

Raamleping avaliku halduse dokumentide tõlgete tulemusena loodud tõlkemälu andmete jagamiseks

1. SISSEJUHATUS

Pärast seda, kui 2003. aastal võeti vastu esimesed avaliku sektori teabe taaskasutamist käsitlevad eeskirjad, on kogutud andmete maht kogu maailmas hüppeliselt kasvanud; samal ajal kui luuakse ja kogutakse uut tüüpi andmeid.

Kogu selle aja oleme tunnistanud andmete analüüsimiseks, kasutamiseks ja töötlemiseks mõeldud keeletöötlemistehnoloogiate pidevale arengule.

See kiire tehnoloogiline areng võimaldab luua uusi teenuseid ja rakendusi, mis põhinevad andmete kasutamisel, liitmisel või kombineerimisel.

Suur hulk andmeid või suuri andmeid, mis on kogutud erinevate tehnoloogiliste vahendite kaudu või mis on saadud eri vormingutes suurtest andmekogudest, on võimalised genereerima uusi teadmisi kõige erinevates sektorites, kuid samal ajal tekitavad probleeme nii andmete omandamisel kui ka selle hilisemal kasutamisel.

Riiklikud haldusorganid (PA-d) ning avaliku sektori asutused ja üksused on peamised andmetootjad, ja see on väga kasulik (loomulikku) keelt töötleva tööstuse jaoks. Riiklikud haldusorganid esitavad oma kohustuste täitmisel tohutul hulgal andmeid. Nende andmete kasutamine ja rakendamine võib olla väga huvipakkuv nii PA-le endale kui ka tööstusele.

Nende andmete kasutamine avalike haldusasutuste poolt koos andmekaevandamise ning suurte andmetehnikate ja loomuliku keeletöötlemise tehnoloogiatega võib hõlbustada avaliku sektori otsuste tegemist ja avaliku korra tõhusust.

Käesolev raamleping soovib anda liikmesriikidele soovitusel rakendada avatud litsentse, mis kõrvaldavad võimalikud õiguslikud takistused, et lisada seda tüüpi andmed uute teadmiste loomisse.

Sellest on saanud praegune tava, et eri valitsused, korporatsioonid ja avalikke andmeid tootvad rahvusvahelised organisatsioonid toetuksid nendele litsentsidele (eelkõige PPDL-i, GPL-i ja Creative Commons'i (CC) korrale, milles loobutakse kõikidest õigustest andmebaasidele) kui asjakohasele mehhanismile, mis soosib nende suurte andmemahtude kasutamist ja taaskasutamist.

2. Miks on andmete jagamise haldamiseks vaja Euroopa raamlepingut?

Tõlkemäludest pärit andmete jagamise raamlepingu põhjused.

2.1. Üldine lähenemine

FATDS (**TMX-andmete jagamise raamleping**) ei ole juhuslik ega puhtalt episoodiline ega punktuaalne. Vastupidi, see vastab vajadusele, mis leiab erinevaid motivatsioone. Euroopa riikide haldusorganid loovad iga päev tuhandeid tõlkeid ja need tõlked loovad andmebaase, mille andmete avaldamine on juba seaduslik volitus (Open Data Europe'i direktiiv).

On kaks väga olulist punkti, mis peegeldavad põhilise konsensusi puudumist piisava reguleerimise osas ja haldusasutustest pärinevate tekstide tõlkimisel tekkinud andmete jagamisega seotud probleemide käsitlemist. See on **üks sektoreid**, mis loob oma igapäevases äritegevuses suurima hulga selliseid andmeid. See on ka sektor, mis investeerib kõige rohkem andmesalvestusinfrastruktuuridesse kogu maailmas. Me keskendume eelkõige andmetele, mida tekitavad riiklike haldusorganite valduses olevate dokumentide tõlkemälud, arvestades hallatavate vajadusi, ja arvukatele keeltele mitte ainult ühel territooriumil või riigis, vaid kõigis riikides, mis moodustavad EL-i.

Väga oluline on järgida soovitusi andmete taaskasutamise ja koostalitlusvõime kohta avaliku halduse sektoris. Sama tähtis on ka taaskasutamise ja koostalitlusvõime tingimuste sätestamine.

Kahtlemata oleks optimaalne olukord, kus andmed oleksid täielikult kättesaadavad ja nende taaskasutamine ei nõuaks erilubasid, nagu on sätestatud avatud andmeid käsitlevas Euroopa direktiivis.

Selleks et juurdepääs ja taaskasutamine toimiksid, on siiski vaja teatud tingimusi ning seetõttu on litsentsid vajalikud – antud juhul peaksid tõlkemälud olema võimalikult avatud. Meie esmaeesmärk on luua tegevusraamistik, mis soodustab selle ulatuslikumat levitamist ja taaskasutamise võimet.

Väga soovitatav on mitte luua rohkem litsentse, vaid hoopis vastupidi, lihtsustada tüüpe, nagu on soovitatud Euroopa digitaalarengu tegevuskavas (Euroopa Komisjon, 2010).

Praegu on kaks peamist rahvusvahelist alternatiivi, mis edendavad litsentside kasutamist andmete ja teabe kontekstis ning mis reguleerivad ja edendavad teabe vaba juurdepääsu ja kasutamist. Need litsentsid on Creative Commons (CC), Apache 2 ning GPL ja ODC. Creative Commons kehtib nii andmete kui ka muude dokumentide kohta, samas kui viimaseid kasutatakse ainult andmealal. Rahvusvaheline suundumus näitab, et andmete juurdepääsu ja kasutamise tingimuste reguleerimiseks kasutatakse avatud litsentsimudeleid, mis põhinevad peamiselt CC-l, kuigi variatsioonid on kohandatud iga andmeportaali ja iga riigi omadustega, kus neid kohaldatakse.

See kõik on kasulik üha enam digitaalsele administratsioonile, kes on võimeline **mõõtma oma tegevuse mõju seoses sotsiaalse tasuvusega investeringutele**, mis on selliste meetmete eesmärk nagu riiklik ja Euroopa keskne tõlkemälu (NECTM) käesoleva raamlepingu kaudu, et jagada andmeid, mis on loodud tõlkemälu abil riiklike haldusorganite jaoks tõlgitud tekstidest.

Seetõttu, kuna praegu puudub üksmeel selle üle, kuidas reguleerida avaliku halduse avaldamisvajaduste tõttu loodud tõlkemäludest saadud andmete töötlemist, kasutamist ja jagamist, võiks selle välja pakkuda kas või keeldumise võimalusega, viidates

konkreetsetele olemasolevatele üldistele seadustele igas liikmesriigis. Parem alternatiiv on selle küsimuse jaoks konkreetne regulatsioon, mis on välja pakutud regulatiivse, „pehmetel standarditel” põhineva mudel kujul; see on ühepoolne, mittesiduv ja praktiline lähenemisviis, nagu RAAMLEPING.

FATDS asub mõlema protsessis osaleja, PA ja hankijate poolel. Esiteks on selge, et tõlkemälu andmete jagamise probleem on olemas, et see on asjaomaste organisatsioonide jaoks reaalne ja et on vaja sekkuda, lükates samas kõrvale liialdused ja kuritarvitused. Teiseks pakub FATDS välja täiendava ja spetsiifilise reguleerimise raamistiku, ja ka süstemaatilise juhtimise, mis on samuti paindlik ja kohandatav kõigile haldusasutustele, eraõiguslikele üksustele ja liikmesriikidele.

Ühiskonna, kes tunneb üha enam muret andmete väärtuse pärast, mida ta genereerib, on väga soovitatav ja vajalik, et andmed tehakse kättesaadavaks vastutavale töövõtjale või kesksele organile igas liikmesriigis koos iga sõlmitud tõlketeenuse lepinguga, mis teeb kättesaadavaks mitte ainult kõnealuse tõlketeksti või -dokumendid, vaid ka protsessi tulemusena loodud tõlkemälu.

Samuti kehtestatakse tänu NEC TM andmebaasi rakendamisele tingimused, et hõlbustada nende tõlkemälude taaskasutamist tänu nende juurdepääsetavusele, ning eelkõige ei kohaldata nende suhtes tehnilisi või õiguslikke piiranguid, mis piiraksid või takistaksid sellist taaskasutamist.

3. Riiklike haldusorganite loodud tõlkemälude andmete jagamise probleemi mitmemõõtmelisus: FATDS-i pakutav kaitse

Õiguslikus mõttes on käesoleva raamlepingu eesmärk olla leping, mille rakendamine toimub vastavalt iga liikmesriigi kogemustele ja menetlustele ning võttes arvesse tavaid, mis on juba ühised ja mis ei ole tõlketööstuse hankijatele tundmatud.

Eespool esitatud kaalutluste põhjal tundub selge, et raamlepinguid ei ole võimalik tunnustada otsese õigusliku tõhususega, nagu oleks see vaid järjekordne ühenduse õigusakt (käesoleval juhul autonoomsest allikast), kuid samuti ei ole võimalik selle tõhusust alahinnata. Leping on puhtvabatahtlikust vastavusest, rangelt sõltuv allakirjutanud subjektide võimust organisatsioonide siduda.

Seega on FATDS-i, mida mõistetakse autonoomse lepinguna, eesmärk olla pigem praktiline vahend, see tähendab juhtsuuniste kogum, milles on sätestatud tõhusad eeskirjad andmete haldamiseks, tsentraliseerimiseks ja jagamiseks, ning raamleping täiendab rangelt ühenduse ja siseriiklike õigusakte.

Sel viisil võib mittenormatiivne leping, näiteks FATDS, kujutada endast kasulikku vahendit, mis seisneb pakkujate/hankijate tehtud töö tulemusena loodud tõlkemälude abil loodud andmete jagamise täpsustamises ja selgitamises.

Seetõttu on FATDS ülesanne selgitada ja täpsustada riiklike haldusorganite õigust nõuda kõiki tõlkelepingutes saadud andmeid, nende omandiõigust originaaltekstidele ja tõlget kui teenust, mis on sõlmitud ja tuletatud originaalist, peegeldades osalejate konsensust, mõistes, et need on seotud ühelt poolt pakumise teinud äriühingutega (hankijad) ja teiselt poolt iga liikmesriigi haldusorganiga ning lõpuks EL-i keskasutusega, mis määratakse kindlaks sel eesmärgil.

Sellega seoses tuleb meele pidada, et selle mudeli juriidiline paradigma kõige akrediteeritud konfiguratsioon näitab instrumentide olemasolu, mida ei saa ilma traditsioonilistele õigusnormidele vastamata välja jätta ka õigusmaailmast ja seega kohustuslikust olemusest, et väljendada kindlat pühendumust allakirjutanud subjektidele, kes täidavad teenuste hankelepingut oma pädevuse teostamisel õiguslikult ettenähtud menetluse alusel, kehtestades praktikas minimaalse või suhtelise seaduslikkusega instrumendi.

FATDS-i põhisisu keskendub rohkem mudelite ja suuniste kehtestamisele, moodustades tüüpilise soovitava õigusvahendi.

Lõppkokkuvõttes on see reguleerimise paradigma, mida FATDS realiseerib, raamistatud uute lähenemisviiside ning uute reguleerimise ja meetmete meetodite raames, samuti EL-i poolt, mis tugineb rohkem sellistele vahenditele, mis muudavad õigusnormi praktiliseks, mitte ei kehtesta rohkem eeskirju, määratledes meie puhul seadusega reguleeritud andmebaasides sisalduva teabe kättesaadavuse, kasutamise ja võimaliku taaskasutamise, samuti lepingute ja/või litsentside kaudu.

EMÜ on avatud andmeid käsitleva teatise ja Euroopa Komisjoni muutmissetepaneku avaldamine poliitika, mille eesmärk on avada andmeid ja edendada teabeturgu, mille keskne punkt on Euroopa Parlamendi ja nõukogu 17. novembri 2003. aasta direktiivi 2003/98/EÜ (avaliku sektori teabe taaskasutamise kohta) heakskiitmine.

Selleks peame lisama 2011. aasta lõpus kasutusele võetud avaliku sektori teabe taaskasutamist käsitleva direktiivi läbivaatamise, mis on osa Euroopa digitaalarengu tegevuskavast, Euroopa Komisjoni algatatud algatusest (2010), mille eesmärk on soodustada võrguteenuseid liidus ning mille prioriteet on avalike andmete taaskasutamiseks avamine, infosisu vahetamise litsentsimissüsteemi lihtsustamine ja koostalitlusvõime standardite rakendamine.

5. FATDS-i regulatiivne valik

Euroopa Liidu avalik-õiguslike asutuste valduses olevaid avalikke andmeid käsitletakse konkreetselt 2003. ja 2013. aasta avaliku sektori teabe taaskasutamise direktiivide alusel, millega nähakse ette, et PA dokumendid tehakse üldsusele kättesaadavaks „mis tahes sisu (mis on kirjutatud paberile või salvestatud elektroonilisel kujul või heli -, visuaal - või audiovisuaalsalvestisena), mida avaliku sektori asutused hoiavad kaubanduslikel või mittekaubanduslikel eesmärkidel” taaskasutamiseks. Seetõttu on üldpõhimõtteks sellise sisu vaba kättesaadavus; kui kohaldatakse tariife, „piirdatakse selle reprodutseerimise, kättesaadavaks tegemise ja levitamise piirkuludega” (art. 6,1).

Lühidalt, FATDS avaldub põhimõtteliselt „soovitusena”, mis põhineb minimaalsel või suhtelisel juriidilisusel, tuginedes olemasolevatele teenuslepingutele ja *ex profeso* genereeritud andmetele (tõlkemälud). Käesolev soovitus on riiklike haldusorganite sõlmitud tõlketeenustest pärinevate tõlkemälude kujul loodud andmete jagamise kohustuse praktiline kirjeldus, olenemata nende andmekandjast või vormingust.

6. Raamlepingu sisu: lepinguga edendatud tõlkemäludest andmete edastamise haldussüsteem

Tõlkemälu on keeleandmebaas, mis talletab pidevalt professionaalide loodud paralleeltõlkeid, et neid tulevikus kasutada, saamaks suuremat terminoloogilist ja stiililist kooskõla, samuti sääste, mis on tingitud täielikest või osalistest vastetest uute tekstide ja juba lõpetatud vanade tõlgete vahel.

Tõlkemälud on digitaalsed hoidlad, mis koosnevad algkeele sisu tekstiridadest, mis on joondatud selle tõlkega teistes keeltes. Neid tekste saab tõhusalt joondada ka tõlkeüksustega. Tõlkeühikud, mida säilitatakse koos nende ekvivalentidega, määratletakse erinevalt (fraasi, lõigu, sõna või sõnarühma vms järgi), segmenteeritakse pärast kirjavahemärki, mis tähistab lause lõppu või lõigu tagastust, mis on arvutipõhiste tõlkesüsteemide keskkonnas tavaliselt kõige sagedasem eraldaja.

Tõlkemälude (edaspidi „TM-id”) põhifunktsioon on eraldada lausest ja terminitest täielikud või osalised soovitused või vasted. Tõlkimise käigus otsitakse TM-i andmebaasist lähtekeele segmente. Kui TM-il on lähtekeeles segment, mis vastab täpselt või osaliselt, soovitatakse seda segmenti tõlkijale koos sellest andmebaasist toodud tõlkega ja mis tahes lisateabega, mis on salvestatud andmebaasi segmendiga. Arvutipõhised tõlketööriistad (CAT-tööriistad) näitavad sarnasuse astet (tööstuses tuntud kui *hägusad* vasted). Tõlketööstuses võivad need vasted (vahemikus 50–65%) olla kasulikud lingvistidetele ja hinnaskaalad on kehtestatud ka selleks, et tõlkida täiesti uusi fraase või teisi, mille kohta on sarnased soovitused ja protsentuaalsed vasted, samas kontekstis.

Tmx-Andmete jagamise raamleping (FATDS) - avaliku halduse dokumentide tõlkimise käigus loodud andmed. NEC TM PROJEKT.

100% sarnasuse astet peetakse täielikuks vasteks taotletud lause ja andmebaasis oleva identse lause vahel.

Tõlkemälude vahetamise tasuta formaat on TMX (Translation Memory eXchange), üldjuhul versioonis 1,4 b. See on XML-standard tüübiga DTD (*dokumenditüübi* definitsioon). Selle lõi OSCAR-i komitee (Open Standards for Container / Content Allowing Re-use).

TMX-formaadi rakendamise kaudu on inimestel või ettevõtetel lihtsam tõlkeprojektides koostööd teha. TMX-vorming lihtsustab ka ühest arvutipõhisest tõlkesüsteemist teise siirdumist, mis soodustab konkurentsivõimet pakutavate tehnoloogiate ja nende pideva arengu vahel, et oma konkurentide suhtes vahet teha. Sarnaselt teiste avatud standarditega töötati see vorming välja selleks, et vähendada ühilduvusküsimusi, edendada keeleliste ressursside taaskasutamist, lihtsustada andmevahetust ja seega ergutada tehnoloogilist innovatsiooni.

Käesoleva raamlepingu eesmärk on luua, nagu me oleme hästi määratlenud, „heade tavade raamistik”, et saada Euroopa riikide haldusorganite tõlketeenuste lepingutes genereeritud andmeid, selle korraldust NEC TM-i tarkvarasüsteemi rakendamise kaudu ja Euroopa Komisjoni määratud andmete tsentraliseerimise algatusi, et võtta vastu protokoll kakskeelsete andmete tsentraliseerimise kohta riiklike haldusorganite poolt, võimaldada selle kasutamist ja kasu ühiskonnale üldiselt ja liikmesriikidele endile, et aidata kaasa suurte riiklike andmete koondamisele ja jagamisele liikmesriikide tasandil.

Selleks haldab riiklikes arvutitaristutes hoitavate andmete tsentraliseerimist kõnealuse riigi pädev asutus (näiteks spetsiaalne digitaalarengu riiklik sekretariaat) või asjaomase liikmesriigi pädev ministeerium. Neid andmeid võib jagada käesoleva raamlepinguga ühinemise korral ja kui pädev asutus on selle kindlaks määranud, kõrgemal tasandil Euroopa Komisjoni asjaomase organiga (nt ELRC-Share).

Olles seda öelnud, on esimeseks sammuks riiklike haldusorganite poolt käesoleva lepingu vastuvõtmise kaudu loodud tõlkemälu andmete tsentraliseerimine riiklikul tasandil ning pärast seda ja kooskõlas sellega nende jagamine üleeuroopalise organiga teiste EL-i liikmesriikide PA-de hüvanguks. Tarkvara „National and European Central Translation Memory” (NEC TM) kasutamine annab sisendi (väljundühenduse (I/O)) haldusasutustele, kes saavad oma tõlkemälu privaatselt talletada, majutada keskserveris ja jagada nendega, kes otsustavad muuta oma töö riiklikul tasandil tõhusamaks ja kulusäästlikumaks (tavaliselt majasisesed tõlkijad või välised hankijad), luua riiklikud suured andmed ja valida andmed, mida nad soovivad keskserverist, kus asub Euroopa keskne tõlge.

Suhe eraõiguslike üksustega ja TMX-i määramine

Pakkumisettevõtete (hankijate) teostatav tõlketöö toob kaasa nende poolt oma teenuse osutamise tulemusena loodud tõlkemälude pakkumise. Sel eesmärgil tuleks liikmesriigi, piirkonna või omavalitsuse ametlikus väljaandes, st töövõtjaportaalis sisalduvas teates lisada klausel, mis viitab NEC TM FATDS-i, sealhulgas CPV tähisele, millega määratakse kindlaks, et TÕLKEMÄLU ANDMED ESITATAKSE PÄDEVALE RIIGIASUTUSELE eesmärgil, et riiklik asutus võib neid jagada Euroopa Komisjoni või samalaadse algatusega (nt ELRC-Share).

Sel eesmärgil edastavad pakkumise teinud äriühingud (hankijad) koos oma tööga loodud tõlkemälud, kohaldades tagasiulatuva jõu puudumise põhimõtet, mis tähendab, et käesolev raamleping ei tohi ajas tagasi mõjuda ning et ainult see kohustuslik iseloom toimib selle klausli haaramise ja allkirjastamise tõttu, millega nad kohustuvad koos tehtud töödega esitama nende tõlgete abil loodud paralleelseid andmeid (TMX- või sarnane ühilduv vorming).

7. INTELLEKTUAALOMAND JA ANDMETE KASUTAMISE PÕHITINGIMUSED

Intellektuaalomandiga seoses ei saa olla kahtlust, et andmed, mis on saadud lepingu kohaselt tehtud töö käigus, mille jaoks kõnealune tõlge on sõlmitud, on täielikult ja

eranditult avaliku halduse andmed, mille autor on tõlke objektiks oleva originaalteksti autor või haldaja.

Ei saa eirata asjaolu, et kuivõrd suurem osa ajast pärineb info avalikest haldusasutustest, tuleb silmas pidada ka avaliku teabe taaskasutamise reguleerimist ja avalike andmete avamist.

Seega määrab edukas pakkuja (hankija) ainuisikuliselt ja ilma mis tahes aja- või territoriaalsete piiranguteta mis tahes liiki dokumentide või andmete õigused, mis on saadud sõltumatult nende toetusest või vormingust, mõistmaks levitamist ja ümberkujundamist selles reprodutseerimisõiguste üleandmisel.

Järelikult ei tohi edukas pakkuja selliste tõlkemälude abil saadud andmeid kasutada ühelgi tulutooval eesmärgil ja neid võib kasutada teabe esitamiseks ainult juhul, kui see on tehtud töö puhul nii.

Intellektuaalomandi ja andmebaaside kasutustingimused

DGT-TM andmebaas on Euroopa Komisjoni ainuomand. Komisjon annab nende õiguste puhul kogu perioodi jooksul tasuta ja ülemaailmse kasutusõiguse taaskasutajatele kõikide kasutusviiside jaoks, mis vastavad komisjoni 12. detsembri 2011. aasta otsuses (komisjoni dokumentide taaskasutamise kohta, avaldatud Euroopa Liidu Teatajas L 330, 14. detsember 2011, lk 39–42) sätestatud tingimustele.

Andmebaasi või selles sisalduvate struktureeritud elementide mis tahes taaskasutamise peab kindlaks tegema taaskasutaja, kes on kohustatud osutama kasutatud dokumentide allikale: veebisaidi aadressile, viimase ajakohastamise kuupäevale ja asjaolule, et Euroopa Komisjon säilitab andmete omandiõiguse.

Seetõttu on see andmebaas optimaalne, et algselt sobida liikmesriigi NEC TM versiooni.

8. ANDMEKAITSE. MÄÄRUSE EL 2016/679 JA MÄÄRUSE 2018/1725 KOHALDAMINE

Isikuandmete puhul, mida riiklik haldusorgan ja edukas pakkuja peavad lepingu täitmise alusel töötleva, on mõlemad pooled kohustatud järgima allpool esitatud andmekaitse üldmäärsi.

Euroopa Parlamendi ja nõukogu 27. aprilli 2016. aasta määrus (EL) 2016/679, samuti käesoleva raamlepingu vastu võtnud liikmesriikide andmekaitset käsitlevate kehtivate siseriiklike õigusaktide ning Euroopa Parlamendi ja nõukogu 23. oktoobri 2018. aasta määruse (üksikisikute kaitse kohta isikuandmete töötlemisel liidu institutsioonides, asutustes, ametites ja asutustes ning selliste andmete ringluse kohta) järgimine.

LÖPPJÄRELDUSED.

Riiklikud haldusorganid kui keeletehnoloogiate valdkonna edendajad on kohustatud välja töötama andmejägamispoliitika ning seadma selle jagamise aluse, mis muudaks andmete jagamise reaalseks ja tõhusaks kõigi protsessis osalevate üksuste vahel, tagades ühtlasi, et väljaspool administratsiooni toimivas protsessis osalejad omandavad kohustuse allkirjastada oma pakkumuste aktsepteerimise klausli, kohustudes esitama selle koos oma töödega, edastama nende loodud tõlgetest (tõlkemäludest) saadud andmed, mis kokku moodustavad hindamatu väärtusega keeleressursside kogumi.

VIITED. Ferrer-Sapena ja teised (2011) oma artiklis juurdepääsu kohta avalikele andmetele; Ferrer-Sapena ja Peset (2012) kultuuriandmete taaskasutamise kohta või Ramos Simón ja teised (2012) oma uuringus Euroopa andmeportaalide kohta.

Avatud andmete litsentsimine: Inglismaa kõrghariduse rahastamise nõukogu praktiline juhend (Korn ja Oppenheim, 2011). Märkimisväärsed on ka *avatud andmete litsentsimise juhend* (Open Knowledge Foundation, n.d.) ja *Euroopa digitaalarengu tegevuskavast* (Euroopa Komisjon, 2010) tulenevad suunised.