

Okvirni sporazum o izmenjavi podatkov iz prevajalskih pomnilnikov, ustvarjenih pri prevajanju dokumentov javne uprave.

1. UVOD

Od sprejetja prvega sklopa pravil o ponovni uporabi podatkov javnega sektorja leta 2003 se je obseg ustvarjenih podatkov po vsem svetu znatno povečal, ustvarjajo in zbirajo pa se tudi nove vrste podatkov.

Obenem smo priča nenehnemu razvoju tehnologij obdelave jezika za namen analize, uporabe in obdelave podatkov.

Ta hiter tehnološki razvoj omogoča oblikovanje novih storitev in aplikacij na podlagi uporabe, zbiranja ali združevanja podatkov.

Ogromne količine podatkov oz. masovni podatki, zbrani prek različnih tehnoloških orodij ali prevzeti iz velikih zbirk informacij v različnih formatih, omogočajo oblikovanje novega znanja v najrazličnejših sektorjih, vendar obenem povzročajo težave v smislu ugotavljanja lastništva nad podatki in njihove naknadne uporabe.

Javne uprave (JU), organi in subjekti, ki sestavljajo javni sektor, so največji proizvajalci podatkov, ki imajo v panogi obdelave (naravnega) jezika izjemno uporabno vrednost. Javne uprave pri izvajanju svojih dolžnosti proizvajajo ogromne količine podatkov. Uporaba in izkoriščanje teh podatkov sta močno v interesu tako JU kot panoge.

V primeru javne uprave lahko uporaba teh podatkov v kombinaciji s podatkovnim rudarjenjem in tehnologijami obdelave masovnih podatkov ter tehnologijami za obdelavo naravnega jezika olajša javno odločanje in učinkovitost javnih politik.

Ta Okvirni sporazum je namenjen državam članicam kot priporočilo za uvedbo odprtih licenc, ki z namenom vključevanja teh vrst podatkov v oblikovanje novega znanja odpravljajo morebitne pravne ovire.

Različne vlade, korporacije in mednarodne organizacije, ki ustvarjajo podatke javnega značaja, te licence v praksi uporabljajo kot mehanizem za uporabo in ponovno uporabo velikih količin podatkov – zlasti to velja za metode PPDL, GPL in Creative Commons (CC), pri katerih so vse pravice do podatkovnih baz odvzete.

2. Zakaj Evropski okvirni sporazum za urejanje izmenjave podatkov

Razlogi za nastanek Okvirnega sporazuma za izmenjavo podatkov iz prevajalskih pomnilnikov

2.1. Splošni pristop

Sporazum FATDS (Framework **Agreement** on **TMX** Data **Sharing**) ni niti naključen niti povsem občasen ali časovno povsem določen. Ravno nasprotno: odziva se na potrebo, ki izhaja iz številnih motivov. Evropske javne uprave vsak dan opravijo na tisoče prevodov, ki predstavljajo podlago podatkovnim bazam, katerih objava v obliki podatkov je že zakonsko določena (evropska Direktiva o odprtih podatkih).

Zlasti pomembni sta dve točki, ki odražata pomanjkanje osnovnega soglasja o ustrezni zakonski ureditvi in obravnavi problematike izmenjave podatkov, ustvarjenih s prevodi besedil javnih uprav. To je **eden od sektorjev, ki ustvari največjo količino podatkov** te vrste pri vsakodnevnem poslovanju. Ta sektor pa tudi največ vlaga v infrastrukture za shranjevanje podatkov po vsem svetu. Zlasti se osredotočamo na podatke, ustvarjene s prevajalskimi pomnilniki dokumentov v lasti javne uprave v skladu s potrebami upravljanega subjekta, in številne jezike ne samo znotraj določenega ozemlja ali države, pač pa v vseh državah EU.

Nadvse pomembno je upoštevati priporočila za ponovno uporabo in interoperabilnost podatkov v sektorju javne uprave. Prav tako pomembno je določiti pogoje za ponovno uporabo in interoperabilnost.

Brez dvoma bi bilo optimalno, če bi bili podatki v celoti dostopni in če za njihovo ponovno uporabo ne bi potrebovali posebnih dovoljenj, kot to ugotavlja evropska Direktiva o odprtih podatkih.

Vendar pa so za omogočanje dostopnosti in ponovne uporabe zahtevani določeni pogoji, zato pa so potrebne licence – v tem primeru bi morali biti prevajalski pomnilniki čim bolj odprti. Naš primarni cilj je oblikovati okvir za ukrepanje, ki bi omogočilo večje razširjanje in možnost ponovne uporabe.

Ustvarjanje več vrst licenc ni priporočljivo; ravno nasprotno, vrste bi poenostavili tako, kot to priporoča *Evropska digitalna agenda* (Evropska komisija, 2010).

Na mednarodni ravni trenutno obstajata dve glavni alternativni, ki na področju podatkov in informacij spodbujata uporabo licenc ter urejata in spodbujata prost dostop ter uporabo informacij. Gre za licence Creative Commons (CC), Apache 2, GPL in ODC. Licenca Creative Commons se nanaša na podatke in druge dokumente, medtem ko se slednje tri uporabljajo samo na področju podatkov. Na mednarodni ravni se trend nagiba k določanju pogojev dostopa in uporabe podatkov na podlagi modelov odprtih licenc, večinoma licence CC, čeprav obstajajo različice, prilagojene značilnostim posameznega podatkovnega portala in države, v kateri se uporabljajo.

Vse to govori v prid vedno bolj digitalni upravi, ki je zmožna **meriti učinek lastnih ukrepov v smislu družbene donosnosti naložbe**, kar je cilj ukrepov, kot je npr. Nacionalni in evropski centralni prevajalski pomnilnik (NEC TM), določenih s tem Okvirnim sporazumom o izmenjavi podatkov, ustvarjenih s prevajalskimi pomnilniki besedil, prevedenih za javne uprave.

Ker trenutno ni soglasja o tem, kako urediti obdelavo, uporabo in izmenjavo podatkov iz prevajalskih pomnilnikov, ustvarjenih kot rezultat potreb objav javne uprave, bi bilo mogoče predlagati vzdržno možnost, ki se bi urejala določene obstoječe zakone splošne

narave v posamezni državi članici. Boljšo možnost pa predstavlja specifična ureditev te problematike v obliki predlaganega zakonodajnega modela, ki temelji na »mehkih standardih« enostranske in nezavezujoče narave, ter praktičnega pristopa, kakršen je OKVIRNI SPORAZUM.

Sporazum FATDS upošteva obe v proces vključeni strani, tako JU kot izvajalce storitev. Prvič, jasno priznava obstoj problematike izmenjave podatkov iz prevajalskih pomnilnikov, s katero se soočajo sodelujoče organizacije in ki terja ukrepe, hkrati pa zavrača čezmernost in zlorabe. Drugič, sporazum FATDS zagotavlja dopolnilen in specifičen okvir urejanja ter systemskega upravljanja, ki ga je mogoče prilagajati vsem upravam, zasebnim subjektom ter državam članicam.

V ta namen je zelo priporočljivo in v interesu družbe, ki je glede vrednosti ustvarjenih podatkov vse bolj zaskrbljena, na podlagi pogodbe o prevajalskih storitvah prepustiti podatke odgovornemu pogodbenemu izvajalcu ali centralnemu organu vsake države članice. Pri tem mu niso na voljo le besedilo oz. dokumenti zadevnega prevoda, temveč tudi prevajalski pomnilnik, ustvarjen kot rezultat procesa.

Prav tako so z uvedbo podatkovne baze NEC TM ustvarjeni pogoji, ki lajšajo ponovno uporabo teh zelo dostopnih prevajalskih pomnilnikov, predvsem pa zanje ne bodo veljale tehnične ali pravne omejitve, ki bi takšno ponovno uporabo omejevale ali ovirale.

3. Večdimenzionalnost problematike izmenjave podatkov iz prevajalskih pomnilnikov, ki jih ustvari javne uprave: zaščita, ki jo ponuja FATDS.

V strogo pravnem smislu je ta Okvirni sporazum mišljen kot sporazum, ki se bo izvajal v skladu s praksami in postopki na temelju izkušenj posamezne države članice in ob upoštevanju že uveljavljenih praks, ki so za izvajalce v prevajalski industriji že običajni.

Na podlagi zgoraj povedanega se zdi jasno, da Okvirnih sporazumov ni mogoče priznati z neposrednim pravnim učinkom, kot če bi šlo za nov del zakonodaje Skupnosti, v tem

primeru iz avtonomnega vira; vseeno pa ni mogoče podcenjevati njegovega učinka. Pristop k sporazumu je povsem prostovoljne narave in strogo odvisen od moči, s katero podpisnik razpolaga pri uveljavljanju zavez k upoštevanju sporazuma pri organizacijah.

Sporazum FATDS je kot avtonomni sporazum zato mišljen kot praktični instrument, tj. skupek smernic, ki določajo učinkovita pravila za upravljanje, centralizacijo in izmenjavo podatkov, in je zgolj namenjen dopolnitvi zakonodaje Skupnosti in nacionalnih zakonodaj.

Sporazum takšne, ne-normativne narave, kakršen je FATDS, predstavlja uporaben instrument za udejanjanje in pojasnjevanje izmenjave podatkov, ustvarjenih s prevajalskimi pomnilniki, kot rezultata dela ponudnikov/izvajalcev storitev.

Namen sporazuma FATDS je zato razjasniti in določiti pravice javnih uprav, da zahtevajo vse podatke, ustvarjene v okviru pogodb za prevajanje, lastništvo nad izvornimi besedili in iz izvornika izpeljanimi prevodi kot storitvijo zunanjega izvajalca, ki odraža soglasje vključenih deležnikov: na eni strani ponudnikov (izvajalcev) storitev, na drugi strani upravnih organov posameznih držav članic ter v ta namen določenega centralnega organa EU.

V tem smislu si velja zapomniti, da najbolj merodajna konfiguracija tega modela oz. paradigme legislativnosti odraža obstoj instrumentov, ki jih brez skladnosti s tradicionalnimi pravnimi standardi ni mogoče izključiti iz pravnega konteksta in ostajajo zato zavezujočega značaja, za izražanje trdne zavezanosti podpisnikov, ki izvajajo pogodbo o storitvah, pri opravljanju svojih pristojnosti v skladu z zakonsko predvidenim postopkom, in pri tem v praksi pomeni instrument minimalne ali relativne legislativnosti.

Osnovna vsebina sporazuma FATDS je v večji meri osredotočena na oblikovanje modelov in smernic, ki oblikujejo značilni instrument reflektivnega prava.

Končno je ta paradigma, ki jo udejanja sporazum FATDS, ozko uokvirjena z novimi pristopi, novimi metodami zakonskega urejanja in trenutno predlaganimi ukrepi; medtem

pa se EU bolj zanaša na instrumente, ki prenašajo pravila v prakso in le v majhni meri oblikujejo nova pravila; v našem primeru določa dostop, uporabo in možno ponovno uporabo informacij, vključenih v podatkovne baze na podlagi zakonodaje ter pogodb in/ali licenc.

Objava sporočila EGS o odprtih podatkih in predlog dopolnitve s strani Evropske komisije oblikujeta politiko odprtih podatkov in spodbujanja trga informacij, ki temelji na Direktivi 2003/98/ES Evropskega parlamenta in Sveta o ponovni uporabi informacij javnega sektorja z dne 17. novembra 2003.

Dodati moramo še revizijo Direktive o ponovni uporabi informacij javnega sektorja s konca leta 2011, ki predstavlja del Evropskega programa za digitalne tehnologije, iniciative Evropske komisije (2010), katere cilj je spodbujanje spletnih storitev znotraj Evropske unije, njena osrednja naloga pa je odpiranje javnih podatkov za ponovno uporabo, poenostavitev sistema izdajanja licenc za izmenjavo vsebin in uvajanje standardov interoperabilnosti.

5. Regulativna opcija sporazuma FATDS

Javni podatki v lasti javnih organov Evropske unije so predmet posebne obravnave v skladu z Direktivama o ponovni uporabi informacij javnega sektorja iz let 2003 in 2013, ki določata, da je treba dokumente JA dati na voljo javnosti za ponovno uporabo, in sicer »kakršnokoli vsebino ne glede na nosilec podatkov (bodisi na papirju bodisi v elektronski obliki ali kot zvočni, vizualni ali avdiovizualni zapis), ki jih hranijo organi javnega sektorja za komercialne ali nekomercialne namene«. Splošno načelo tako ostaja prosta razpoložljivost tovrstnih vsebin; če je predpisana taksa, »ta ne sme presegati mejnih stroškov reproduciranja, dajanja na razpolago in razširjanja dokumentov« (člen 6.1).

Skratka, sporazum FATDS je načeloma objavljen kot priporočilo, ki temelji na minimalni ali relativni legislativnosti na podlagi obstoječih pogodb o izvajanju storitev in podatkih

(prevajalskih pomnilnikih), ustvarjenih v okviru izvajanja storitev. To priporočilo predstavlja praktično specifikacijo dolžnosti izmenjave podatkov, ustvarjenih v obliki prevajalskih pomnilnikov, ki izhajajo iz prevajalskih storitev, ki jih razpišejo javne uprave, ne glede na medij ali obliko zapisa.

6. Vsebina Okvirnega sporazuma: sistem upravljanja prenosa podatkov iz prevajalskih pomnilnikov, ki ga spodbuja sporazum.

Pri prevajalskem pomnilniku gre za jezikovno bazo podatkov, ki sproti shranjuje vzporedne prevode, ki jih ustvarijo strokovnjaki in tako omogoča njihovo prihodnjo uporabo z namenom zagotavljanja popolnejše terminološke in slogovne usklajenosti ter na podlagi popolnih ali delnih ujemanj med novimi besedili in starimi, že zaključenimi prevodi, ustvarja prihranke.

Prevajalski pomnilniki so digitalne shrambe, ki jih sestavljajo vrstice besedila v izvirnem jeziku, ki so poravnane s prevodi v druge jezike. Ta besedila je mogoče učinkovito poravnati tudi s pomočjo prevajalskih enot. Prevajalske enote, shranjene skupaj s svojimi ustreznici, so opredeljene na različne načine (v obliki stavka, odstavka, besede ali skupine besed itd.), segmentacija, ki sledi ločilu, ki označuje konec povedi ali odstavka, pa najpogosteje predstavlja privzeti ločevalni element v okolju računalniško podprtih prevajalskih sistemov.

Glavna funkcija prevajalskih pomnilnikov (v nadaljevanju "TM") je ugotavljanje popolnih ali delnih predlogov oz. ujemanj iz povedi in skladnosti izrazov. Med prevajanjem poteka iskanje segmentov izvirnega jezika v podatkovni bazi TM. Če vsebuje TM v izvirnem jeziku segment s popolnim ali delnim ujemanjem, je ta segment predlagan prevajalcu skupaj s prevodom, pridobljenim iz te podatkovne baze; prikazane pa so tudi dodatne informacije, shranjene v podatkovni bazi skupaj s segmentom. Računalniško podprta prevajalska orodja (CAT) prikažejo stopnje ujemanja (v strokovnem jeziku znane kot *megleni zadetki*). Pri prevajanju se lahko ti zadetki izkažejo za uporabne od stopnje ujemanja 50–65 % dalje, v skladu s trdom, vloženim v prevajanje popolnoma novih

stavkov ali stavkov, za katere so znotraj istega konteksta že na voljo podobni predlogi in zadetki z določenim odstotkom ujemanja, pa so oblikovane tudi cene. 100-odstotna stopnja ujemanja se šteje za popoln zadetek zahtevane povedi in enake povedi v podatkovni bazi.

Odperta oblika zapisa za izmenjavo prevajalskih pomnilnikov se imenuje TMX (Translation Memory eXchange) in je običajno na voljo v različici 1.4b. Gre za XML-standard vrste DTD (*document type definition*). Ta format je oblikoval odbor OSCAR (Open Standards for Container/Content Allowing Re-use).

Uporaba formata TMX omogoča tako posameznikom kot podjetjem preprostejše sodelovanje pri prevajalskih projektih. Format TMX poleg tega olajša prehod z enega računalniško podprtega prevajalskega sistema na drugega, kar spodbuja konkurenčnost razpoložljivih tehnologij in njihov nenehni razvoj, s katerim se skuša razlikovati od tekmecev. Tako kot drugi odprti standardi je bil tudi ta format razvit z namenom odpravljanja težav z združljivostjo, spodbujanja ponovne uporabe jezikovnih virov, preprostejše izmenjave podatkov in spodbujanja tehnoloških inovacij.

Kot že rečeno je namen tega Okvirnega sporazuma oblikovanje okvira dobrih praks na področju pridobivanja podatkov, ustvarjenih v okviru pogodb o izvajanju prevajalskih storitev, ki jih sklepajo evropske javne uprave, njihovo organiziranje na podlagi uvajanja sistema programske opreme NEC TM in spodbujanje centralizacije podatkov, kot jo predpisuje Evropska komisija z namenom sprejetja protokola za centralizacijo dvojezičnih podatkov v javnih upravah, njegove uporabe in zagotavljanja splošne družbene koristi ter koristi za države članice in pomoči pri oblikovanju korpusa nacionalnih masovnih podatkov in izmenjave delov, za katere posamezna nacionalna javna uprava meni, da so ustrezni na evropski ravni.

V ta namen bo centralizacijo podatkov na nacionalnih računalniških infrastrukturah urejal pristojni organ posamezne države – npr. državni sekretariat, pristojen za digitalni napredek – ali pristojno ministrstvo v posamezni državi članici. Izmenjava teh podatkov

je mogoča po pristopu k Okvirnemu sporazumu in če tako določi pristojni organ, oz. ustrezní organ Evropske komisije (npr. ELRC-Share).

Ob upoštevanju tega bo s sprejetjem tega sporazuma prvi korak predstavljala centralizacija podatkov iz prevajalskih pomnilnikov, ki jih ustvarjajo javne uprave na nacionalni ravni, zatem pa izmenjava podatkov s panevropskim organom v korist JU drugih držav članic EU v skladu s tem sporazumom. Programska oprema za Nacionalni in evropski centralni prevajalski pomnilnik (NEC TM) bo omogočala vhodno/izhodno povezavo (I/O) upravam, ki bodo tako lahko v okviru zasebnosti shranjevale svoje prevajalske pomnilnike, ki jih bo gostil osrednji strežnik, in jih delile s tistimi, ki si bodo na nacionalni ravni prizadevali za več učinkovitosti in nižje stroške dela (običajno torej z internimi prevajalci ali zunanjimi izvajalci storitev), ustvarjale nacionalne masovne podatke in po želji izbirale podatke z osrednjega strežnika, na katerem bo shranjen osrednji evropski prevajalski pomnilnik.

Razmerje z zasebnimi subjekti in dodelitev TMX

Prevajanje, ki ga opravljajo ponudniki (izvajalci) storitev, bo vključevalo predložitev prevajalskih pomnilnikov, ki jih ustvarijo v okviru izvajanja storitve. V ta namen je treba v objavo v Uradnem listu države članice, regije ali občine, tj. na portalu pogodbenice, vključiti klavzulo, ki se nanaša na sporazum FATDS o NEC TM, vključno s KODO CPV, ki določa, DA BODO PODATKI IZ PREVAJALSKEGA POMNILNIKA DOSTAVLJENI NA STREŽNIK PRISTOJNEGA JAVNEGA ORGANA, da se bo v prihodnje po potrebi oz. želji nacionalnega organa lahko delil z ustrezno določenim pristojnim organom Evropske komisije (npr. ELRC-Share ali podobna iniciativa).

V ta namen bodo ponudniki (izvajalci) storitev skupaj s svojim delom predložili ustvarjene prevajalske pomnilnike in to ob upoštevanju načela neretroaktivnosti, kar pomeni, da ta okvirni sporazum ne učinkuje za nazaj in da velja ta obveznost od sprejetja in podpisa klavzule, s katero se zavezujejo, da bodo skupaj s svojim delom predložili tudi vzporedne podatke, ustvarjenje pri prevajanju (v TMX ali podobnem združljivem format).

7. INTELEKTUALNA LASTNINA IN POGOJI ZA UPORABO PODATKOVNIH BAZ

V pogledu intelektualne lastnine ni nikakršnega dvoma, da so podatki, ustvarjeni pri izvajanju dela v skladu s pogodbo, s katero je bilo dodeljeno naročilo za izvedbo zadevnega prevoda, v popolni izključni lasti javne uprave, ki je avtor oz. upravitelj izvirnega besedila, ki je predmet prevoda.

Pri tem je treba upoštevati tudi Uredbo o ponovni uporabi javnih informacij in odpiranju javnih podatkov, saj so javne uprave tiste, ki večinoma ustvarjajo informacije.

Zato bo izbrani ponudnik (izvajalec) storitev ekskluzivno in brez kakršnihkoli časovnih ali ozemeljskih omejitev odstopil pravice do ustvarjene dokumentacije ali podatkov kakršnekoli vrste in to neodvisno od podpore ali oblike zapisa ter ob popolnem razumevanju distribucije in preoblikovanja pravic reprodukcije pri tem prenosu.

Izbrani izvajalec storitev posledično ne sme uporabljati s takšnimi prevajalskimi pomnilniki ustvarjenih podatkov za pridobitni namen, temveč izključno za zagotavljanje informacij, če je to potrebno za izvedbo dela.

Intelektualna lastnina in pogoji za uporabo podatkovnih baz.

Podatkovna baza DGT TM je v izključni lasti Evropske komisije. Komisija za obdobje zaščite teh pravic ponovnim uporabnikom po vsem svetu brezplačno podeljuje pravice do vseh vrste uporabe, ki izpolnjujejo pogoje, določene v Odločbi Komisije o ponovni uporabi dokumentov Komisije z dne 12. decembra 2011, objavljeni v Uradnem listu Evropske unije L330 dne 14. decembra 2011 na str. 39 do 42.

Ponovni uporabnik mora ponovno uporabo podatkovnih baz ali strukturiranih elementov v njih identificirati in navesti vir uporabljenih dokumentov: spletni naslov, datum zadnje posodobitve in dejstvo, da ostaja lastništvo nad podatki v rokah Evropske komisije.

Zato ta podatkovna baza predstavlja optimalno osnovo za začetno dodajanje podatkov v pomnilnik NEC TM države članice.

8. VARSTVO PODATKOV. IZVAJANJE UREDBE EU 2016/679 IN UREDBE 2018/1725

V zvezi z osebnimi podatki, ki jih je morata v okviru izvajanja pogodbe obdelati javna uprava in izbrani ponudnik, sta obe pogodbenici zavezani k upoštevanju spodaj navedenih Splošnih uredb o varstvu podatkov.

Uredba (EU) 2016/679 Evropskega parlamenta in Sveta z dne 27. aprila 2016 in upoštevanje veljavne zakonodaje o varstvu podatkov v vsaki državi članici, ki izvaja ta Okvirni sporazum, ter Uredba Evropskega parlamenta in Sveta z dne 23. oktobra 2018 o varstvu posameznikov pri obdelavi osebnih podatkov v institucijah, organih, uradih in agencijah Unije in o prostem pretoku takih podatkov.

KONČNE UGOTOVITVE

Javna uprava je kot spodbujevalec panoge jezikovnih tehnologij z oblikovanjem splošnih platform za obdelavo naravnega jezika in samodejnega prevajanja ter razvojem virov za ponovno uporabo informacij javnega sektorja (RISP) obvezana razvijati politike o izmenjavi podatkov in oblikovati temelje za njihovo udejanjanje ter učinkovito izmenjavo med vsemi subjekti, ki sodelujejo v procesu, in zagotoviti, da so tudi udeleženci procesa zunaj javne uprave zavezani k podpisu klavzule o sprejemu ustreznih razpisov in predložitvi podatkov, ustvarjenih pri prevajanju (prevajalski pomnilniki) skupaj s svojim delom, ki predstavljajo skupek jezikovnih virov neprecenljive vrednosti.

REFERENCE. Ferrer-Sapena in drugi (2011) v članku o dostopu do javnih podatkov; Ferrer-Sapena in Peset (2012) o ponovni uporabi kulturnih podatkov ali Ramos Simón in drugi (2012) v študiji o evropskih podatkovnih portalih.

Licensing Open Data: A Practical Guide for the Higher Education Funding Council for England (Korn and Oppenheim, 2011). Omeniti velja tudi *Vodič za licenciranje odprtih podatkov – Guide to Open Data Licensing* (Open Knowledge Foundation, n.d.) in smernice, ki izhajajo iz *Evropskega programa za digitalne tehnologije* (Evropska komisija, 2010).