

POZVÁNKA NA PŘEDBĚŽNÉ TRŽNÍ KONZULTACE

k veřejné zakázce:

„Tvorba analytických nástrojů“

Zadavatel veřejné zakázky:

Česká republika – Ministerstvo práce a sociálních věcí
se sídlem Na Poříčním právu 376/1, 128 01 Praha 2
IČO: 00551023

Ministerstvo práce a sociálních věcí (dále jen „zadavatel“) připravuje zadávací řízení na veřejnou zakázku „**Tvorba analytických nástrojů**“.

Před vyhlášením veřejné zakázky budou představeny předpokládané zadávací podmínky veřejné zakázky potenciálním dodavatelům v rámci předběžné tržní konzultace.

Předmětem plnění veřejné zakázky bude (1) prověření datové kvality, tj. proces analýzy nepřesností, chybějících nebo logicky nesprávných dat v databázi nárokových podkladů (INEP), (2) datovou přípravu, tj. proces odstranění chybných a nepotřebných dat z INEP, (3) provedení analýzy obsahu upravené INEP, tj. identifikace míst v databázi, která potřebují doplnit, (4) návrh postupů, identifikace použitelných dat a doplnění potřebných údajů, jež zkompletují INEP, (5) ověření kvality provedeného doplnění a (6) vybudování nástroje, prostřednictvím něhož bude upravená INEP převedena do podoby databáze model-pointů (DMP). Tato automatizace tvorby DMP bude zahrnovat i funkcionalitu v podobě flexibilního doplňování DMP o tzv. new-business (modelování osob, které do systému pojištění teprve vstoupí). Výstupem VZ bude metodika tvorby DMP z INEP, pilotní DMP vytvořená dle manuálu a nový nástroj (open-source SW) pro automatizovaný převod upravené INEP do podoby DMP.

Podrobnější specifikace předmětu plnění veřejné zakázky je uvedena níže.

Úvodní schůzka předběžné tržní konzultace se bude konat dne **14. 6. 2017 v 09:00 hodin** v budově sídla zadavatele **Na Poříčním právu 1/376, 128 01 Praha 2**.

Následná komunikace v předběžné tržní konzultaci bude probíhat elektronicky.

Předběžná tržní konzultace je podle § 33 zákona č. 134/2016 o zadávání veřejných zakázek (dále jen „ZZVZ“) možností zadavatele předtím, než vyhlásí veřejnou zakázku, komunikovat s dodavateli a zjišťovat jejich možnosti a návrhy řešení. V rámci předběžné tržní konzultace bude představen záměr zadavatele, včetně některých navrhovaných detailů jak předmětu veřejné zakázky, tak zadávacího

řízení. Dodavatelé se pak budou moci k navrhovaným parametrům zakázky vyjádřit. Dojde tak ke zvýšení transparentnosti zadávacího řízení a získání relevantních a objektivních informací o možnostech trhu, tak aby mohl zadavatel optimálně nastavit zadávací podmínky veřejné zakázky.

Možnost účastnit se tohoto jednání je otevřená. Účastnit se mohou zejména všichni potenciální dodavatelé o tuto veřejnou zakázku.

Dodavatelé, kteří mají o účast na předběžných tržních konzultacích zájem, potvrdí svou účast nejpozději **do 12. 6. 2017** na e-mailové adrese jana.jandova@mpsv.cz. V registraci dodavatelé uvedou název dodavatele a jména zúčastněných osob.

Pozvánka na předběžnou tržní konzultaci k této veřejné zakázce byla zároveň uveřejněna ve Věstníku veřejných zakázek (<https://www.vestnikverejnychzakazek.cz/Form01/Display/25325>).

Předběžné tržní konzultace budou vedeny tak, aby to nenarušilo hospodářskou soutěž ani zásady zadávání veřejných zakázek ve smyslu § 6 ZZVZ.

Podrobnější specifikace předmětu plnění veřejné zakázky

Hlavním výstupem veřejné zakázky bude manuál (metodika) tvorby databáze modelových bodů ze vstupních databází administrativních údajů a doplňkových statistických ukazatelů nebo výsledků výběrových šetření a pilotní databáze modelových bodů vytvořená pomocí postupu popsaného v manuálu, aktuální ke konci roku 2015.

Databáze modelových bodů je klíčovým vstupem dynamického mikrosimulačního modelu důchodového systému MPSV. Obsahuje základní informace o modelových bodech (modelová osoba). Tyto informace zahrnují informace obecné povahy (pohlaví, věk, rodinný stav ad.), osobní historii důchodového pojištění (historické výdělky, pojištěné doby ad.) a informace o dalších, z hlediska důchodového systému, významných osobách (manžel/ka, děti). Plný rozsah informace obsažené v modelovém bodu je uveden v příloze 1. Databáze bude vytvořena transformací historických údajů (pro žijící osoby) a odhadem založeným na historických údajích jiných osob (pro tzv. nový business). Současně bude třeba vyřešit párování osob do modelových bodů.

Databáze modelových bodů má řádkovou strukturu, kde každý řádek představuje plné údaje o jedné modelové osobě. Celkový rozsah databáze je přibližně 20 milionů řádků (obsahuje všechny existující osoby – občané ČR – a osoby, které do systému teprve vstoupí – tzv. nový business).

Hlavním vstupem pro tvorbu databáze modelových bodů je kompletní záznam z databáze České správy sociálního zabezpečení (databáze INEP). Databáze INEP má řádkovou strukturu, kde jeden řádek představuje záznam o konkrétním občanu v konkrétním kalendářním roce. Databáze obsahuje údaje významné z pohledu důchodového pojištění, zejména roční výdělek, dobu pojištění a náhradní doby pojištění. Plný rozsah informace obsažené v řádkovém záznamu databáze INEP je uveden v příloze 2.

Databáze INEP je tvořena cca 200 miliony řádků a obsahuje údaje za všechny pojištěnce od okamžiku jejich vstupu do systému důchodového pojištění do konce roku 2012. Databáze byla vytvořena jednorázovým exportem z provozních aplikací České správy sociálního zabezpečení a obsahuje tedy všechny údaje, které ČSSZ o každém jednotlivci eviduje. Veškeré doby, které nejsou v aplikacích ČSSZ identifikovány, jsou v databázi INEP agregovány do položky „nepojištěn“.

Dalším zdrojem administrativních údajů jsou databáze STATMIN VZ a STATMIN ANOD, které obsahují individuální roční údaje o závislé výdělečné činnosti pojištěnce (VZ), resp. roční údaje o vyplácených důchodech (ANOD).

Databáze STATMIN VZ obsahuje řádkové záznamy o jednotlivých zaměstnáních pojištěnce, kdy každý řádek představuje jeden zaměstnanecký vztah daného občana (jeden pojištěnec může mít více záznamů). Databáze obsahuje cca 6 milionů řádků. Databáze STATMIN VZ jsou k dispozici za období 2004 – 2015. Databáze VZ obsahuje elektronizované údaje ze všech evidenčních listů důchodového pojištění, které ČSSZ obdržela za daný kalendářní rok.

STATMIN ANOD obsahuje řádkové informace o důchodech vyplácených danému důchodci, přičemž každý důchodce má přiřazen právě jeden záznam. Rozsah této

databáze je cca 3 miliony řádků. Popis informace v databázích je obsahem příloh 3 a 4. Databáze STATMIN ANOD jsou dostupné za roky 2008 – 2015.

Všechny uvedené databáze (INEP, VZ a ANOD) používají pro své záznamy stejný unikátní identifikátor a jejich údaje jsou jednoznačně propojitelné.

Řešitel veřejné zakázky provede pilotní tvorbu databáze modelových bodů spočívající v prověření datové kvality, tj. proces analýzy nepřesností, chybějících nebo logicky nesprávných dat v databázi INEP, datové přípravě, tj. proces odstranění chybných dat z databáze INEP, provedení analýzy obsahu upravené databáze INEP, tj. identifikace míst v databázi, která potřebují doplnit (přestože data nejsou chybná), návrh postupů, identifikace použitelných dat a doplnění potřebných údajů, jež zkompletují databázi INEP (tj. databáze bude obsahovat všechny údaje potřebné pro tvorbu databáze modelových bodů) a ověření kvality takto provedeného doplnění. Databáze STATMIN VZ a STATMIN ANOD budou použity pro kontrolu údajů v databázi INEP a pro případné doplnění chybějících nebo nepřesných údajů. V případě doplňování údajů za použití databází STATMIN VZ a STATMIN ANOD, budou tyto údaje rovněž prověřeny z pohledu datové kvality.

Současně budou navrženy postupy doplnění chybějících údajů. Tyto postupy budou mít podobu algoritmů nastavených k zaplnění prázdných míst v databázi. Dodavatel pro nastavení těchto algoritmů v maximální míře vytěží informace obsažené ve vstupních databázích INEP, STATMIN VZ a STATMIN ANOD. S pomocí veřejně dostupných zdrojů, zejména statistik, dojde k verifikaci použitých algoritmů tak, aby výstupy algoritmů odpovídaly v agregované podobě skutečnosti.

Pro účely zakázky bude možné zajistit i přístup k obvykle neveřejným informačním zdrojům, např. k mikrodátům EU-SILC, která mohou řešení zpřesnit a usnadnit. Data EU-SILC budou využitelná mj. k tvorbě algoritmu, který bude vytvářet z individuálních záznamů společně žijící dvojice.

Použité postupy a algoritmy budou zadavateli předány v podobě skriptu nebo kódu pro volně dostupný software nebo software, který MPSV již používá. Funkcí skriptu/kódu je jednak vytvoření upravené databáze INEP, která bude obsahovat plné informace (bez chybějících nebo chybných záznamů) a, za druhé, tvorba databáze modelových bodů.

Současně budou použité postupy popsány v manuálu. Manuál bude mít takovou formu a podobu, aby zaměstnanci zadavatele byli schopni po aktualizaci databáze INEP (jednou za několik let) schopni replikovat použité postupy a vytvořit aktualizovanou verzi databáze modelových bodů vlastními silami, včetně případných dílčích úprav kódu, pokud by si ji budoucí aktualizace vyžádala.

Součástí plnění bude také dokumentace, která bude obsahovat všechny provedené testy, úpravy a odvození algoritmů a jejich popis a provede diskuzi výsledků.

Předběžná představa zadavatele o postupu řešení

Dodavatel v plnění bude primárně vycházet z metodiky „Kontrola kvality vstupních dat a jejich použitelnost“, která je v příloze 5.

Dodavatel provede ověření datové kvality obsahu databáze INEP, nalezne chybně vyplněné, nepřesné nebo logicky nesprávné údaje a provede analýzu residuální informace obsažené v „DOBA NEPOJIŠTĚN“ a „SKR_NEPOJIŠTĚN“, kde je záznam v případě, že není v ČSSZ k dané situaci žádný záznam. Stejnou analýzu provede s dalšími datovými zdroji dodanými MPSV, konkrétně databázemi STATMIN VZ a STATMIN ANOD. Nedostatky v databázích budou klasifikovány podle typu nepřesnosti a detailně popsány.

Následně proběhne příprava databáze pro další úpravy. V tomto kroku dodavatel odstraní chybné, nepotřebné a nadbytečné údaje z databáze INEP. Odstraněné chybné údaje, pokud půjde o individuální, nikoli systematické chyby, budou nahrazeny údaji, které nebudou chyby vykazovat. Systematické chyby budou v tomto kroku vyřešeny vymazáním datových záznamů.

Následně primárně na základě analýzy residuální informace nalezne místa, kde chybí relevantní data. Pro takto identifikovaná místa provede rešerši dostupných datových zdrojů, jiných než databáze INEP, a identifikuje datový nebo statistický zdroj nejvhodnější k doplnění potřebných informací na tato „bílá místa“.

S využitím nalezených datových zdrojů dojde k doplnění údajů do databáze INEP upravené podle předchozích kroků. V případě, že budou využívána data neindividuálního charakteru, dodavatel navrhne a použije algoritmy, kterými budou potřebné údaje odvozeny a do upravené databáze INEP doplněny.

Dodavatel provede další kontrolu upravené a doplněné databáze z pohledu datové kvality. Pokud budou nalezeny další nebo nové chyby v datové kvalitě, bude opakován popsáný postup, dokud nebude databáze plně opravena. Současně budou ověřeny výsledky použitých algoritmů, porovnáním se statistickými údaji, aby bylo zajištěno, že algoritmy „trefují“ správné počty a rozložení osob, podle jednotlivých proměnných.

K celému plnění připraví dodavatel zprávu, která shrne postup prací v jednotlivých krocích, vysvětlí zvolená řešení a zhodnotí výsledky plnění.

Zadavatel zároveň přikládá ke specifikaci plnění veřejné zakázky následující přílohy:

- Příloha č. 1
- Příloha č. 2
- Příloha č. 3
- Příloha č. 4
- Příloha č. 5