere end at etablere flere regler. Hvilket i vores tilfælde definerer adgang, brug og mulig genbrug af oplysningerne indeholdt i databaserne reguleret ved lov samt gennem kontrakter og / eller licenser.

EØF's offentliggørelsen af meddelelsen om åbne data og Europa-Kommissionens forslag til ændring udgør en politik for åbning af data og fremme af et informationsmarked, der har sit centrale punkt i godkendelsen af Europa-Parlamentets direktiv 2003/98 / EF og af Rådet fra 17. november 2003 om genbrug af information fra den offentlige sektor.

Til dette må vi tilføje revisionen af direktivet om genanvendelse af offentlig sektorinformation, der blev indført i slutningen af 2011, som er en del af den europæiske digitale dagsorden, et initiativ, der er iværksat af Europa-Kommissionen (2010), idet denne dagsorden er rettet mod begunstiggelse af onlinetjenester inden for Unionen og har som en prioritet åbning af offentlige data til genbrug, forenkling af licenssystemet til udveksling af indhold og implementering af interoperabilitetsstandarder.

## **5. FATDS' reguleringsmulighed**

Offentlige data, der besiddes af offentlige organer i Den Europæiske Union, er underlagt særlig behandling i henhold til direktiverne om offentlig genbrug af informationssektorer fra 2003 og 2013, som indeholder bestemmelser om, at PA-dokumenter skal stilles til rådighed for offentligheden til genbrug "ethvert indhold uanset medium (skrevet på papir eller opbevaret i elektronisk form eller som en lyd, visuel eller audiovisuel optagelse) opbevares af offentlige organer til kommercielle eller ikke-kommercielle formål ". Derfor er det generelle princip den frie tilgængelighed af sådant indhold. Hvis der anvendes en told, skal det "være begrænset til de marginale omkostninger, der er afholdt for dets reproduktion, tilgængeliggørelse og formidling" (artikel 6.1).

Kort sagt afsløres FATDS grundlæggende som en "anbefaling" baseret på en minimal eller relativ juridisk handling, baseret på eksisterende servicekontrakter og data (oversættelseshukommelser) genereret ex profeso. Denne "anbefaling" er den praktiske specifikation af pligten til at dele data, der er genereret i form af oversættelseshukommelser, der stammer fra oversættelsestjenester, der er indgået af offentlige administrationer, uanset deres medium eller format.

## **6. Rammeaftalens indhold: styringssystemet til overførsel af data fra oversættelseshukommelserne, der er fremmet af aftalen.**

En oversættelseshukommelse er en sprogdatabase, som løbende lagrer parallelle oversættelser genereret af fagfolk for at være i stand til at bruge dem i fremtiden med henblik på at opnå større terminologisk og stilistisk konsistens, samt besparelser som følge af hel eller delvis matches mellem nye tekster og gamle oversættelser, der allerede er afsluttet.

**Oversættelseshukommelser er digitale opbevaringssteder, der består af tekstlinjer fra indholdet på originalsproget, der er på linje med dets oversættelse til andre sprog.** Disse tekster kan også tilpasses effektivt af oversættelsesenheder. De oversættelsesenheder, der er gemt sammen med deres ækvivalenter, er defineret på forskellig måder (ved sætning, afsnit, ord eller gruppe af ord, etc.). Segmentering efter et tegnsætningstegn markerer slutningen af sætningen eller et paragraf afkast bliver typisk den

hyppigste separator som standard i miljøet i computerassisterede oversættelsessystemer.

Hovedfunktionen ved oversættelseshukommelser (i det følgende TM'er) er at udtrække samlede eller delvise forslag eller matches fra en sætning og konkordanser for termer. Under oversættelse søges segmenter af kildesproget i TM-databasen. Hvis TM har et segment i kildesproget, der matcher nøjagtigt eller delvist, foreslås dette segment til oversætteren sammen med oversættelsen hentet fra den database og eventuelle yderligere oplysninger, der er gemt med segmentet i databasen. Computerassisterede oversættelsesværktøjer (CAT-værktøjer) viser grader af lighed (kendt i branchen som *fuzzy matches*). I oversættelsesbranchen kan disse matches være nyttige for sprogfolk fra et interval, der starter med 50% - 65% op, og der oprettes også prisskalaer for indsatsen, der er involveret i at oversætte helt nye sætninger eller andre, som der er lignende forslag og procentvise matchninger på, inden for samme kontekst. En 100% lighed betragtes som et komplet match mellem en anmodet sætning og en identisk sætning i databasen.

Det gratis format til udveksling af oversættelseshukommelser er TMX (Translation Memory eXchange), generelt i version 1.4b. Dette er en XML-standard af typen DTD (*dokument type definition*). Det blev oprettet af OSCAR-udvalget (Open Standards for Container / Content Allowing Re-use).

Gennem anvendelsen af TMX-formatet er det mere muligt for mennesker eller virksomheder at samarbejde i oversættelsesprojekter. TMX-formatet gør det også nemmere at migrere fra et computerassisteret oversættelsessystem til et andet, som favoriserer konkurrenceevnen mellem de teknologier, der tilbydes, og deres konstante udvikling for at gøre en forskel i forhold til deres konkurrenter. Ligesom andre åbne standarder, blev dette format udviklet med henblik på at reducere kompatibilitetsproblemer, fremme genbrug af sproglige ressourcer, forenkle udveksling af data og dermed stimulere den teknologiske innovation.



Denne rammeaftale har til formål at etablere, som vi har defineret tydeligt, en "ramme for god praksis" til at få data genereret i oversættelsestjenestekontrakter af europæiske offentlige administrationer, dens organisering gennem implementeringen af NEC TM-software systemet og datacentraliseringsinitiativer udpeget af Europa-Kommissionen med henblik på vedtagelse af en protokol til centralisering af tosprogede data fra offentlige administrationer, dens anvendelse og fordel for samfundet generelt og medlemsstaterne selv, hjælp til oprettelse af et korpus af nationale Big Data og deling af de dele, som hver nationale administration finder relevant på europæisk niveau.

Med henblik herpå forvaltes centraliseringen af data, der er hostet på nationale computerinfrastrukturer, af det kompetente organ i den pågældende stat - for eksempel det specifikke statssekretariat for digital fremskridt - eller det kompetente ministerium i den pågældende medlemsstat. Disse data kan deles i tilfælde af tiltrædelse af denne rammeaftale og, hvis det er bestemt af den kompetente myndighed, på et højere niveau med det relevante organ i Europa-Kommissionen (f.eks. ELRC-Share).

Når det er sagt, vil det første af trinnene være centraliseringen af dataene fra oversættelsestjenestekontrakterne genereret af de offentlige administrationer på nationalt plan gennem vedtagelsen af denne aftale, og derefter, og i overensstemmelse hermed, deres deling med et paneuropæisk organ til fordel for PA'er i de andre EU-medlemsstater. Brug af "National og European Central Translation Memory" (NEC TM)-softwaren giver input / output (I/O)-forbindelsen til administrationer, som vil være i stand til privat at gemme deres oversættelsestjenestekontrakter, hoste på en central server og dele dem sammen med dem, der beslutter at gøre deres arbejde mere effektivt og omkostningsbesparende på nationalt plan (typisk interne oversættere eller eksterne leverandører), oprette nationale Big Data og vælge de data, de ønsker, med den centrale server, hvor den europæiske centrale oversættelsestjenestekontrakt befinder sig.

#### Forhold til private enheder og tildeling af TMX

Oversættelsestjenestearbejdet, der udføres af budfirmaer (leverandører), vil medføre levering af de oversættelsestjenestekontrakter, der genereres af dem som

et resultat af levering af deres tjeneste. Til dette formål i meddelelsen, der optræder i Statstidende i den medlemsstat, region eller kommune, det vil sige i entreprenørens portal, bør en klausul indgå, der henviser til NEC TM FATDS , herunder CPV-koden, som vil afgøre, AT DATAENE FRA OVERSÆTTELSESHUKOMMELSEN LEVERES TIL SERVEREN AF DEN KOMPETENTE OFFENTLIGE MYNDIGHED til de formål, der i fremtiden, og hvis den nationale myndighed anser det, kunne deles med det organ, som Kommissionen skal bestemme (ELRC-Share eller lignende initiativ, for eksempel).

**Til dette formål vil de bydende selskaber (leverandører) levere de oversættelseshukommelser genereret sammen med deres arbejde, anvendt af princippet om forbud mod tilbagevirkende kraft, hvilket betyder, at denne rammeaftale ikke må have effekter i fortiden, og at kun denne obligatoriske karakter vil fungere fra vedhæftningen og underskriften af klausulen, hvorved de forpligter sig til at levere, sammen med det udførte arbejde, de parallelle genererede data ved deres oversættelser (TMX eller lignende kompatibelt format) .**

## **7. BETINGELSER FOR BRUG AF INTELLEKTUEL EJENDOM OG DATABASE**

Med hensyn til intellektuel ejendomsret kan der ikke være nogen tvivl om, at de data, der genereres i udførelsen af arbejdet i overensstemmelse med den kontrakt, som den pågældende oversættelse er tildelt for, udelukkende er oplysningerne fra den offentlige administration, der er forfatter eller manager af den originale tekst, der er genstand for oversættelsen.

Faktum kan ikke ignoreres, at også i den udstrækning informationerne stammer fra de offentlige administrationer, skal reguleringen af genbrug af offentlig information og åbningen af offentlige data huskes.

Således tildeler den vellykkede bud (sælger) udelukkende og uden nogen tids- eller territoriale grænser rettighederne til enhver form for dokumentation eller data, der genereres uafhængigt af deres support eller

format, forståelse af distribution og transformation i denne overførsel af reproduktionsrettigheder.

Derfor kan de data, der genereres af sådanne oversættelseshukommelser, ikke bruges af den vellykkede tilbudsgiver til noget lukrativt formål og må kun bruges til at give information, hvis dette er tilfældet for det udførte arbejde.

#### Betingelse for brug af intellektuel ejendomsret og databaser.

DGT-TM-databasen er Europa-Kommissionens eksklusive ejendom. Kommissionen tildeler gratis og på verdensbasis i hele beskyttelsesperioden for disse rettigheder sine ikke-eksklusive rettigheder til genbrugere til alle typer anvendelser, der opfylder betingelserne i Kommissionens beslutning af 12. december 2011 om genbrug af Kommissionens dokumenter, der er offentliggjort i Den Europæiske Unions Tidende L 330 af 14. december 2011, side 39 til 42.

Enhver genbrug af databasen eller de strukturerede elementer deri skal identificeres af genbrugeren, der er forpligtet til at angive kilden til de anvendte dokumenter: webstedets adresse, datoen for den sidste opdatering og det faktum, at Europa-Kommissionen bevarer ejerskabet af dataene.

Denne database er derfor optimal til oprindeligt at udfylde NEC TM-versionen af en medlemsstat.

## **8. DATA BESKYTTELSE. ANVENDELSE AF FORORDNING EU 2016/679 OG FORORDNING 2018/1725**

I relation til personoplysninger, der skal behandles på grundlag af kontraktens opfyldelse af den offentlige administration og den vellykkede tilbudsgiver, er begge parter forpligtet til at overholde nedenstående almindelige forskrifter for databeskyttelse.

Europa-Parlamentets og Rådets forordning (EU) 2016/679 af 27. april 2016 samt overholdelse af den nuværende statslige lovgivning om databeskyttelse i

hver af de medlemsstater, der vedtager denne rammeaftale, samt forordningen om Europa-Parlamentet og Rådet af 23. oktober 2018 om beskyttelse af enkeltpersoner med hensyn til behandlingen af personoplysninger fra Unionens institutioner, organer, kontorer og agenturer og om distribution af sådanne data.

## **ENDELIGE KONKLUSIONER .**

Den offentlige administration som promotor for sprogteknologisektoren med oprettelse af fælles platforme til naturlig sprogbehandling og automatisk oversættelse og udvikling af ressourcer til genbrug af offentlig sektorinformation (RISP), er forpligtet til at udvikle datadelingspolitikker og lægge fundamentene for, at denne deling er reel og effektiv blandt alle de enheder, der deltager i processen, sikrer også, at deltagere i processen uden for administrationen erhverver forpligtelsen til at underskrive klausulen om accept af deres bud, og påtage sig at levere sammen med deres arbejde, de data, der genereres af deres oversættelser (oversættelseshukommelser), der danner dem alle et sæt sproglige ressourcer af uvurderlig værdi.

**Referencer.** Ferrer-Sapena m.fl. (2011) i deres artikel om adgang til offentlige data; Ferrer-Sapena og Peset (2012) om genbrug af kulturelle data eller Ramos Simón m.fl. (2012) i deres undersøgelse af europæiske dataportaler.

*Licenser til åbne data: En praktisk vejledning* til Highing Education Funding Council for England (Korn og Oppenheim, 2011). Også bemærkelsesværdig er *Vejledningen til Open Data Licensing* (Open Knowledge Foundation, nd) og retningslinjerne fra den europæiske digitale dagsorden (Europa-Kommissionen, 2010).