



## ZÁPIS Z PŘEDBĚŽNÉ TRŽNÍ KONZULTACE K VEŘEJNÉ ZAKÁZCE – SCHŮZKA S ÚČASTNÍKEM (DODAVATELEM)

### Tvorba analytických nástrojů

Schůzka se konala dne **3. 7. 2017 od 10:00** v sídle účastníka (dodavatele) předběžné tržní konzultace (dále jen „PTK“) společnosti KPMG Česká republika, s. r. o., Pobřežní 1a, 186 00 Praha 8.

Za zadavatele se schůzky účastnili:

**Ing. Jan Škorpík** – vedoucí oddělení pojistně matematického MPSV

**Ing. Marek Suchomel** – referent oddělení pojistně matematického MPSV

Za účastníka (dodavatele) se schůzky účastnili:

**Ondřej Kulháněk**

**Boris Halata**

**David Dušek**

#### 1. Předmět schůzky s účastníkem (dodavatelem)

Zástupce dodavatele Boris Halata poděkoval zástupcům zadavatele za možnost realizovat tuto schůzku a představil své kolegy – Ondřeje Kulhánka a Davida Duška. Zástupce zadavatele – Ing. J. Škorpík – na úvod pouze upozornil všechny účastníky schůzky, že pro potřeby zpracování zápisu ze schůzky, která je součástí PTK, bude pořízen z průběhu schůzky audiozáznam. Nikdo z účastníků schůzky neuplatnil vůči tomuto opatření žádné námítky.

Předmětem schůzky byly především doplňující dotazy ze strany dodavatele k požadavkům zadavatele veřejné zakázky „Tvorba analytických nástrojů“ a zároveň i získání další informací, které budou zadavateli sloužit pro řádné nastavení zadávacích podmínek a konkrétnější specifikaci předmětu plnění veřejné zakázky. Tato zpětná vazba, tj. doplňující informace od účastníka PTK (dodavatele) budou sloužit jako podpůrný nástroj a umožní zadavateli co nejsrozumitelněji specifikovat předmět plnění.

Zadavatel bude postupovat v souladu s zák. č. 134/2016 Sb., o zadávání veřejných zakázek, ve znění pozdějších předpisů (zejména v souladu na ustanovení § 36 odst. 4 uvedeného zákona). V zadávací dokumentaci bude označeno vše, co vzešlo z PTK (zadavatel bude v maximální možné míře dodržovat zásadu transparentnosti). Tento zápis bude uveřejněn na profilu zadavatele MPSV.

#### 2. Diskuse účastníků schůzky

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

Na úvodu znovu stručně představil zástupcům dodavatele předmět plnění, zejména uvedl, že hlavním vstupem pro tvorbu databáze modelových bodů (dále jen „DMP“) je kompletní záznam z databáze

České správy sociálního zabezpečení (dále jen „databáze INEP“). Databáze INEP obsahuje údaje za pojištěnce od okamžiku jejich vstupu do systému důchodového pojištění do konce roku 2012. Databáze INEP má pevně definovanou strukturu. Tato struktura byla předem definována ve spolupráci s Českou správou sociálního zabezpečení (dále jen „ČSSZ“) na základě požadavku MPSV. Do jisté míry je struktura databáze INEP odrazem struktury DMP, která je klíčovým vstupem dynamického mikrosimulačního modelu důchodového systému MPSV.

Ani databáze INEP neobsahuje kompletní informaci o pojištěncích, která by se z hlediska možnosti analýz jevila jako ideální, neboť hlavní činností ČSSZ není analytická činnost, ale především zajištění výpočtu a výplaty dávek důchodového pojištění a tomu je primárně přizpůsoben rozsah informace o pojištěncích, kterou je ČSSZ oprávněna disponovat. Databáze INEP má řádkovou strukturu, kde jeden řádek představuje záznam o konkrétním pojištěnci v konkrétním roce. Vzhledem k tomu, že v databázi je hierarchie jednotlivců a kalendářní rok, tak je zde rovněž obsažena reziduální informace v podobě proměnných „DOBA NEPOJIŠTĚN“ a „SKR\_NEPOJIŠTĚN“. Tyto proměnné obsahují informaci, pokud v evidenci ČSSZ není k dané situaci žádný záznam včetně stavu na konci kalendářního roku.

Databáze INEP je anonymizována. Unikátní identifikátor umožňuje pouze spojení jednotlivých řádků za konkrétního jedince v databázi INEP a případně propojení údajů z databáze INEP s údaji z databáze STATMIN VZ a STATMIN ANOD. Těmito databázemi MPSV rovněž disponuje. Jakákoliv jiná databáze administrativních dat již není propojitelná s výše uvedenými databázemi.

V první fázi předmětu plnění veřejné zakázky dojde k prověření datové kvality databáze INEP, zejména by mělo dojít k tzv. „rozklíčování“ informace v reziduální proměnné. Databáze INEP obsahuje údaje o jedincích do konce roku 2012 a částečně za rok 2013. Nicméně údaje za rok 2013 již nejsou kompletní. MPSV má rovněž k dispozici databázi STATMIN VZ za období 2004 – 2015 a STATMIN ANOD za období 2008 – 2015. Databáze STATMIN VZ obsahuje údaje ze všech evidenčních listů důchodového pojištění, které ČSSZ obdržela za daný kalendářní rok. Databáze STATMIN VZ a STATMIN ANOD by rovněž měly sloužit v další fázi pro doplnění databáze INEP po roce 2012.

Ve finální fázi předmětu plnění veřejné zakázky by měla být doplněná databáze, která vznikla v předchozích dvou fázích, transformována do podoby DMP. Databáze DMP obsahuje další doplňující informace, které nejsou v databázi INEP k dispozici (např. údaje o vzdělání) a rovněž obsahuje informaci o slučování jedinců do párů (tzv. párování osob). Dotaz související s párováním osob v databázi DMP již zaslal zadavateli jiný účastník PTK (PricewaterhouseCoopers Česká republika, s. r. o.). Odpověď zadavatele na tento dotaz byla dne 27. 6. 2017 uveřejněna na profilu zadavatele MPSV.

#### **Zástupce dodavatele (Ondřej Kulhánek, KPMG Česká republika, s. r. o.):**

Poděkoval zástupci zadavatele a představil první dotaz ze strany dodavatele.

Hledá zadavatel spíše jednorázové řešení ve vztahu ke zpracování a analýze dat nebo řešení, které nabídne průběžnou kontrolu a bude fungovat tzv. na denní bázi nebo případně kombinaci uvedených řešení?

#### **Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

K tomuto dotazu uvedl (1) databáze INEP byla vytvořena jednorázovým exportem z provozních aplikací ČSSZ na základě ad hoc požadavku MPSV. Cílem zadavatele v rámci plnění veřejné zakázky je „datově“ zkvalitnit databázi INEP, jedná se tedy především „o jednorázovou činnost“. Nicméně nelze zcela vyloučit, že v dlouhodobém horizontu by mohl být opět vznesen směr k ČSSZ ad hoc požadavek na jednorázový export z provozních aplikací ČSSZ, tj. vznikla by další databáze INEP.

Dále zástupce zadavatele doplnil, že se předpokládá, že jedním z výstupů veřejné zakázky bude manuál či nástroj, který umožní zadavateli opakovat bez vzniku dalších dodatečných nákladů prověření datové kvality u databáze INEP. Databázi STATMIN VZ získává zadavatel každoročně a v rámci plnění veřejné zakázky předpokládá, že vznikne metodika či nástroj, který umožní zadavateli každoročně transformovat data z databáze STATMIN VZ do datové struktury upravené databáze INEP.

**Zástupce dodavatele (Ondřej Kulhánek, KPMG Česká republika, s. r. o.):**

Druhý dotaz ze strany dodavatele se týkal dokumentace a metadat k datovým zdrojům, kterými zadavatel disponuje. (1) Co je v současné době k dispozici? a (2) Dodavatel vnímá datovou kvalitu nejen z pohledu toho, co je obsaženo v datech, ale rovněž jak jsou strukturována metadata, jak je vše popsáno?

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

MPSV obdrželo z ČSSZ výstupní databázi v podobě databáze INEP. Přístup ke zdrojovým datům nemá. Z hlediska dokumentace má MPSV k dispozici pouze doplňující informace z ČSSZ vztahující se k databázi INEP.

**Zástupce zadavatele (Ing. Marek Suchomel, MPSV):**

K databázi INEP neexistuje souhrnná dokumentace, která by popisovala způsob vzniku informací, které jsou v databázi INEP a co se za nimi skrývá, tj. metadata v obecné rovině k dispozici nejsou. Zadavatel při převzetí databáze INEP od ČSSZ neobdržel tedy „balík dat“ a k tomu dokumentaci obsahující kompletní popis dat v balíku.

**Zástupce dodavatele (Ondřej Kulhánek, KPMG Česká republika, s. r. o.):**

Další dotaz ze strany dodavatele směřoval na oblast datové kvality. Dodavatel uvedl pro lepší porozumění příklad. Např. proměnná „rodné číslo“. Při prověřování datové kvality je možné si popsat strukturu této proměnné a jakým způsobem provedu kontrolu této proměnné. Dále je možné prověření ve vazbě na jiné proměnné, např. zda proměnná „rodné číslo“ odpovídá datu narození či pohlaví. Je výše uvedený příklad postupu prověřování datové kvality požadován ze strany zadavatele?

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

V obecné rovině ano. Kontrola datové kvality např. výše uvedeným postupem se zadavateli jeví jako vhodná. Zadavatel nicméně odkázal na studii s názvem „Kontrola kvality vstupních dat a jejich použitelnost“, kterou pro MPSV v roce 2015 realizovala společnost Ernst & Young, s. r. o. Tato studie byla a je k dispozici všem uchazečům v rámci PTK na profilu zadavatele. Studie popisuje metodiku datové kvality ve vztahu k databázím, kterými MPSV disponuje. Zástupce zadavatele jako příklad ve kterém bude nutné prověřit datovou kvalitu je např. v rámci database STATMIN VZ proměnná „poštovní směrovací číslo“.

**Zástupce dodavatele (David Dušek, KPMG Česká republika, s. r. o.):**

Bez ohledu na výsledek prověřování datové kvality databáze INEP není cesta k přegenerování databáze INEP. Stávající podoba databáze INEP je definitivní a neměnná?

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

Ano. Vynečování databáze INEP byl jednorázový krok, který je nereplikovatelný.

**Zástupce dodavatele (Ondřej Kulhánek, KPMG Česká republika, s. r. o.):**

Další dotaz dodavatele směřoval na oblast nástrojů pro prověření datové kvality. Pokud jde tedy o jednorázovou akci, lze tuto oblast vyřešit např. (1) vytvořením nějakého skriptu, který se v případě potřeby spustí a provede určité kroky nebo druhou možností (2) je nástrojová podpora softwarového řešení, které již dodavateli slouží pro prověření datové kvality. Jaká je v tomto směru představa zadavatele?

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

V této otázce nemá zadavatele ještě zcela konečnou představu. Tato otázka bude ještě předmětem interní diskuse zadavatele. Otázka softwarové podpory již byla v rámci PTK otevřena na jednání, které se uskutečnilo dne 14. 6. 2017. Zadavatel nemá k dispozici ve svém týmu datové analytiky a tak není jednoznačně kvalifikován na to, aby jednoznačně určil, která softwarová podpora se z hlediska prověřování datové kvality jeví jako nejvhodnější. Ani nebude možné tuto otázku jednoznačně vymezit v rámci zadávací dokumentace. Vše tedy bude odvislé od nabídek, které budou předloženy v rámci zadávacího řízení a budou předmětem hodnocení ze strany zadavatele.

**Zástupce dodavatele (Ondřej Kulhánek, KPMG Česká republika, s. r. o.):**

Je tedy v možné v nabídce, kterou by dodavatel předložil ve lhůtě pro podání nabídek v rámci zadávacího řízení, předložit i několik variant řešení? Tím ovšem může dojít např. k rozdílu v celkové finanční hodnotě veřejné zakázky.

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

V aktuální rozpracované podobě zadávací dokumentace nepřipouští zadavatel variantní řešení nabídky. Zadavatel předpokládá, že výstupem veřejné zakázky bude kompletní řešení, které nezatíží zadavatele z hlediska dalších finančních nákladů, které by mohly vzniknout s souvislostí s užíváním výstupu veřejné zakázky (např. nejprve než začne zadavatele používat výstup veřejné zakázky, bude si muset např. zakoupit licenci softwarového nástroje, který umožní aplikaci výstupu veřejné zakázky). Licence nutné k používání výstupu předmětu plnění musí být již součástí řešení, tj. být součástí výstupu veřejné zakázky.

**Zástupce dodavatele (David Dušek, KPMG Česká republika, s. r. o.):**

Došlo ze strany zadavatele k posunu v otázce zveřejnění seznamu softwarových nástrojů, kterými v současnosti zadavatele disponuje a případně, kdy bude tento seznam zveřejněn?

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

V tento moment nemá ještě zadavatel seznam dokončen.

**Zástupce dodavatele (David Dušek, KPMG Česká republika, s. r. o.):**

V jakém softwarovém nástroji momentálně zadavatel databázi INEP zobrazuje? Např. v MS SQL (Microsoft SQL)?

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

Data, tj. databáze INEP, které má zadavatel k dispozici jsou ve formátu csv (textový soubor) a pro zobrazení a jakoukoliv další práci s nimi zadavatel používá Microsoft SQL.

**Zástupce dodavatele (Ondřej Kulhánek, KPMG Česká republika, s. r. o.):**

Jak bude v rámci zadávací dokumentace a v rámci realizace veřejné zakázky řešena otázka ochrany osobních údajů? Zejména s ohledem na nařízení Evropského parlamentu a Rady (EU) o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a volném pohybu těchto údajů, které bylo zveřejněno v Úředním věstníku EU v květnu 2016 a u kterého běží dvouletá lhůta na zajištění souladu stávající regulace s novou regulací při zpracování osobních údajů.

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

Předmět plnění veřejné zakázky se neprolíná v uvedeném nařízení. Databáze, s níž bude v průběhu řešení veřejné zakázky jakkoliv nakládáno, nejsou předmětem výše uvedené regulace. Nicméně otázka disponování s databázemi bude ještě předmětem interní diskuse zadavatele a konkrétní vymezení pro disponování s databázemi bude uvedeno v zadávací dokumentaci, resp. v návrhu smlouvy. V otázce zabezpečení při disponování s databázemi, bude primárně zadavatel operovat se smluvními sankcemi.

**Zástupce dodavatele (David Dušek, KPMG Česká republika, s. r. o.):**

Zadavatel uvedl, že databáze INEP prošla procesem anonymizace. Obsahuje však databáze INEP určité osobní údaje? Např. rodné číslo, jméno a příjmení nebo jméno zaměstnavatele?

**Zástupce zadavatele (Ing. Jan Škorpík + Ing. Marek Suchomel, MPSV):**

Databáze INEP je plně anonymizována a identifikace jednotlivce je vyloučena. Rodné číslo, jméno a příjmení, ani jméno zaměstnavatele se v databázi INEP nenachází. Pro kumulaci údajů (např. pokud chci zjistit dobu pojištění a každý řádek v databázi INEP obsahuje záznam pro jednoho jednotlivce za jeden kalendářní rok), musím použít unikátní identifikátor, který databáze INEP obsahuje a který znemožňuje identifikaci konkrétní osoby (jedná o desetimístný kód, který nemá vazbu na další proměnné v databázi INEP). Rovněž informace o zaměstnavateli je anonymizována. Anonymizované údaje obsahuje rovněž databáze STATMIN VZ.

**Zástupce dodavatele (Ondřej Kulhánek, KPMG Česká republika, s. r. o.):**

Jaký předpokládá zadavatele další postup v rámci PTK? Zejména z časového hlediska?

**Zástupce zadavatele (Ing. Jan Škorpík, MPSV):**

Probíhá proces přípravy zadávací dokumentace. Do 7. července 2017 očekává zadavatel od dodavatelů zaslání (1) připomínek k představě zadavatele o postupu řešení, resp. k navrhovaným parametrům veřejné zakázky a (2) písemných podkladů obsahujících představu dodavatele o postupu řešení, která se zcela nebo pouze částečně může odlišovat od průběžné představy zadavatele.

Zadavatel předpokládá finalizaci zadávací dokumentace na přelomu srpna a září t. r. a vyhlášení veřejné zakázky na přelomu měsíců září a říjen 2017. PTK pravděpodobně poběží až do okamžiku vyhlášení veřejné zakázky.