**Marketingový průzkum pro nákup SW řešení**

**pro správu dat v resortu MZe**

Obsah

[1 Úvod 3](#_Toc16234841)

[2 Obecné požadavky na technologii 3](#_Toc16234842)

[3 Data Quality Tool (DQT) – požadované vlastnosti a funkcionality 3](#_Toc16234843)

[3.1 Data Quality Tool - požadované vlastnosti 3](#_Toc16234844)

[3.2 Data Quality Tool - požadované funkcionality 4](#_Toc16234845)

[5 Plánovaná výsledná architektura DG 5](#_Toc16234846)

[6.1 Legenda ke schématu 5](#_Toc16234847)

[6 Položkový rozpočet 7](#_Toc16234848)

# **1 Úvod**

Cílem tohoto průzkumu je zjistit finanční požadavky na nákup softwarového řešení, které zajistí vytěžování metadat z resortu MZe, jejich validaci, vyčištění a odstranění duplicit.

# **2 Obecné požadavky na technologii**

Z hlediska potřeb MZe musí navrhované technologické řešení splňovat následující požadavky:

**Adaptace na měnící se podmínky** - technologieumožní podporu dnešních i budoucích požadavků organizace s využitím dostupných technologických komponent. Použité technologické komponenty musí poskytovat možnost vysoké adaptability na změny v organizaci i IT architektuře.

**Integrace do existující infrastruktury -** uvažované technologické produkty a řešení umožní zahrnout stávající technologické komponenty již použité v resortu MZe. Navržený produkt musí zvládnou komunikovat (předávat data) s ostatními IS MZe prostřednictvím webových služeb.

**Úložiště dat pouze na MZe** – veškerá data budou ukládána na vnitřních úložištích MZe. Data nelze ukládat do úložišť, která nejsou vlastněna MZe.

**Stabilita řešení, podpora a dlouhodobá perspektiva** – technologické řešení bude projektem v dlouhodobém horizontu, proto bude požadována záruka dlouhodobé a stabilní podpory (bude upřednostněno řešení dodavatele s odpovídajícím zázemím a zdroji, garantující trvající podporu a rozvoj).

# **3 Data Quality Tool (DQT) – požadované vlastnosti a funkcionality**

Nástroj DQT se bude skládat minimálně z následujících komponent:

**Front-end nástroj:** umožní vytvářet, vyvíjet a testovat jednotlivé úlohy datové kvality, spravovat všechny zdroje, objekty, komponenty, plánovat a spouštět jednotlivé úlohy.

## **3.1 Data Quality Tool - požadované vlastnosti**

Nástroj DQT bude mít tyto vlastnosti:

Otevřenost pro:

datové zdroje: DB/soubory/XML, XLS …

logiku, semantiku, pravidla

podporu rozvoje řešení

Intuitivní a strukturované grafické uživatelské rozhraní (GUI)

Univerzálnost pravidel

Podpora lokálních prostředí

Výhled na rozvoj produktu – existence vývojové roadmapy, financování vývoje, zázemí

Podpora národních a mezinárodních standardů

## **3.2 Data Quality Tool - požadované funkcionality**

Nástroj DQT bude vybaven těmito funkcionalitami:

Zjištění struktury dat

Zkoumání obsahu dat

Zkoumání vztahů v datech

Analýza stavu dat

Vytvoření profilu dat

Identifikace, kvantifikace, klasifikace, analýza chyb a problémů v datech

Automatické rozpoznávání obsahu datových položek

Automatické rozpoznávání struktury dat

Základní čištění dat, opravy, překlepy

Převedení na požadované (standardizované) hodnoty

Převedení na požadovanou strukturu a formát

Identifikace nulitních položek

Doplnění chybějících položek

Rozpoznání typu entity (FO, PO, pohlaví, typ subjektu atd.)

Kontrola přístupných hodnot

Kontrola vůči standardům

Porovnávání, párování záznamů

Fuzzy matching

Vytváření skupin záznamů

Výběr, vytvoření referenčního, zlatého master záznamu

Výběr nejlepších atributů záznamu

Ověření existence v referenčním zdroji (registru)

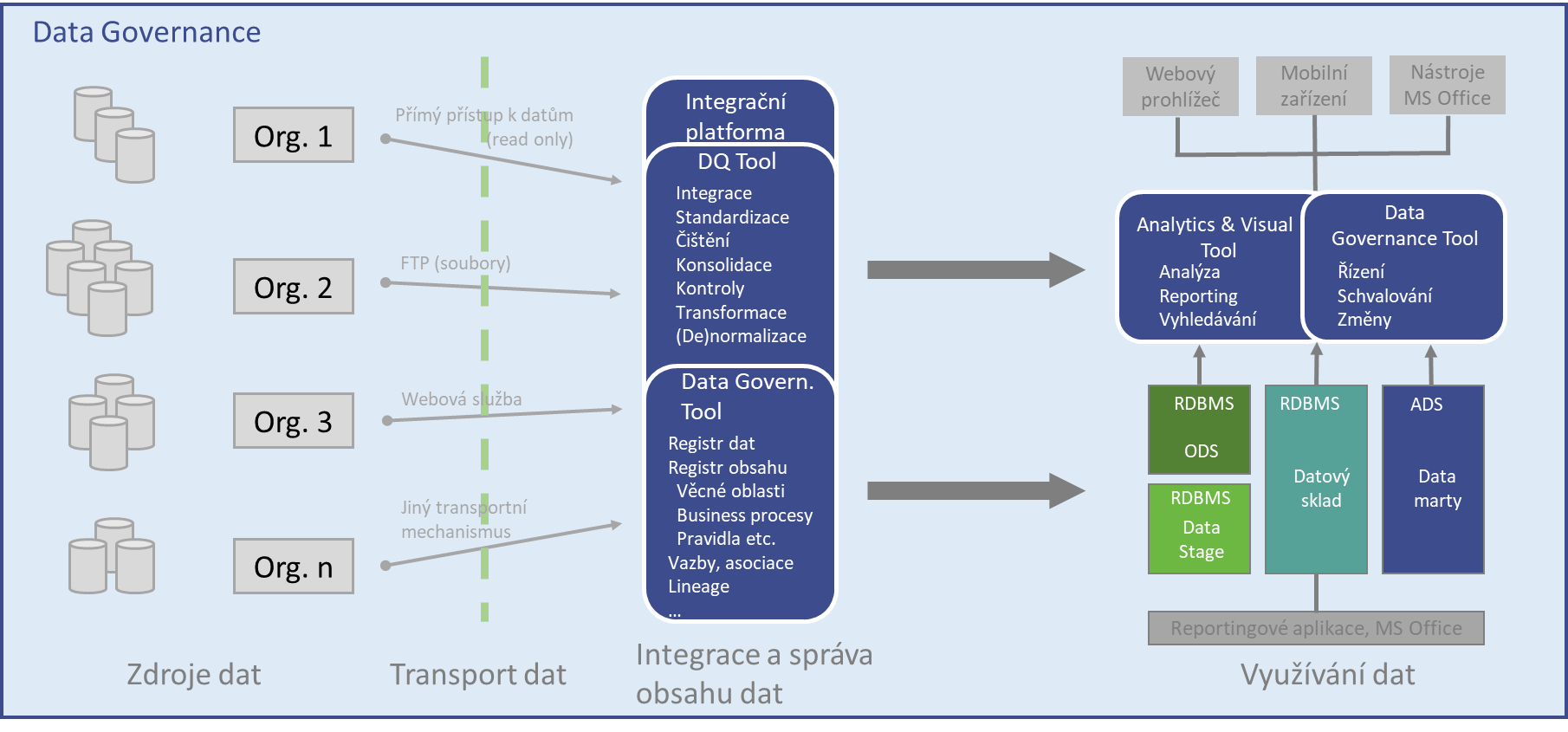
Adresní registry (UIR-ADR, RUIAN atd.)

Registr ekonomických subjektů (RES), Obchodní rejstřík atd.

Specializované registry (možnost napojení přes SOAP)

# **4 Plánovaná výsledná architektura DG**

Na obrázku níže je znázorněna plánovaná výsledná architektura SW řešení pro Data Governance, do které má být Data Quality Tool implementován.



*Ukázka plánované architektury*

## **4.1 Legenda ke schématu**

**Zdroje dat**

MZe

resortní organizace

ostatní organizace a subjekty

**Transport dat (různé mechanismy přenosu dat od organizací do centrály)**

iniciovaný centrálou: přímý přístup k datům

iniciovaný centrálou: přenos FTP apod. (organizace připraví datový extrakt „k vyzvednutí“)

iniciovaný centrálou: systém/aplikace (předá data prostřednictvím volání webové služby)

iniciovaný organizací: přenos FTP apod. (organizace odešle datový extrakt do centrály)

iniciovaný organizací: systém/aplikace (předá data prostřednictvím volání webové služby)

další mechanismy: off-line na médiu, elektronická pošta atd.

**Integrace dat („Integrační ústředna“ (ETL+))**

načtení/příjem dat z organizací prostřednictvím různých transportních mechanismů

integrační a konsolidační operace s daty (profiling, standardizace, normalizace, datová kvalita, kontroly dat atd.)

distribuce a synchronizace dat do cílových systémů a úložišť

řízení a monitorování procesů integrace dat, integrace s workflow atd.

**Využívání dat**

možnosti využití: reporting

analýza

monitorování

vyhledávání

poskytování informací do resortu i mimo něj

typy úložišť: Data Stage („vyrovnávací“ datové úložiště)

ODS (Operational Data Store)

Datový sklad (centralizovaná konsolidovaná informace)

Datamarty (specializovaný reporting a analýzy)

typy přístupu k informacím: webové rozhraní

mobilní aplikace

nástroje MS Office

speciální BI nástroje

konzumenti informací: uživatelé s vyšší úrovní přístupu (privilegovaní)

interní uživatelé

uživatelé v rámci resortu

veřejnost

**Nástroj pro integraci a konsolidaci dat (Integrační platforma)**

načítání dat z různých zdrojů různými způsoby

integrace, transformace, normalizace, denormalizace a anonymizace dat

kontroly, kvalita dat, deduplikace, unifikace, profiling, monitorování dat

**Reportovací a analytický nástroj (Vizualizace)**

**Běžný relační databázový systém RDBMS (např. MS SQL Server, Oracle, Sybase atd.)**

ODS - úložiště pro operativní reporting v reálném čase

Data Stage - vyrovnávací datové úložiště mezi zdrojovými systémy a datovým skladem atd. (L1)

Datový sklad - centralizované normalizované úložiště konsolidovaných dat pro reporting a analýzy (L2)

Datamarty - specializovaná datová úložiště pro konkrétní oblast (L2)

**Nástroj Data Governance Tool**

registr dat

registr obsahu (věcné oblasti, business procesy, pravidla atd.)

vazby a asociace

lineage

**Nástroj Data Governance Tool**

řízení

schvalování

změny

# **5 Položkový rozpočet**

Ceny uvádějte v Kč bez DPH

1. Implementace řešení do prostředí MZe dle výše uvedených požadavků
2. Licenční náklady pro možnost využívání systému v rozsahu
3. Školení 2x/rok v rozsahu 40 uživatelů
4. Dokumentace k systému, uživatelská, systémová
5. Rozvojové činnosti v rozsahu 1000 MD (1MD=8hodin)
6. Zajištění podpory navrženého řešení (cena za 24 měsíců provozu) v níže definovaném SLA

SLA:

Dostupnost / Podpora aplikace v pracovních dnech 5 x 10 = od 8:00 do 18:00, při SLA 99%

| ***Priorita*** | ***Definice priority požadavku*** | ***Parametry řešení požadavku*** |
| --- | --- | --- |
| Priorita 1  Kritická | Některé nebo všechny části plnění selhaly a jsou zcela nefunkční nebo je jejich funkčnost omezena, zároveň je kritickým způsobem ovlivněna činnost Systému. | Odezva: 0,5 hodiny  Obnovení služby: 4 hodiny |
| Priorita 2  Vysoká | Plnění je funkční pouze částečně, zároveň je Systém ovlivněn selháním nebo omezením některé z komponent nebo funkcí podporujících důležité činnosti Systému. | Odezva: 2 hodiny  Obnovení služby: 2 pracovních dnů |
| Priorita 3  Střední | Vyskytují se nedostatky nepodstatné povahy, které způsobují například nekomfortnost obsluhy nebo zvyšující pracnost činností než v běžném provozu plnění. | Odezva: 1 pracovní den  Obnovení služby: 10 pracovních dnů |