



Riadiaci orgán OPIS



Európska únia



MINISTERSTVO FINANCIÍ
SLOVENSKEJ REPUBLIKY

Sprostredkovateľský orgán OPIS

**Európsky fond regionálneho rozvoja
„Tvoríme vedomostnú spoločnosť“**

Projekt je spolufinancovaný Európskou úniou

www.opis.gov.sk

www.informatizacia.sk

Národná agentúra pre sieťové a elektronické služby

**Projekt Elektronické služby Úradu vlády SR –
eDemokracia a otvorená vláda**

PRÍRUČKA POUŽÍVATEĽA MODULU MOD - TRANSFORMAČNÝ MODUL

Verzia dokumentu	1.1
Dátum vydania	16. 11. 2015
ID dokumentu	DOV-1-5-Prirucka_pouzivatela_MOD_Transformacny_modul-v1_1
Autor	Michal Rajniak
Vlastník	NASES

Denník zmien:

Dátum vydania	Verzia	Popis verzie a zmien oproti predošlej verzii	Autor zmeny
22. 9. 2015	1.0	Aktualizácia dokumentu	Michal Rajniak
16. 11. 2015	1.1	Korekcie, doplnenia	Michal Rajniak

Schválenia:

Funkcia	Meno	Verzia	Dátum	Podpis
Programový manažér GlobalTel, a. s.	Peter Marman	1.0	22. 9. 2015	
Projektový manažér NASES	Mário Kulašik	1.0		

Obsah

1	ÚVOD	5
1.1	O DOKUMENTE	5
1.2	PRE KOHO JE DOKUMENT URČENÝ	5
1.3	ČO JE TRANSFORMAČNÝ MODUL	5
1.4	PREČO TRANSFORMAČNÝ MODUL	6
2	SLOVNÍK POJMOV A SKRATIEK.....	6
3	PREHĽAD FUNKCIONALITY	8
3.1	ZÁKLADNÉ ENTITY V TRANSFORMAČNOM MODULE	8
3.2	ROLY V TRANSFORMAČNOM MODULE.....	8
3.3	OBRAZOVKY	8
4	DETAILNÝ POPIS JADRA TRANSFORMAČNÉHO MODULU	9
4.1	PRIHLÁSENIE	9
4.2	ODHLÁSENIE	9
4.3	OBRAZOVKA „DOMOV“ A NAVIGÁCIA V TM	10
4.4	OBRAZOVKA „PROCESY“	10
4.5	OBRAZOVKA EDITOVANIA PROCESU (CANVAS)	12
4.6	OBRAZOVKA „ŠABLÓNY KROKOV“	14
4.7	OBRAZOVKA „MONITOR BEHU“	15
4.8	OBRAZOVKA „PLÁNOVAČ“	16
4.9	OBRAZOVKA „NASTAVENIA“	18
5	INTEGRÁCIA TRANSFORMAČNÉHO MODULU S KATALÓGOM OTVORENÝCH DÁT	19
6	ŽIVOTNÝ CYKLUS POUŽÍVATEĽA V TRANSFORMAČNOM MODULE	19
7	DOKUMENTÁCIA KROKOV	20
7.1	GENEROVANIE METADÁT DÁTOVÉHO ZDROJA.....	20
7.2	STIAHNUTIE SÚBOROV	20
7.3	STIAHNUTIE TABUĽKOVÝCH DÁT Z DATABÁZY	22
7.4	STIAHNUTIE RDF DÁT Z TRIPLESTORE	23
7.5	ULOŽENIE SÚBOROV DO VIRTUOSO TRIPLESTORE	24
7.6	ULOŽENIE DÁT DO DATABÁZY	26
7.7	FILTER SÚBOROV.....	27
7.8	NAHRADENIE VZORU V SÚBORE.....	27
7.9	ZJEDNOTENIE SÚBOROV	28
7.10	PREMENOVANIE SÚBOROV	29
7.11	NAČÍTANIE RDF ZO SÚBORU	30
7.12	VALIDÁCIA XML.....	31
7.13	ZLÚČENIE RDF GRAFOV	31
7.14	ULOŽENIE RDF DO SÚBORU	32

7.15	VALIDÁCIA RDF DÁT	33
7.16	SQL TRANSFORMÁCIA	34
7.17	KONVERZIA TABUĽKY DO RDF	35
7.18	RDF VÝBER POMOCOUP SPARQL CONSTRUCT	36
7.19	RDF VÝBER POMOCOUP SPARQL SELECT	37
7.20	RDF ÚPRAVA POMOCOUP SPARQL UPDATE	37
7.21	NAČÍTANIE TABUĽKOVÝCH DÁT ZO SÚBORU	38
7.22	ROZBALENIE SÚBOROV	39
7.23	XSTL TRANSFORMÁCIA	40
7.24	ZBALENIE SÚBOROV	42
7.25	STIAHNUTIE SÚBORU Z KATALÓGU	42
7.26	STIAHNUTIE TABUĽKOVÝCH DÁT Z KATALÓGU	43
7.27	ULOŽENIE SÚBORU DO KATALÓGU	44
7.28	ULOŽENIE RDF DO KATALÓGU	45
7.29	AKTUALIZÁCIA TABUĽKY V KATALÓGU	45
7.30	ULOŽENIE TABUĽKY DO KATALÓGU	46
7.31	PUBLIKÁCIA DÁT NA ELEKTRONICKÝ PODPIS	47
7.32	EXTRAKCIA METADÁT ZO SÚBORU	48
7.33	EXTRAKCIA TEXTU ZO SÚBORU (OCR)	49
8	SÚVISIACE DOKUMENTY	50

1 Úvod

1.1 O dokumente

Tento dokument (Príručka používateľa MOD – Transformačný modul) slúži pre používateľov Transformačného modulu ako **návod na použitie a referenčná dokumentácia**.

1.2 Pre koho je dokument určený

Dokument je určený pre používateľov Transformačného modulu (ďalej aj „TM“):

- **Povinné osoby**
 - Budú používať Transformačný modul na prípravu a publikáciu svojich dát do Katalógu otvorených dát.
- **Správca transformácií**
 - Správca transformácií je špeciálna rola, ktorá pripravuje predlohy transformácií, ktoré potom môžu Povinné osoby čiastočne prepoužiť alebo sa nimi inšpirovať pri vytváraní vlastných procesov v TM.

Tieto dve skupiny zároveň zodpovedajú dvom používateľským rolám, ktoré sú v TM dostupné: Povinná osoba a Správca transformácií.

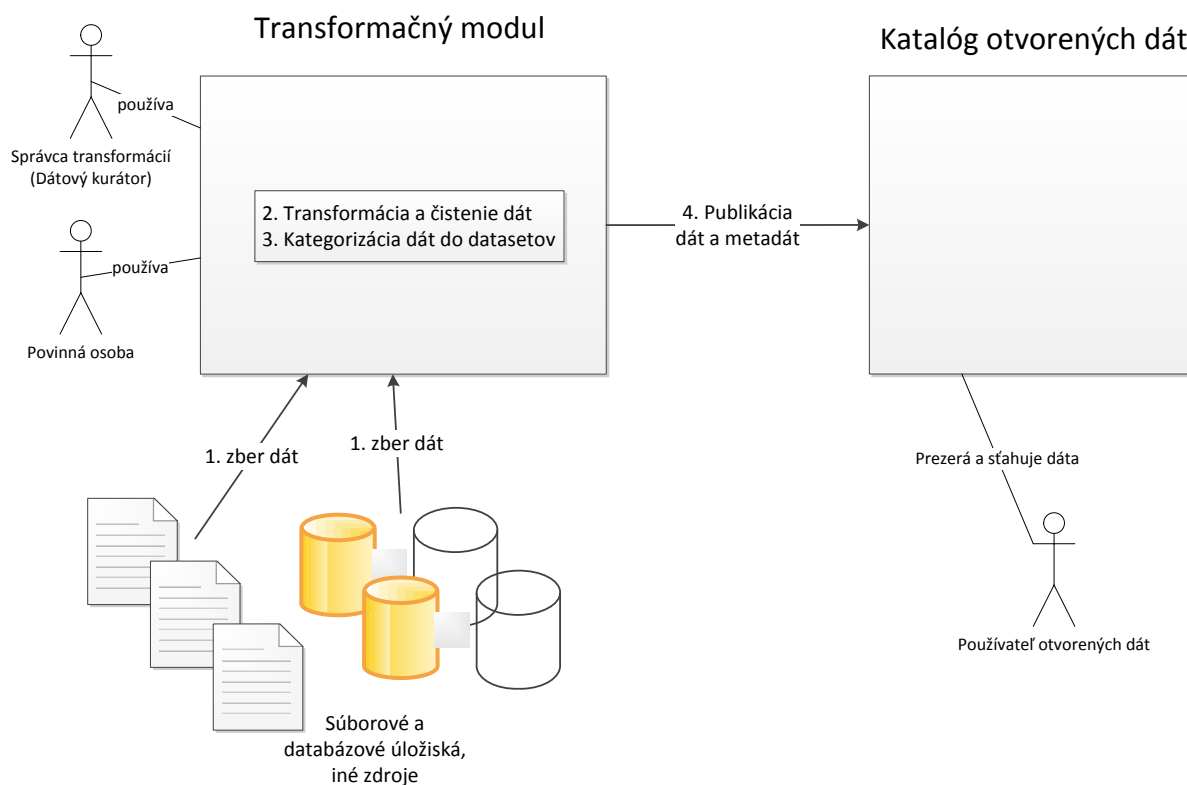
Transformačný modul nie je akýmkoľvek spôsobom sprístupnený pre občanov a verejnosť.

1.3 Čo je Transformačný modul

Transformačný modul je webová aplikácia prepojená s Katalógom otvorených dát, ktorá slúži na prípravu dát a automatizované publikovanie dát do Katalógu otvorených dát.

Transformačný modul umožňuje:

- sťahovanie dát z rôznych zdrojov (databázy, http, ftp, súbory z Katalógu otvorených dát...)
- transformáciu dát (premenovanie, filtrovanie, spájanie, obohacovanie)
 - podpora pre súbory, tabuľkové dáta (SQL tabuľky) a linkované dáta (RDF dáta)
- publikovanie a aktualizáciu dát a metadát do katalógu (súborov, tabuľkových dát, rdf dát)



Obrázok 1 – Prehľad funkcionality Transformačného modulu

1.4 Prečo Transformačný modul

- Automatická publikácia pravidelne vytváraných dát
 - Namiesto ručného nahrávania dát do katalógu je možné definovať v Transformačnom module proces, ktorý bude nahrávať dáta automaticky (napr. každý deň).
- Príprava dát pred publikáciou
 - V prípade, že dáta nie sú predpripravené vo forme vhodnej na publikovanie ako otvorené dáta, môžu byť definované transformačné predpisy, ktoré dáta „pripraví“. Toto je hlavná funkcionality Transformačného modulu a bude rozpísaná v tejto príručke detailne.

2 Slovník pojmov a skratiek

Pojem	Význam
Beh, Beh procesu	Reprezentuje jedno konkrétne spustenie procesu. Alternatívne názvy: exekúcia, exekúcia procesu; anglický názov: pipeline execution.

Katalóg	pozri Katalóg otvorených dát.
Katalóg otvorených dát	Webová aplikácia dostupná na https://data.gov.sk , kde sú publikované otvorené dáta Povinných osôb.
Krok, Krok procesu	Stavebná jednotka v procese v TM. Krok je jedna aktivita (jedna „krabíčka“) v procese, napr. načítanie dát z databázy. Alternatívny názov: DPU Anglický názov: DPU (Data Processing Unit)
Linkované dáta	Dáta vo formáte RDF, kde jednotlivé prvky dát sú prepojené väzbami. Pozri aj: https://en.wikipedia.org/wiki/Linked_data
Metadáta	Informácie o dátach. Metadáta sú napr. názov, popis, autor, dátum zverejnenia, stav, dátum poslednej aktualizácie, dátum platnosti od-do, dátum účinnosti od-do, formát dát, licencia.
Načasovanie, Načasovanie procesu	Pravidlo vytvorené v TM, podľa ktorého je konkrétny proces spustený buď jednorazovo v konkrétnom čase, alebo je spúšťaný pravidelne v definovaných intervaloch (napr. raz denne o 2.00 ráno).
Povinná osoba	Jedna z používateľských rol v Transformačnom module. Tento termín má zároveň aj legislatívnu definíciu, pozri § 2 zákona č. 211/2000 (zákon o slobodnom prístupe k informáciám).
Predloha procesu	Ide o štandardný proces s jedným rozdielom: je viditeľný pre všetkých používateľov. Takéto procesy môže vytvárať iba rola Správca transformácií. Úmysel je, že Správca transformácií bude vytvárať užitočné predlohy procesov, ktoré si môžu Povinné osoby skopírovať a inšpirovať sa nimi pri tvorbe vlastných procesov.
Proces	Postupnosť krokov, ktorej cieľom je vypublikovať dáta a metadáta do niektorého datasetu v katalógu. Procesy v TM sú vytvárané používateľmi. Alternatívny názov: transformačný proces, transformačný predpis Anglický názov: pipeline
RDF	(Resource Description Framework) Štandardný model a formát používaný pre linkované dáta .
Správca transformácií	Jedna z používateľských rol v Transformačnom module.
Šablóna kroku	Predkonfigurovaný krok.
TM	Transformačný modul.
Transformačný modul	Pozri 1.3 Čo je Transformačný modul
UPVS	Ústredný portál verejnej správy, https://www.slovensko.sk

3 Prehľad funkcionality

3.1 Základné entity v Transformačnom module

- Proces
- Krok procesu
- Šablóna kroku
- Beh procesu
- Načasovanie procesu

Definície týchto pojmov sú uvedené v kapitole 2 Slovník pojmov a skratiek.

3.2 Roly v Transformačnom module

Používateľ môže pracovať v TM v nasledujúcich rolách:

- Rolu **Povinná osoba** majú používatelia z organizácií, ktoré su Povinnými osobami a publikujú prostredníctvom TM dáta do Katalógu otvorených dát.
- Rola **Správca transformácií** je podporná rola, ktorá ma pomáhať Povinným osobám s vytváraním a so správou procesov. Túto rolu bude pravdepodobne zastrešovať prevádzkovateľ MOD-u. Oproti funkcionalitám dostupným Povinným osobám má táto rola navyše oprávnenia vytvárať predlohy procesov a krokov, ktorými sa môže Povinná osoba inšpirovať pri vytváraní vlastných procesov.

Kombinácia rolí sa nevylučuje, t. j. používateľ môže byť súčasne Povinná osoba aj Správca transformácií.

3.3 Obrazovky

TM sa skladá z nasledujúcich obrazoviek:

- Domov
 - úvodná stránka TM, obsahuje základné info a link na dokumentáciu
- Procesy
 - zoznam procesov a možnosť ich vytvárania, editovania a spúšťania
- Šablóny krokov (iba pre rolu Správca transformácií)
 - zoznam dostupných krokov, možnosť konfigurácie šablón krokov
- Monitor behu
 - zoznam behov procesov a inšpekcia detailov jednotlivých behov
- Plánovač
 - zoznam načasovaní procesov, možnosť ich vytvárania a editovania
- Nastavenia
 - nastavenia e-mailových notifikácií

Jednotlivé obrazovky a funkcionality dostupné na jednotlivých obrazovkách budú detailnejšie popísané v ďalších kapitolách.

4 Detailný popis jadra Transformačného modulu

4.1 Prihlásenie

4.1.1 Prihlásenie cez data.gov.sk

Transformačný modul je súčasťou Ústredného portálu verejnej správy (ďalej aj „ÚPVS“), s ktorým zdieľa aj prihlasovací mechanizmus.

Existuje niekoľko postupov, ako sa prihlásiť.

4.1.2 Prihlásenie cez Katalóg otvorených dát <https://data.gov.sk/>

Najbežnejší spôsob prihlásenia je cez Katalóg otvorených dát <https://data.gov.sk/>. Postup je nasledujúci:

1. Používateľ otvorí v prehliadači adresu <https://data.gov.sk/> a klikne na Prihlásiť.
2. Používateľ je presmerovaný na portál ÚPVS.
3. Používateľ sa prihlási na ÚPVS štandardným spôsobom pomocou eID (elektronický občiansky preukaz) alebo iným mechanizmom.
 - a. V druhom kroku prihlasovania **je potrebné vybrať organizáciu, za ktorú sa do Transformačného modulu chce používateľ prihlásiť**.¹
4. Po úspešnom prihlásení cez ÚPVS je používateľ presmerovaný späť na <https://data.gov.sk/>.
5. Na úvodnej stránke klikne na link Transformačný modul.
6. Používateľ je presmerovaný do Transformačného modulu a môže pracovať.

4.1.3 Priame prihlásenie

Rýchlejší spôsob prihlásenia – obíde sa Katalóg otvorených dát. Postup je nasledujúci:

1. Používateľ otvorí v prehliadači adresu <https://data.gov.sk/transformacnymodul>, čo ho presmeruje na portál ÚPVS.
2. Používateľ sa prihlási na ÚPVS štandardným spôsobom pomocou eID alebo iným mechanizmom.
 - a. V druhom kroku prihlasovania **je potrebné vybrať organizáciu, za ktorú sa do Transformačného modulu chce používateľ prihlásiť**.
3. Po úspešnom prihlásení cez ÚPVS je používateľ presmerovaný do Transformačného modulu a môže pracovať.

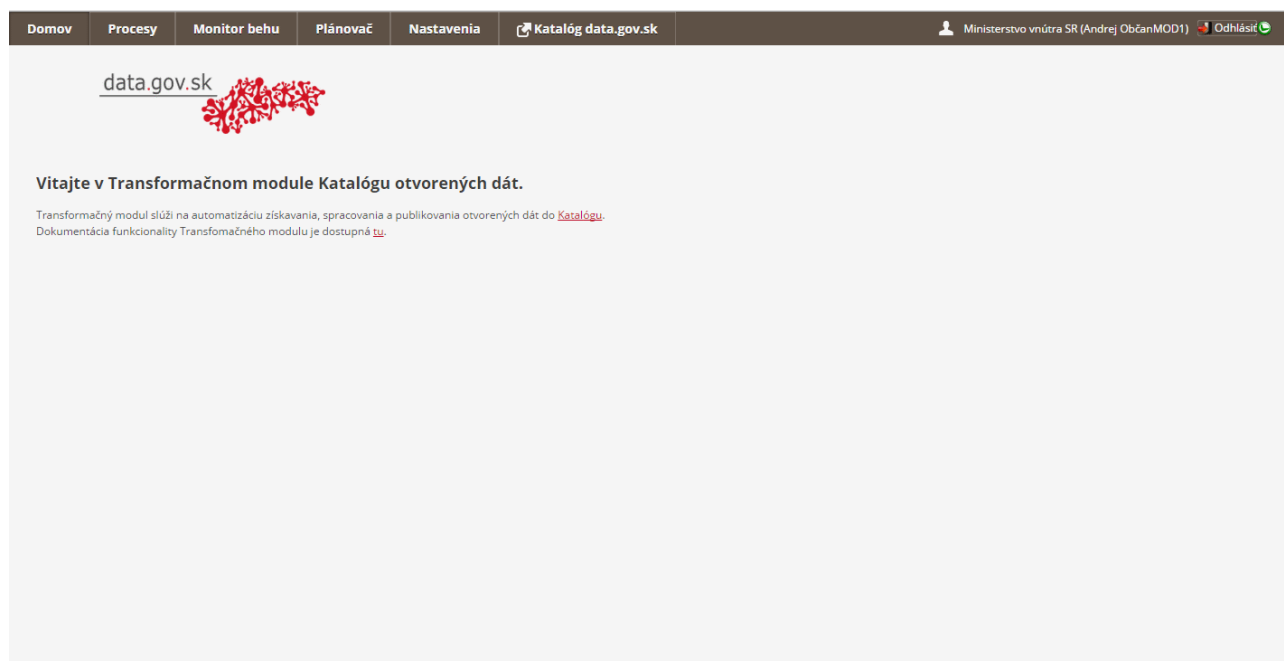
4.2 Odhlásenie

Používateľ sa odhlási kliknutím na link Odhlásiť umiestnený v pravom hornom rohu na každej obrazovke Transformačného modulu.

Upozornenie: Týmto sa používateľ odhlási nielen z Transformačného modulu, ale z celého portálu ÚPVS. V prípade, že chce ukončiť prácu v Transformačnom module bez odhlásenia, stačí jednoducho v prehliadači zavrieť okno (resp. záložku) s Transformačným modulom.

¹ Prihlásiť sa do TM ako fyzická osoba (občan sám za seba) nie je možné. Prihlásiť a pracovať v TM je možné iba za organizáciu.

4.3 Obrazovka Domov a navigácia v TM



Obrázok 2 – Obrazovka Domov

Úvodná stránka TM, ktorá obsahuje základné info a link na dokumentáciu.

V hornej časti sa nachádza menu.

Posledný link v menu je presmerovanie na Katalóg otvorených dát (<https://data.gov.sk/>).

V pravej hornej časti obrazovky je informácia o aktuálne prihlásenom používateľovi a link na odhlásenie z aplikácie.

4.4 Obrazovka Procesy

Akcie	Názov	Vytvoril	Začiatok posledného spracovania	Trvanie posledného spracovania	Posledný stav
[Icons]	[predloha) Načítanie RDF dát zo súboru a uloženie do katalógu	NASES (Andrej SprávcaTransformáciíMOD)	27.10.2015 9:46:01	0:00:08	✓
[Icons]	Geo data test proces	Ministerstvo vnútra SR (Andrej ObčanMOD1)			
[Icons]	[AutoTest) CSV to relational to CKAN	Ministerstvo vnútra SR (Andrej ObčanMOD1)	16.11.2015 0:10:44	0:02:41	✓
[Icons]	[predloha) nový proces	NASES (Fero FO)			
[Icons]	[AutoTest) RelationalFromSql-Relational-RelationalToCkan	Ministerstvo vnútra SR (Andrej ObčanMOD1)	16.11.2015 0:51:00	0:03:43	✓
[Icons]	[Autotest) FilesDownload - OCR - FilesToCKAN	Ministerstvo vnútra SR (Andrej ObčanMOD1)	16.11.2015 1:00:42	0:00:13	✗
[Icons]	[performance) Spracovanie veľkej tabuľky	Ministerstvo vnútra SR (Andrej ObčanMOD1)	11.11.2015 17:31:11	20:13:54	✗
[Icons]	Meteorologické dáta - aktualizácia	Ministerstvo vnútra SR (Andrej ObčanMOD1)			
[Icons]	[AutoTest) RDF from file - Merge RDF - Validate RDF - RDF to CKAN	Ministerstvo vnútra SR (Andrej ObčanMOD1)	16.11.2015 15:06:54	0:01:08	✗
[Icons]	Periodické sťahovanie meteo dát	Ministerstvo vnútra SR			
[Icons]	Priebežná aktualizácia dátového zdroja (MV)	Ministerstvo vnútra SR			
[Icons]	[AutoTest) FilesDownload to FilterValidom to CKAN	Ministerstvo vnútra SR (Andrej ObčanMOD1)	16.11.2015 0:35:43	0:01:50	⚠
[Icons]	[AutoTest) FilesDownload - XSLT - file to CKAN	Ministerstvo vnútra SR (Andrej ObčanMOD1)	16.11.2015 0:55:33	0:02:00	⚠
[Icons]	[performance) OCR veľkého a komplexného dokumentu	Ministerstvo vnútra SR (Andrej ObčanMOD1)	3.11.2015 14:48:39	7:36:54	✓
[Icons]	[AutoTest) RelationalFromSql-Relational-RelationalToCkan	Ministerstvo vnútra SR (Andrej ObčanMOD1)	16.11.2015 0:45:45	0:01:54	✓

Obrázok 3 – Obrazovka Procesy

Táto obrazovka obsahuje zoznam procesov a možnosť ich vytvárania, editovania, spúšťania, filtrovania atď.

Zoznam procesov je zobrazený v tabuľke.

V prípade, že je procesov viac, ako sa zmestí na jednu obrazovku, je možné sa cez procesy stránkovať.

Pod horným menu sú dostupné tlačidlá.

- Vytvoriť proces
 - presmeruje používateľa na obrazovku vytvorenia nového procesu
- Vyčistiť filtre
 - vyčistí filtre zadané na obrazovke, ak boli nejaké zadané
- Vyčistiť triedenie
 - zruší používateľom zvolené triedenie podľa vybraného stĺpca

V stĺpci Akcie sú dostupné tlačidlá na vykonávanie akcií nad jednotlivými procesmi:



- Spustiť
- Načasovať
- Skopírovať
- Editovať
- Zmazať

Jednotlivé tlačidlá a akcie sú popísané nižšie.

4.4.1 Filtrovanie a usporadúvanie procesov

Procesy je možné **filtrovať** zadáním hodnoty do poľa nad daným stĺpcom. Takto sa dá filtrovať napr. podľa názvu procesu alebo stavu posledného spracovania procesu.

Procesy je možné **usporiadať** podľa hodnôt vo vybranom stĺpci kliknutím na hlavičku príslušného stĺpca.

4.4.2 Vytvorenie nového procesu

Na obrazovke Procesy klikne používateľ na tlačidlo Vytvoriť proces.

Otvorí sa obrazovka vytvárania/editácie procesu. Táto je popísaná v kapitole 4.5 Obrazovka editovania procesu .

4.4.3 Vytvorenie predlohy procesu (iba Správca transformácií)

Na obrazovke Procesy klikne používateľ na tlačidlo Vytvoriť proces s tým, že na obrazovke vytvárania procesu nastaví parameter **Viditeľnosť** na **verejný (len na čítanie)**. To spôsobí, že proces budú vidieť aj ostatní používatelia TM a budú si ho môcť aj skopírovať.

Obrazovka vytvárania/editácie procesu je detailne popísaná v kapitole 4.5 Obrazovka editovania procesu

4.4.4 Skopírovanie existujúceho procesu

Na obrazovke Procesy je potrebné nájsť príslušný proces v zozname a pri danom procese kliknúť na tlačidlo Kopírovať.

Týmto sa vytvorí kópia procesu a bude pomenovaná ako Kópia #(číslo kópie) z (názov kopírovaného procesu).

4.4.5 Spustenie procesu

Na obrazovke Procesy je potrebné nájsť príslušný proces v zozname a pri danom procese kliknúť na tlačidlo Spustiť.

Týmto sa proces spustí a na obrazovke bude zobrazená informácia, že proces bol úspešne spustený. Po spustení je beh procesu možné zobrazíť na obrazovke Monitor behu.

4.4.6 Načasovanie procesu

Na obrazovke Procesy je potrebné nájsť príslušný proces v zozname a pri danom procese kliknúť na tlačidlo Naplánovať spustenie.

Zobrazí sa obrazovka Načasovania procesu, pozri 4.8.1 Obrazovka Načasovanie procesu.

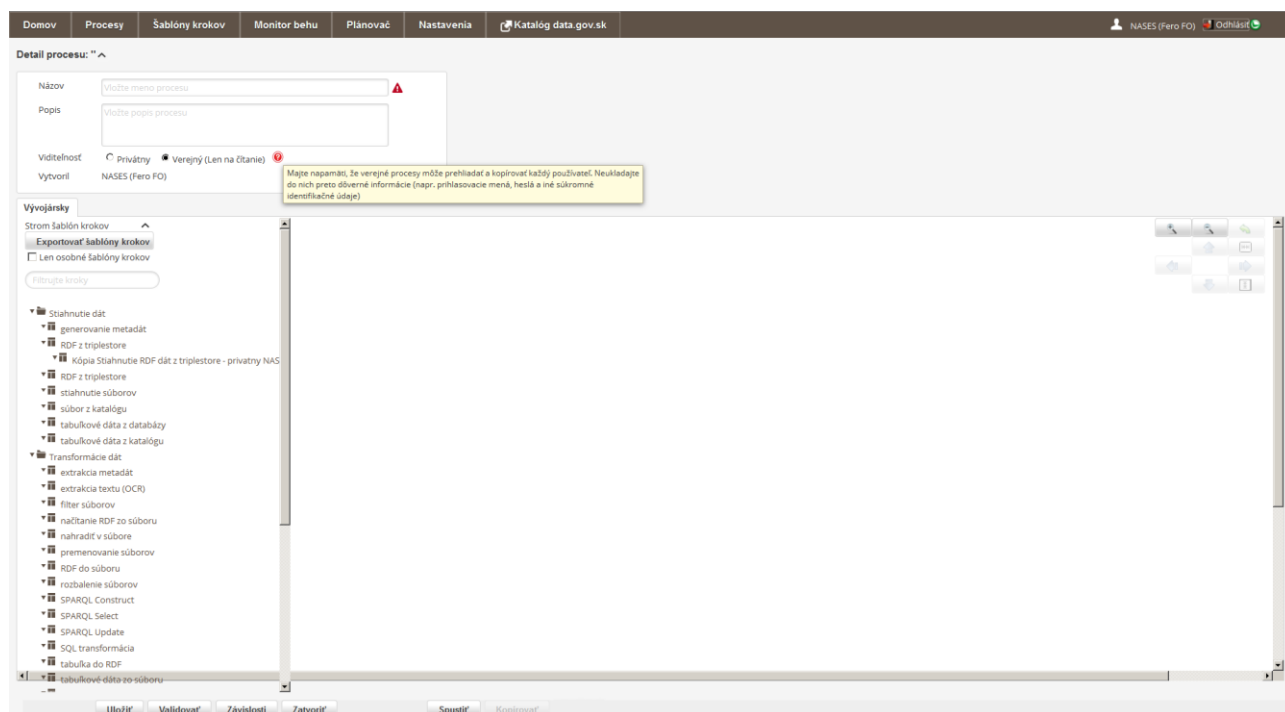
4.4.7 Zmazanie procesu

Na obrazovke Procesy je potrebné nájsť príslušný proces v zozname a pri danom procese kliknúť na tlačidlo Zmazať.

4.5 Obrazovka editovania procesu (Canvas)

Na tejto obrazovke² sa proces vytvára aj edituje.

V tejto podkapitole je popísaný len všeobecný spôsob editácie procesu: pridávanie krokov, prepájanie krokov hranami, mazanie krokov atď. Konfigurácia *obsahu* jednotlivých krokov je popísaná v samostatnej kapitole 7 Dokumentácia krokov.



Obrázok 4 – Obrazovka editovania procesu

V ľavej hornej časti je možné zadať/zmeniť názov a popis procesu.

² Táto obrazovka nie je dostupná priamo v menu. Používateľ sa na ňu dostane, keď zvolí vytvorenie nového procesu alebo editáciu existujúceho procesu.

Rola *Správca transformácií* tu takisto môže zvoliť **Viditeľnosť** procesu: **privátny** alebo **verejný (len na čítanie)**. Verejný proces môžu vidieť všetci používatelia TM a môžu sa daným procesom inšpirovať pri vytváraní vlastných procesov alebo si ho skopírovať a upraviť pre vlastné potreby. Rola *Povinná osoba* môže vytvárať iba privátne procesy, t. j. procesy, ktoré budú vidieť a môcť editovať iba používatelia v rámci danej Povinnej osoby.

Na **ľavej strane** je zoznam dostupných krokov a ich predkonfigurovaných šablón. Je možné medzi nimi vyhľadávať (filtrovať) pomocou textového poľa nad ich zoznamom.

Hlavná časť obrazovky je priestor na grafickú editáciu procesu – **editačný priestor**.

Kroky v zozname sú združené do 4 skupín:

- Stiahnutie dát
 - tieto kroky slúžia na sťahovanie dát z rôznych zdrojov, databáz, z internetu, ale napríklad aj z katalógu
- Transformácia dát
 - tieto kroky slúžia na vykonávanie rôznorodých transformácií nad dátami, ktoré už boli stiahnuté do procesu
- Publikácia dát
 - tieto kroky slúžia na uloženie pripravených dát a metadát do Katalógu otvorených dát
- Kontrola kvality
 - tieto kroky slúžia na kontrolu kvality a validáciu dát, napr. na kontrolu, či sú vstupné XML dáta validné voči schéme a pod.

Editačné akcie pre proces:

- **Pridať krok do procesu** je realizované výberom kroku v zozname a presunutím ho (drag-and-drop) do editačného priestoru.

Ostatné akcie sú dostupné z **kontextového menu každého kroku**, ktoré sa dá zobrazíť presunutím myši na daný krok v editačnom priestore:

- Vytvoriť novú hranu
 - pridá hranu z vybraného kroku do iného kroku v procese
- Detaily
 - zobrazí konfiguráciu kroku. Detaily konfigurácie jednotlivých krokov sú uvedené v samostatnej kapitole 7 Dokumentácia krokov
- Kopírovať krok
 - skopíruje daný krok a umiestni ho do editačného priestoru
- Upraviť polohu
 - upraví polohu daného kroku v editačnom priestore
- Odstrániť krok
 - odstráni krok z procesu, rovnako ako aj všetky jeho hrany

Tlačidlá v dolnej časti:

- Uložiť
 - uloží proces
- Validovať

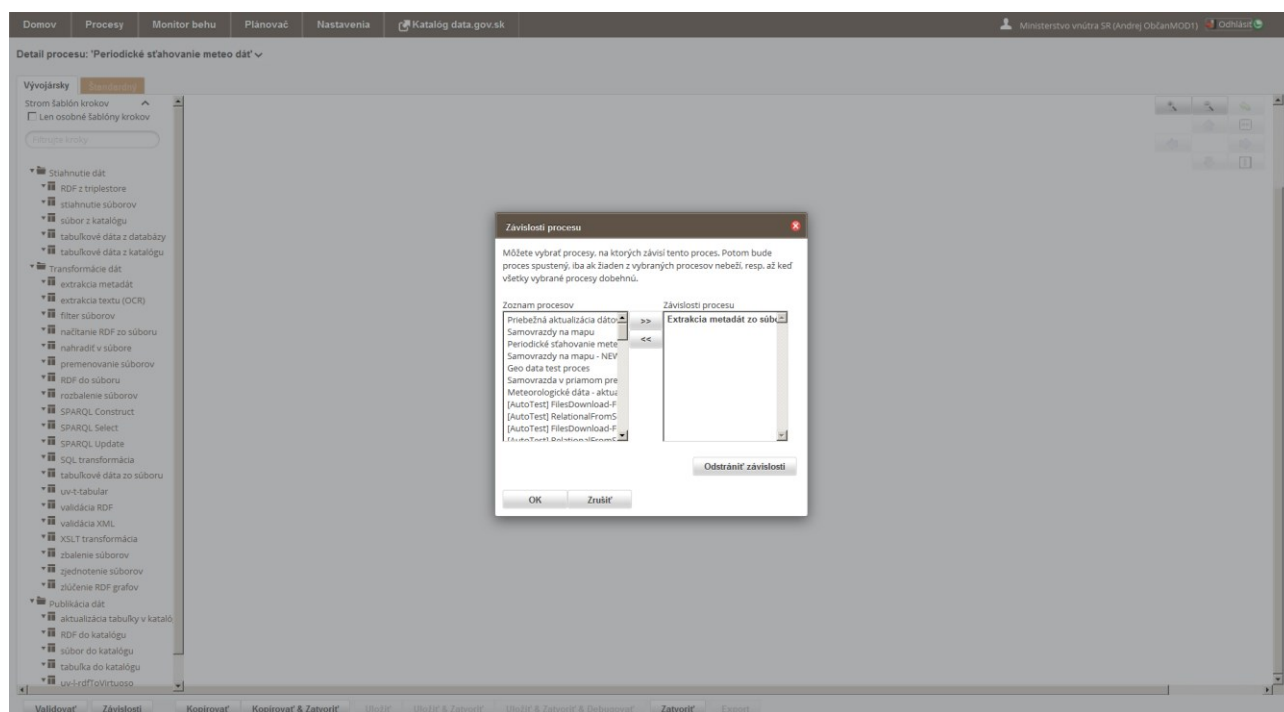
- skontroluje, či sa v definícii procesu nenachádzajú zjavné chyby (to negarantuje, že proces je „bezchybný“, iba skontroluje niektoré zjavné chyby)
- Závislosti
 - zobrazí obrazovku na zadefinovanie závislostí procesu
- Zatvoriť
 - zatvorí proces a obrazovku editácie procesu
- Spustiť
 - spustí proces
- Kopírovať
 - vytvorí nový proces skopírovaním tohto procesu

4.5.1 Definovanie závislosti medzi procesmi

Na tejto obrazovke je možné definovať „závislosti“ procesu na iných procesoch.

To znamená, že proces bude proces spustený, iba ak žiaden z vybraných procesov nebeží, resp. až keď všetky vybrané procesy dobehnú.

Definovať takúto závislosť môže byť užitočné napríklad v prípade, keď vykonanie aktuálneho procesu závisí od výstupov iného procesu.



Obrázok 5 – Definovanie závislostí procesu

4.6 Obrazovka Šablóny krokov

(iba pre rolu Správca transformácií)

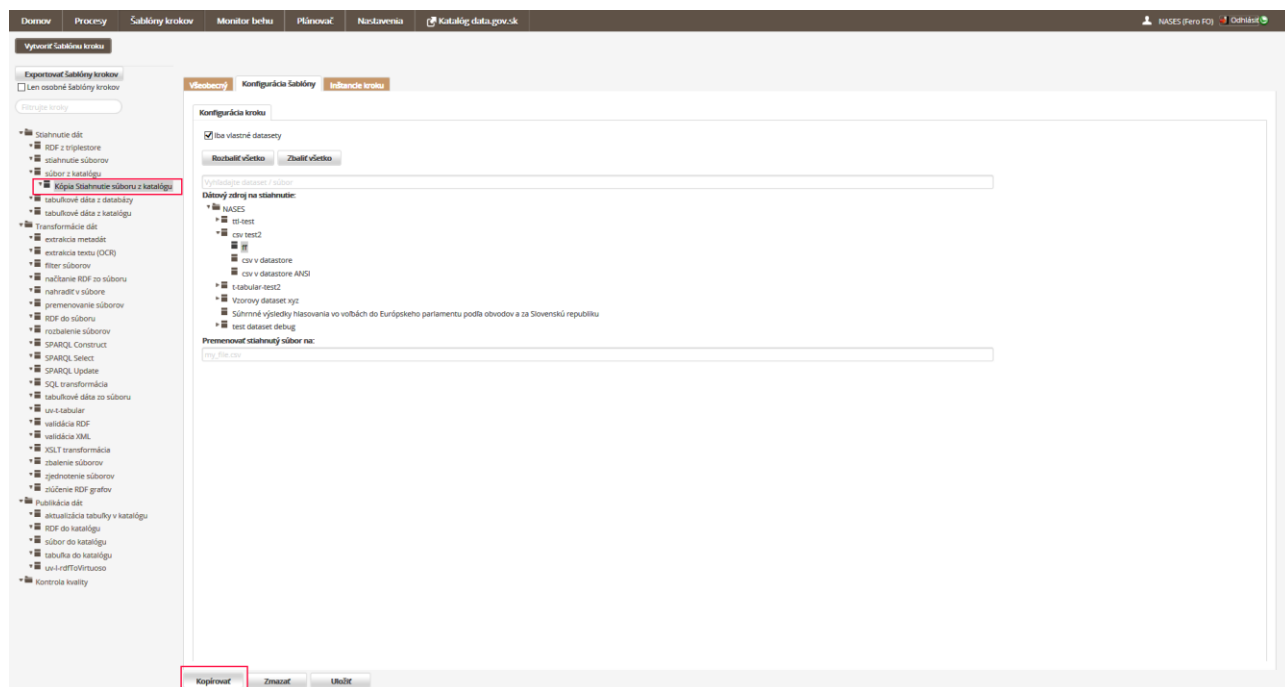
Táto obrazovka obsahuje zoznam všetkých dostupných krokov použiteľných pri vytváraní jednotlivých procesov. Je tu možné definovať *šablóny krokov*. Šablóna kroku je, zjednodušene povedané,

nakonfigurovaný krok. Prínos: ak existuje krok, ktorý má zložitú konfiguráciu a používateľ ho používa často, nakonfigurovaním šablóny si ušetrí čas.

Na tejto obrazovke je možné:

- zobraziť zoznam všetkých dostupných krokov
- prezerať detail kroku
- vytvárať šablóny kroku a konfigurovať ich

Šablónu kroku je možné vytvoriť skopírovaním ľubovoľnej inej šablóny kroku zo stromu. Kópia šablóny kroku bude umiestnená do stromu po nadriadenú šablónu, pozri obrázok.



Obrázok 6 – Vytvorenie šablóny kroku skopírovaním

4.7 Obrazovka Monitor behu

Obrazovka obsahuje zoznam všetkých behov usporiadaných v tabuľke od najnovšieho behu.

Po vybraní konkrétneho behu sú na pravej strane obrazovky zobrazené detailné informácie o behu daného procesu. Pozri obrázok.

Obrázok 7 – Obrazovka *Monitor behu*

Na obrazovke Monitor behu je možné vykonávať nasledujúce:

- Zobrazíť a prezerať zoznam behov
- Filtrovať a usporiadať behy podľa názvu, autora, stavu a iných parametrov
 - podobne ako na obrazovke Procesy
- Zobrazíť detail behu
 - zobrazí sa pravá časť obrazovky – t. j. logy a udalosti
- Zobrazíť udalosti v rámci behu
 - v záložke na pravej časti obrazovky; zobrazia sa dôležité udalosti v behu procesu
- Zobrazíť logy v rámci behu
 - na pravej časti obrazovky; zobrazia sa detailnejšie záznamy (logy) o behu procesu
- Vyhľadávať v udalostiach a logoch behu podľa viacerých parametrov
 - podobne ako na obrazovke Procesy

4.8 Obrazovka Plánovač

Obrazovka Plánovač obsahuje zoznam načasovaní procesov, možnosť ich vytvárania a editovania

Domov

Procesy

Súhrnný kľúčok

Monitor behu

Plánovač

Nastavenia

Katalóg dát: gov.uk

Nové načasovanie

Vytvoriť filter

Akcie	Stav	Proces	Pravidlo	Posledný	Nasledujúci	Trvanie	Naplánovať
<div><div>🛑</div><div>🔄</div><div>🗑️</div></div>	<div>✅</div>	Dlžnica sociálnej poisťovne	Spustiť 12.8.2015 8:30:00 s periódou opakovania: deň	12.8.2015 22:12:41	13.8.2015 8:30:00	0:00:05	NADES (Petro FQ)
<div><div>🛑</div><div>🔄</div><div>🗑️</div></div>	<div>✅</div>	Skutočný transf. predpis	Spustiť po procese Dlžnica sociálnej poisťovne.				NADES (Petro FQ)

1

Strana: 1 / 1

22

Počet záznamov: 2

Obrázok 8 – Obrazovka *Plánovač*

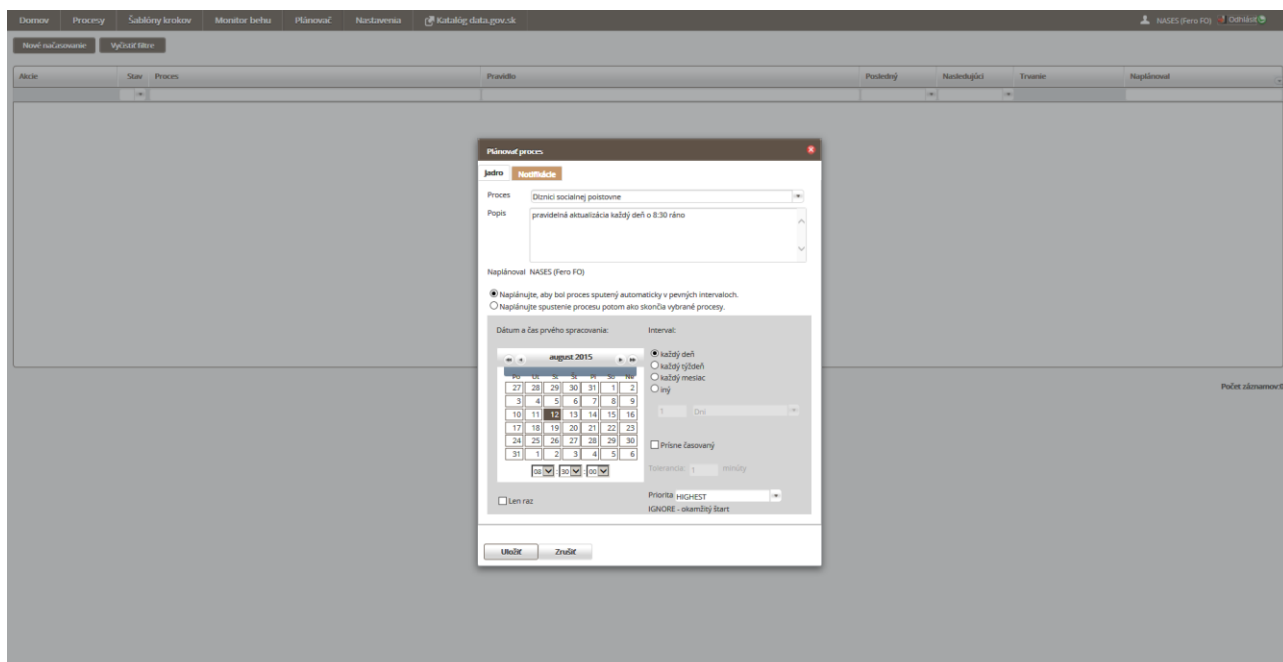
Dostupné funkcionality:

- Zobrazíť prehľad existujúcich načasovaní
 - tabuľkové zobrazenie podobne ako na ostatných obrazovkách
- Filtrovať a usporiadať načasovania procesov podľa viacerých parametrov
 - podobne ako na ostatných obrazovkách
- Vytvoriť nové načasovanie procesu
 - detaily pozri 4.8.1 Obrazovka Načasovanie procesu
- Vytvoriť nové načasovanie procesu – pravidelné spúšťanie procesu
- Zobrazíť detail načasovania procesu
- Upraviť načasovanie procesu
- Zmazať načasovanie procesu

4.8.1 Obrazovka Načasovanie procesu

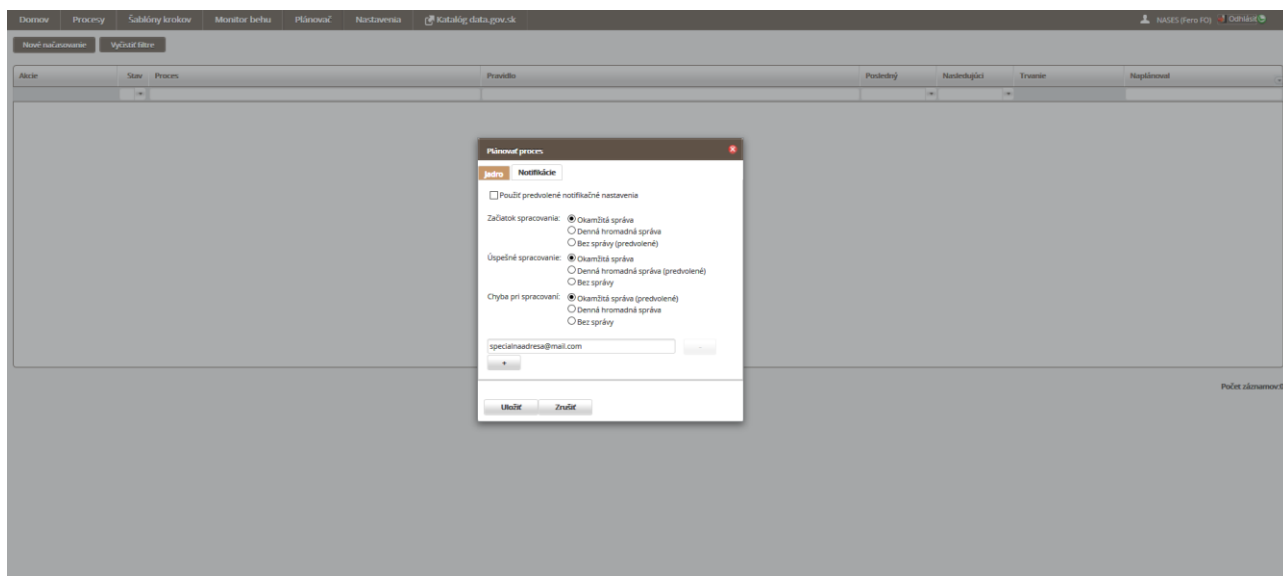
Na obrazovke Načasovanie procesu je možné napláňovať jednorazové alebo pravidelné spustenie procesu. Funkcionality:

- načasovanie na jednorazové spustenie v konkrétny dátum a čas
- načasovanie na pravidelné spúšťanie každý deň, týždeň, mesiac v zadanom čase
- je možné definovať prioritu spustenia procesu
 - priorita je náporada pre systém, nie je však pre systém záväzná. Odporúča sa ponechať predvolenú prioritu – HIGHEST (najvyššia)



Obrázok 9 – Obrázovka *Načasovanie procesu*

V záložke *Notifikácie* je možné definovať notifikácie špeciálne pre tento proces, ináč sa použijú predvolené notifikačné nastavenia definované na obrazovke *Nastavenia*. V príklade nižšie je uvedená konfigurácia, ktorá pošle namiesto predvolených nastavení okamžitú správu pri akejkoľvek udalosti v procese.



Obrázok 10 – Obrázovka *Načasovanie procesu* – záložka *Notifikácie*

4.9 Obrázovka *Nastavenia*

Na obrazovke *Nastavenia* je možné:

- Nastaviť notifikačné e-mailové adresy
 - je možné zvoliť jednu alebo viac adries
- Nastaviť počet riadkov (záznamov) na zobrazovanie vo všetkých tabuľkách v aplikácii

- Konfigurovať notifikácie pre behy procesov
 - notifikáciu pri spustení a úspešnom aj neúspešnom behu procesov

Obrázok 11 – Obrazovka *Nastavenia* – e-mailové adresy a počet riadkov

Obrázok 12 – Obrazovka *Nastavenia* – notifikácie k behu procesov

5 Integrácia Transformačného modulu s Katalógom otvorených dát

V Transformačnom module používateľ definuje procesy na publikovanie dát, nešpecifikuje v ňom však (ani v jednotlivých procesoch), **do ktorého datasetu** sa majú dáta z procesu vypublikovať.

Toto sa definuje až v Katalógu otvorených dát pri jednotlivých datasetoch. T. j. pri vybranom datasete v katalógu používateľ vyberie, ktorý proces (alebo procesy) majú publikovať dáta do procesu.

Táto funkcionálna je popísaná v Príručke používateľa pre Katalóg otvorených dát, pozri [Katalog].

6 Životný cyklus používateľa v Transformačnom module

- Každému Subjektu (Povinnej osobe, Správcovi transformácií) je zriadený účet v Transformačnom module automaticky po tom, ako sa doňho prvýkrát prihlási niekto (t. j. nejaký Actor) v jeho mene (pozri kapitolu 4.1 Prihlásenie).
- V prípade, že Subjekt už ďalej nemá/nechce používať Transformačný modul, je potrebné, aby pred ukončením činnosti **zrušil všetky plánované behy všetkých svojich procesov** (ak nejaké existujú). Ide o to, aby po ukončení používania Transformačného modulu procesy Subjektu ďalej **neprodukovali žiadne dáta**.

7 Dokumentácia krokov

7.1 Generovanie metadát dátového zdroja

7.1.1 Identifikátor

e-distributionMetadata

7.1.2 Typ

Stiahnutie dát.

7.1.3 Popis

Generuje metadáta na základe konfigurácie.

7.1.4 Konfiguračné parametre

Meno	Popis
Nadpis distribúcie	Meno definované pre distribúciu
Popis	Popis distribúcie
Formát	Formát súboru distribúcie
MIME type	Média typ distribúcie ako je definované IANA
Dátum vytvorenia	Dátum a čas formálneho vytvorenia (resp. publikovania) distribúcie

7.1.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
distributionOutput	výstup	RDFDataUnit	Metadáta distribúcie. Uložené v grafe s menom distributionMetadata	áno

7.2 Stiahnutie súborov

7.2.1 Identifikátor

e-filesDownload

7.2.2 Typ

Stiahnutie dát.

7.2.3 Popis

Stiahne súbory podľa zoznamu definovaného v konfigurácii. Je možné stiahnuť samostatné súbory alebo aj celé adresáre.

7.2.4 Konfiguračné parametre

Meno	Popis
Zoznam súborov a adresárov na stiahnutie	Keď je uvedený adresár, stiahnu sa všetky súbory v adresári a jeho podadresároch.

Ak je zadáný názov každého sťahovaného vstupu, tento názov sa použije na vnútornú identifikáciu daného súboru v ďalšom pokračovaní procesu a tiež ako názov virtuálnej cesty (cieľové umiestnenie súboru pri nahrávaní mimo UnifiedViews na konci procesu). Ak vás nezaujíma vnútorné pomenovanie súboru alebo názov virtuálnej cesty, napr. v prípadoch, keď potrebujete iba prechádzať stiahnuté súbory v pokračovaní procesu tým istým spôsobom, nie je potrebné špecifikovať meno súboru.

7.2.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
config	vstup	RdfDataUnit	Dynamická RDF konfigurácia, pozri Pokročilá konfigurácia	
output	výstup	FilesDataUnit	Stiahnuté súbory	áno

7.2.6 Pokročilá konfigurácia

Krok je možné nakonfigurovať aj dynamicky cez vstup `config` pomocou RDF dát.

Vzor konfigurácie:

```
<http://localhost/resource/config>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://unifiedviews.eu/ontology/dpu/filesDownload/Config>;
  <http://unifiedviews.eu/ontology/dpu/filesDownload/hasFile>
<http://localhost/resource/file/0>.
<http://localhost/resource/file/0>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://unifiedviews.eu/ontology/dpu/filesDownload/File>;
  <http://unifiedviews.eu/ontology/dpu/filesDownload/file/uri>
"http://www.zmluvy.gov.sk/data/att/117597_dokument.pdf";
  <http://unifiedviews.eu/ontology/dpu/filesDownload/file/fileName>
"zmluva.pdf".`
```

7.3 Stiahnutie tabuľkových dát z databázy

7.3.1 Identifikátor

e-relationalFromSql

7.3.2 Typ

Stiahnutie dát.

7.3.3 Popis

Načíta relačné dáta (tabuľku) z externej relačnej databázy.

Používa sa na načítanie dát z relačnej databázy pomocou SQL dotazov a následné uloženie do internej dátovej hrany určenej pre relačné dáta.

Tento krok umožňuje využiť niektoré vlastnosti na načítanie dát z databázy: zoznam tabuliek v zdrojovej databáze, generovanie SELECT dotazu pre zvolenú tabuľku, preview údajov.

Podporované je aj bezpečné pripojenie do externej databázy cez SSL.

Pre bezpečnosť sa veľmi odporúča používať databázového používateľa, ktorý má práva iba na čítanie z danej databázy.

Krok podporuje nasledujúce databázy:

- * PostgreSQL
- * Oracle
- * MySQL
- * Microsoft SQL

7.3.4 Konfiguračné parametre

Meno	Popis
Typ databázy	Typ databázy: PostgreSQL, Oracle, MySQL, MS SQL
Host	Host adresa databázy
Port	Port databázy
Meno databázy	Meno databázy (pre ORACLE SID)
Meno inštalácie	(nepovinné) Meno inštalácie databázy - pre MSSQL
Meno používateľa	Heslo na prihlásenie do databázy
Heslo	Heslo na prihlásenie do databázy

Meno	Popis
Pripojiť cez SSL	Či sa má použiť zabezpečené pripojenie (SSL)
Meno cieľovej tabuľky	Meno tabuľky v procese, do ktorej sa uložia dáta extrahované z externej databázy
SQL dotaz	SQL dotaz, ktorý získa dáta z externej databázy
Stĺpce primárneho kľúča	(nepovinné) Mená stĺpcov, ktoré tvoria primárny kľúč tabuľky, oddelené čiarkami
Indexed columns	(nepovinné) Mená stĺpcov, pre ktoré je potrebné vytvoriť indexy, oddelené čiarkami. Môže zvýšiť výkon databázy

7.3.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
outputTables	výstup	RelationalDataUnit	Naplnené databázové tabuľky	áno

7.4 Stiahnutie RDF dát z triplestore

7.4.1 Identifikátor

e-sparqlEndpoint

7.4.2 Typ

Stiahnutie dát.

7.4.3 Popis

Stiahne RDF dáta z externého SPARQL koncového bodu podľa predpisu definovaného v konfigurácii a výsledok vo forme RDF zašle na výstup `output`.

Dopyt je možné prepísať tak, aby sa koncový bod dopytoval po dávkach. V tomto režime sa ignoruje vonkajší ORDER BY a LIMIT.

7.4.4 Konfiguračné parametre

Meno	Popis
URL koncového bodu	URL SPARQL koncového bodu, z ktorého sa majú extrahovať dáta
SPARQL	SPARQL construct dotaz, ktorým sa dáta zo SPARQL koncového bodu vyextrahujú

Meno	Popis
Construct	
Veľkosť dávky	Na zamedzenie vypršania časového limit alebo obídienie limitu na počet výsledkov koncového bodu, je možné prepísať dopy tak, že dopytovanie prebieha po menších dávkach. Toto nastavenie umožňuje nastavenie veľkosti jednej dávky

7.4.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
output	výstup	RdfDataUnit	Extrahované dáta vo forme RDF	áno

7.5 Uloženie súborov do Virtuoso triplestore

7.5.1 Identifikátor

l-filesToVirtuoso

7.5.2 Typ

Uloženie dát.

7.5.3 Popis

Nahrá RDF dáta do Virtuoso.

Vstup tohto kroku (RDF dáta) sú súbory, ktoré sa musia nachádzať v nahrávanom priečinku na lokálnom disku.

Všetky súbory musia byť vo

formátoch `application/rdf+xml` alebo `text/turtle` (prípony `.rdf`, resp. `.ttl`).

Užívateľ musí zabezpečiť spustenie tohto kroku až potom, čo je priečinok naplnený súbormi.

To je možné dosiahnuť zaradením `l-filesUpload` s „spustiť“ po hranou na toto DPU.

Príklad transformácie:

```

-----
|                                     |                                     |                                     |
|                                     |                                     |                                     |
| e-sparqlEndpoint | ---> | t-rdfToFiles | ---> | l-filesUpload | --spustiť po
hrana---> | l-FilesToVirtuoso |
|                                     |                                     |                                     |
|                                     |                                     |                                     |
-----
-----

```


7.5.4 Konfiguračné parametre

Meno	Popis
Virtuoso JDBC URL	URL pre uskutočnenie JDBC spojenia so serverom Virtuoso
Používateľské meno	Používateľské meno na serveri Virtuoso
Heslo	Heslo príslušné užívateľskému menu
Vyčistenie cieľového grafu pred nahrávaním (checkbox)	Intuitívne
Cesta k nahrávanému priečinku	Cesta k nahrávanému priečinku
Vrátane podadresárov (checkbox)	Ak je checkbox zaškrtnutý, na Virtuoso sa nahrá aj obsah podadresárov
Vzor mena súboru	Vzor pre názvy súborov, ktoré budú nahrávané podľa normy SQL (reťazec znakov nahrádza symbol „%“). Napr. „%.ttl“
Cieľový graf	URI cieľového grafu
Interval aktualizácie stavu (s)	Časový úsek medzi dvoma aktualizáciami stavov (v sekundách)
Počet vlákien	
Vynechať súbor pri chybe (checkbox)	V prípade výskytu chyby pokračuje proces s ďalším súborom (ak je checkbox aktívny)

7.5.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
config	vstup	RdfDataUnit	Dynamická RDF konfigurácia, pozri Pokročilá konfigurácia	
rdfOutput	výstup	RDFDataUnit	Metadata o dátach na úložisku Virtuoso	

7.5.6 Pokročilá konfigurácia

Krok je možné nakonfigurovať aj dynamicky cez vstup `config` pomocou RDF dát.

Vzor konfigurácie:

```
`turtle <http://localhost/resource/config <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://unifiedviews.eu/ontology/dpu/filesToVirtuoso/Config>;  
<http://unifiedviews.eu/ontology/dpu/filesToVirtuoso/config/fileName>  
"dataset.trig";
```

```
<http://unifiedviews.eu/ontology/dpu/filesToVirtuoso/config/graphUri>  
"http://dataset".
```

7.6 Uloženie dát do databázy

7.6.1 Identifikátor

l-relationalToSql

7.6.2 Typ

Uloženie dát.

7.6.3 Popis

Nahrá vstup z internej databázovej tabuľky do vzdialenej SQL databázy (v súčasnosti je podporovaná PostgreSQL).

7.6.4 Konfiguračné parametre

Meno	Popis
URL databázy	JDBC URL zdrojovej databázy (v súčasnosti je podporovaná iba PostgreSQL)
Používateľské meno	Login používateľa databázy
Heslo	Heslo k databáze
Pripojenie cez SSL	Použitie bezpečného pripojenia k databáze
Predpona cieľovej tabuľky	Názov tabuľky použitej na interné uchovanie extrahovaných dát
Vymazať tabuľku pred vložením	Vyprázdni tabuľku pred vkladáním
Znovu vytvoriť tabuľku	Odstráni tabuľku v prípade, ak existuje

7.6.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
inTablesData	vstup	RelationalDataUnit	Dáta z databázovej tabuľky	áno

7.7 Filter súborov

7.7.1 Identifikátor

t-filesFilter

7.7.2 Typ

Transformácia dát.

7.7.3 Popis

Filtruje súbory na základe filtrovacieho vzoru definovaného v konfigurácii. Podporované sú aj regulárne výrazy.

7.7.4 Konfiguračné parametre

Meno	Popis
Použitý filter	Existujú dva filtre na vykonávanie filtrovania: - symbolické meno - virtuálna cesta
Aktuálna šablóna	Filtrovací vzor, napr. „*.csv“
Použiť regulárny výraz (checkbox)	Ak je checkbox aktívny, na vstupe sú akceptované regulárne výrazy

7.7.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	FilesDataUnit	Zoznam súborov na filtrovanie	áno
output	výstup	FilesDataUnit	Zoznam súborov vyhovujúcich filtru	áno

7.8 Nahradenie vzoru v súbore

7.8.1 Identifikátor

t-filesFindAndReplace

7.8.2 Typ

Transformácia dát.

7.8.3 Popis

Nájde a nahradí reťazce (vzory) v súboroch.

7.8.4 Konfiguračné parametre

Meno	Popis
Hľadany vzor	Pôvodný reťazec, ktorý má byť nahradený
Náhrada	Reťazec, ktorý nahradí pôvodný reťazec
Vynechaj súbor pri chybe (checkbox)	Intuitívne

7.8.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
filesInput	vstup	FilesDataUnit	Vstupné súbory, ktoré je potrebné upraviť	áno
filesOutput	výstup	FilesDataUnit	Upravené súbory s nahradenými reťazcami	áno

7.9 Zjednotenie súborov

7.9.1 Identifikátor

t-filesMerger

7.9.2 Typ

Transformácia dát.

7.9.3 Popis

Zjednotí dátové jednotky so súbormi do jednej.

7.9.4 Konfiguračné parametre

Meno	Popis
N/A	

7.9.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
filesInput	vstup	FilesDataUnit	Dátová jednotka so súbormi	áno
filesOutput	výstup	FilesDataUnit	Dátová jednotka obsahujúca súbory zo všetkých vstupných dátových jednotiek	

7.10 Premenovanie súborov

7.10.1 Identifikátor

t-filesRenamer

7.10.2 Typ

Transformácia dát.

7.10.3 Popis

Premenuje súbory.

7.10.4 Konfiguračné parametre

Meno	Popis
Vzor	Regulárny výraz použitý na získanie reťazca, ktorý má byť nahradený v názve súboru. Táto hodnota je použitá ako časť náhrady (druhý argument) v SPARQL REPLACE
Náhrada	Hodnota, ktorou sa bude nahrádzať, môže odkazovať na skupiny vyhovujúce parametru Vzor. Táto hodnota sa používa ako časť náhrady (tretí argument) v SPARQL REPLACE
Pokročilý mód (checkbox)	Ak je checkbox aktívny, používateľom zadany text sa považuje za výraz nie za reťazec

7.10.4.1 Zoznam užitočných príkazov pri neaktívnom pokročilom móde:

Akcia	Vzor	Nahradený výraz
Pridanie prípony „.gml”	^(.+)\$	\$1.gml
Premenovať súbor na „abc”	^.+\$	abc

7.10.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
inFilesData	vstup	FilesDataUnit	Súbory, ktoré majú byť premenované	áno
outFilesData	výstup	FilesDataUnit	Premenované súbory	áno

7.11 Načítanie RDF zo súboru

7.11.1 Identifikátor

t-filesToRdf

7.11.2 Typ

Transformácia dát.

7.11.3 Popis

Získa RDF dáta zo súborov ľubovoľného RDF formátu. Výstupná dátová jednotka RDF obsahuje jeden záznam pre každý vstupný súbor (každý vstupný súbor vyprodukuje jeden graf RDF dát).

Pri predvolenom nastavení je RDF formát detekovaný automaticky na základe prípony vstupného súboru. V prípade, že detekcia bude nesprávna, špecifikujte, prosím, RDF formát vstupných súborov manuálne.

7.11.4 Konfiguračné parametre

Meno	Popis
RDF formát vstupných súborov	RDF formát dát vo vstupných súboroch. AUTO = automatická detekcia RDF formátu vstupných súborov (predvolené)
Veľkosť transakcie	0 znamená, že na spracovanie každého súboru sa použije separátna transakcia. 1 znamená automatický commit každej databázovej operácie (autocommit). 2 a viac znamená, že v transakcii sa spracuje daný počet RDF trojíc
Čo robiť, keď zlyhá transformácia jedného súboru	Zastaviť vykonávanie alebo pokračovať a preskočiť na ďalší súbor
Ako vyberať symbolické meno výstupu	Použiť symbolické mená vstupných súborov alebo použiť jedno pevné symbolické meno

7.11.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
filesInput	vstup	FilesDataUnit	Vstupné súbory s dátami	áno
rdfOutput	výstup	RDFDataUnit	Získané RDF dáta	áno

7.12 Validácia XML

7.12.1 Identifikátor

t-filterValidXml

7.12.2 Typ

Transformácia dát.

7.12.3 Popis

Validuje XML vstupy tromi spôsobmi:

- * skontroluje, či je XML správne formátované
- * skontroluje, či vyhovuje zadanej XSD schéme
- * validuje prostredníctvom špecifikovanej XSLT šablóny

7.12.4 Konfiguračné parametre

Meno	Popis
XSD schéma	XSD schéma, proti ktorej sa vstupné XML súbory validujú
XSLT transformácia	Prázdny výstup XSLT transformácie znamená bezchybnú validáciu. Hociaký neprázdny výstup znamená chybu pri validácii, pričom výstup zároveň obsahuje informáciu o konkrétnej chybe, ktorá ho spôsobila
Zlyhanie procesu pri prvej validačnej chybe	Pri prvom výskyte chyby validácie DPU zastaví vykonávanie procesu

7.12.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	FilesDataUnit	Zoznam súborov určených na validáciu	áno
outputValid	výstup	FilesDataUnit	Zoznam súborov, ktoré vyhovujú validačným kritériám	
outputInvalid	výstup	FilesDataUnit	Zoznam súborov nevyhovujúcich validačným kritériám	

7.13 Zlúčenie RDF grafov

7.13.1 Identifikátor

t-rdfMerger

7.13.2 Typ

Transformácia dát.

7.13.3 Popis

Zlúči RDF dáta.

7.13.4 Konfiguračné parametre

Meno	Popis
N/A	

7.13.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
rdflInput	vstup	RDFDataUnit	Dátová jednotka, kam používateľ zapojí všetky vstupy, ktoré majú byť zlúčené	áno
rdflOutput	výstup	RDFDataUnit	Dátová jednotka, kam sa pošlú zlúčené grafy	

7.14 Uloženie RDF do súboru

7.14.1 Identifikátor

t-rdfToFiles

7.14.2 Typ

Transformácia dát.

7.14.3 Popis

Transformuje RDF grafy do súborov.

7.14.4 Konfiguračné parametre

Meno	Popis
Formát súboru pre RDF dáta	Formát súboru na výstupe: <ul style="list-style-type: none">- Turtle- RDF/XML- N-Triples- N3- RDFa

Meno	Popis
Vygenerovať súbor .graph s názvom výstupného grafu	Je potrebný grafový súbor?
Názov výstupného grafu	Názov výstupného grafu (ak je generovaný grafový súbor)
Názov výstupného súboru (bez prípony, príponu určuje výstupný formát)	zrejme

7.14.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	RDFDataUnit	RDF grafy	áno
output	výstup	FilesDataUnit	Súbory obsahujúce RDF dáta	áno
config	vstup	RdfDataUnit	Dynamická RDF konfigurácia, pozri Pokročilá konfigurácia	

7.14.6 Pokročilá konfigurácia

Krok je možné nakonfigurovať aj dynamicky cez vstup `config` pomocou RDF dát.

Vzor konfigurácie:

```
`turtle <http://localhost/resources/configuration>
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://unifiedviews.eu/ontology/dpu/rdfToFiles/Config>;
<http://unifiedviews.eu/ontology/dpu/rdfToFiles/config/fileFormat> "TriG";
<http://unifiedviews.eu/ontology/dpu/rdfToFiles/config/outputUri>
"http://output-graph/name";
<http://unifiedviews.eu/ontology/dpu/rdfToFiles/config/outputFile> "graph-
output-file".
```

7.15 Validácia RDF dát

7.15.1 Identifikátor

t-rdfValidator

7.15.2 Typ

Transformácia dát..

7.15.3 Popis

Zvaliduje RDF dáta.

7.15.4 Konfiguračné parametre

Meno	Popis
Validačný dopyt	ASK alebo SELECT SPARQL dopyt ASK má byť naformulovaný tak, že vráti TRUE práve vtedy, keď sú dáta NEvalidné SELECT má byť naformulovaný tak, že vráti neprázdny zoznam práve vtedy, ak sú dáta NEvalidné.
Zastav transformáciu, ak validácia nie je úspešná	Pri zaškrtnutí sa proces zastaví, ak je validácia neúspešná Ak nezaškrtnuté, výsledok pri zlyhaní sa zapíše do logu a proces pokračuje

7.15.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
rdflInput	vstup	RDFDataUnit	Vstupné RDF dáta na zvalidovanie	áno
rdfOutput	výstup	RDFDataUnit	Kópia rdflInput dát	

7.16 SQL transformácia

7.16.1 Identifikátor

t-relational

7.16.2 Typ

Transformácia dát.

7.16.3 Popis

Transformuje N vstupných tabuliek do jednej výstupnej tabuľky pomocou SQL SELECT dotazu.

7.16.4 Konfiguračné parametre

Meno	Popis
SQL SELECT dotaz	SQL SELECT dotaz na získanie/transformáciu dát zo vstupných tabuliek
Meno cieľovej tabuľky	Meno tabuľky, kam sa uloží výstup SQL dotazu
Stĺpce primárneho kľúča	Čiarkou oddelený zoznam stĺpcov, ktoré tvoria primárny kľúč výslednej tabuľky
Indexované stĺpce	Čiarkou oddelený zoznam stĺpcov, ktoré budú indexované vo výslednej tabuľke

7.16.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
inputTables	vstup	RelationalDataUnit	Vstupné tabuľky	áno
outputTable	výstup	RelationalDataUnit	Výstupná tabuľka	áno

7.17 Konverzia tabuľky do RDF

7.17.1 Identifikátor

t-relationalToRdf

7.17.2 Typ

Transformácia dát.

7.17.3 Popis

Konvertuje tabuľkové dáta (SQL databázové tabuľky) na RDF dáta.

7.17.4 Konfiguračné parametre

Meno	Popis
Základ URI zdroja	Táto hodnota sa použije pri automatickom generovaní vlastností stĺpcov a tiež na vytvorenie absolútnej URI, ak je v stĺpci „URI vlastností“ poskytnutá relatívna URI
Stĺpec kľúča	Názov stĺpca, ktorý bude pridaný k „Základ URI vlastností“ a použitý ako predmet pre riadky
Trieda pre entitu riadku	Táto hodnota sa použije ako trieda pre všetky entity riadkov
Mapovať všetky stĺpce	Východzie mapovanie je automaticky vygenerované pre každý stĺpec
Generovať stĺpec pre riadok	Stĺpec s číslom riadku je generovaný pre každý riadok
Generovať triedu pre tabuľku/riadok	Pre výrazy tabuľkových entít s typom sú generované triedy
Generovať predmet pre tabuľku	Vytvorí sa predmet pre každú tabuľku, ktorý odkazuje na všetky riadky v danej tabuľke
Generovať označenia	Označenia (rdfs:labels) sú generované do URI stĺpcov
Rozšírený stĺpec kľúča	„Stĺpec kľúča“ je interpretovaný ako šablóna. Experimentálna funkcionálna!

Meno	Popis
Automatický typ ako reťazec	Všetky automatické typy sú považované za reťazce
Jednoduché mapovanie	Vlastné mapovanie stĺpcov na základe názvu stĺpca v zdrojovej tabuľke

7.17.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
tablesInput	vstup	RelationalDataUnit	Vstupné databázové tabuľky	áno
rdfOutput	výstup	RDFDataUnit	RDF dáta na výstupe	áno

7.18 RDF výber pomocou SPARQL Construct

7.18.1 Identifikátor

t-sparqlConstruct

7.18.2 Typ

Transformácia dát.

7.18.3 Popis

Transformuje vstupné dáta pomocou konštrukčného SPARQL dotazu.

7.18.4 Konfiguračné parametre

Meno	Popis
Púšťanie dotazov po grafoch	Ak zaškrtnuté, dotazy sú púšťané po grafoch
SPARQL konštrukčný dotaz	Textové pole určené pre SPARQL konštrukčný dotaz

7.18.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	RDFDataUnit	RDF vstupné dáta	áno
output	výstup	RDFDataUnit	RDF výstupné (transformované) dáta	áno

7.19 RDF výber pomocou SPARQL Select

7.19.1 Identifikátor

t-sparqlSelect

7.19.2 Typ

Transformácia dát

7.19.3 Popis

Transformuje vstupné dáta pomocou výberového SPARQL dotazu a výsledok uloží do CSV súboru. Nerobí validáciu dát.

7.19.4 Konfiguračné parametre

Meno	Popis
Cieľová cesta	Cesta a názov cieľového CSV súboru
SPARQL výberový dotaz	Textové pole určené pre SPARQL výberový dotaz

7.19.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	RDFDataUnit	RDF vstupné dáta	áno
output	výstup	FilesDataUnit	CSV súbor obsahujúci výsledok z SPARQL SELECT dotazu	áno

7.20 RDF úprava pomocou SPARQL update

7.20.1 Identifikátor

t-sparqlUpdate

7.20.2 Typ

Transformácia dát.

7.20.3 Popis

Transformuje vstupné dáta pomocou aktualizáčného SPARQL dotazu.

7.20.4 Konfiguračné parametre

Meno	Popis
------	-------

Meno	Popis
Púšťanie dotazov po grafoch	Ak zaškrtnuté, dotazy sú púšťané po grafoch
SPARQL aktualizací dotaz	Textové pole určené pre SPARQL aktualizací dotaz

7.20.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	RDFDataUnit	RDF vstupné dáta	áno
output	výstup	RDFDataUnit	RDF výstupné (transformované) dáta	áno

7.21 Načítanie tabuľkových dát zo súboru

7.21.1 Identifikátor

t-tabularToRelational

7.21.2 Typ

Transformácia dát.

7.21.3 Popis

Načíta súbor s tabuľovými dátami vo formáte CSV, XLS, XLSX alebo DBF a výsledok dá na výstup vo forme relačných dát.

7.21.4 Konfiguračné parametre

Meno	Popis
Názov tabuľky	Meno tabuľky, do ktorej budú uložené načítané dáta
Typ súboru	Typ vstupného súboru (XLS/X, CSV, DBF)
Dáta sa začínajú na riadku	Definuje, na ktorom riadku sa začínajú dáta
Znaková súprava	Znaková súprava vstupného súboru
Znak úvodzoviek	Znak na ohraničenie jednotlivých polí používaný pri parsovaní CSV súboru, napr. "

Meno	Popis
Oddelovač	Oddelovač polí používaný pri parsovaní CSV súboru, napr. ,
Mapovacia tabuľka	Uvedie sa definícia tabuľky, do ktorej sa majú uložiť tabuľkové dáta načítané zo súboru. Je potrebné uviesť želané názvy stĺpcov. Ich názvy nemusia byť zhodné s tými zo súboru. Stĺpce zo súboru sa premapujú do tabuľky automaticky podľa poradia (prvý stĺpec zo súboru na prvý stĺpec tabuľky, druhý na druhý atď). Dátový typ každého stĺpca je text (bez obmedzenia veľkosti). Voľba <i>Kompozitný kľúč?</i> znamená, že daný stĺpec sa stane súčasťou primárneho kľúča vytvárajanej tabuľky. Vhodná voľba primárneho kľúču tabuľky môže urýchliť nasledujúce spracovanie tabuľky v procese

7.21.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	FilesDataUnit	Zoznam súborov, z ktorých majú byť načítané tabuľkové dáta	áno
output	výstup	RelationalDataUnit	Výstupná tabuľka/tabuľky	áno

7.22 Rozbalenie súborov

7.22.1 Identifikátor

t-unzipper

7.22.2 Typ

Transformácia dát.

7.22.3 Popis

Rozbalí vstupný súbor komprimovaný metódou ZIP na jednotlivé súbory.

7.22.4 Konfiguračné parametre

Názov	Popis
Prevenca duplicitných názvov	Ak je zaškrtnuté, DPU zabráni kolízii názvov súborov na výstupe v prípade, že je rozbalených viac súborov s rovnakou štruktúrou. Každý súbor tak bude identifikovaný nielen názvom a relatívnou cestou archivovaných súborov, ale aj identifikátorom samotného zip archívu. Ak nie je zaškrtnutý, rozbalené súbory sú identifikované iba ich menom a relatívnymi cestami vnútri archívu, čo môže spôsobiť kolízie názvov, ak je rozbalených viac ako jeden archív obsahujúcich súbory s rovnakým názvom. Ak si nie ste istí, ponechajte checkbox zaškrtnutý.

7.22.4.1 Príklady

Vstupný súbor 1: tounzipA.zip

text.txt

Vstupný súbor 2: tounzipB.zip

text.txt

Výsledok pri zaškrtnutom checkboxe:

tounzipA.zip/text.txt tounzipB.zip/text.txt

Výsledok pri nezaškrtnutom checkboxe:

text.txt text.txt

Ak checkbox nie je zaškrtnutý, tento výstup spôsobí kolíziu a zlyhanie procesu.

7.22.5 Vstupy a výstupy

Názov	Typ	Dátová jednotka	Popis	Povinné
input	vstup	FilesDataUnit	Súbory na rozbalenie	áno
output	výstup	FilesDataUnit	Rozbalené súbory	áno

7.23 XSTL transformácia

7.23.1 Identifikátor

t-xslt

7.23.2 Typ

Transformácia dát.

7.23.3 Popis

Vykoná XSL transformáciu vstupných súborov na `files` vstupe s využitím statického predpisu `xslt`.

Transformované súbory dá na výstup `files`.

V XSLT je podporované generovanie náhodných UUID pomocou `randomUUID()` funkcie v mennom priestore `uuid-functions`.

Príklad použitia:

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet
  xmlns:uuid="uuid-functions"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  version="2.0">
  <xsl:template match="/">
    <xsl:value-of select="uuid:randomUUID()" />
  </xsl:template>
</xsl:stylesheet>
```


7.23.4 Konfiguračné parametre

Name	Description
Preskoč súbor, ak sa vyskytne chyba	Pri zaškrtnutej voľbe sa v prípade chyby pri transformácii súboru daná transformácia skončí a pokračuje sa ďalším súborom. Inak sa transformácia zastaví
Prípona výstupného súboru	Pridá uvedenú príponu k menu vstupného súboru pre výstupný súbor
Počet vlákien navyše	Počet vlákien použitých pri transformácii. Ak sa nastaví 0, spustí sa jedno vlákno rovnako ako pri 1. Toto nastavenie je užitočné pri transformáciách, ktoré trvajú dlhší čas. Viac vlákien však znamená nielen zvýšenie rýchlosti, ale aj zvýšenie spotreby pamäte, takže treba zvážiť vhodné nastavenie
XSLT šablóna	XSLT šablóna použitá na transformáciu

7.23.5 Vstupy a výstupy

Name	Type	DataUnit	Description	Mandatory
files	vstup	FilesDataUnit	Súbory určené na transformáciu	áno
files	výstup	FilesDataUnit	Transformované súbory	áno
config	vstup	RdfDataUnit	Dynamická RDF konfigurácia, pozri Pokročilá konfigurácia	

7.23.6 Pokročilá konfigurácia

Krok je možné nakonfigurovať aj dynamicky cez vstup `config` pomocou RDF dát.

Vzor konfigurácie:

```
<http://localhost/resource/config>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://linked.opendata.cz/ontology/uv/dpu/xslt/Config>;
  <http://linked.opendata.cz/ontology/uv/dpu/xslt/fileInfo>
<http://localhost/resource/fileInfo/0>;
  <http://linked.opendata.cz/ontology/uv/dpu/xslt/outputFileExtension> ".ttl".
<http://localhost/resource/fileInfo/0>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://linked.opendata.cz/ontology/uv/dpu/xslt/FileInfo>;
  <http://linked.opendata.cz/ontology/uv/dpu/xslt/param>
<http://localhost/resource/param/0>;
  <http://unifiedviews.eu/DataUnit/MetadataDataUnit/symbolicName>
"smlouva.ttl".
<http://localhost/resource/param/0>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://linked.opendata.cz/ontology/uv/dpu/xslt/Param>;
  <http://linked.opendata.cz/ontology/uv/dpu/xslt/param/name> "paramName";
  <http://linked.opendata.cz/ontology/uv/dpu/xslt/param/value> "paramValue".
```

7.24 Zbalenie súborov

7.24.1 Identifikátor

t-zipper

7.24.2 Typ

Transformácia dát.

7.24.3 Popis

Skomprimuje súbory na vstupe `input` do zip súboru podľa nakonfigurovaného mena a tento dá na výstup `output`.

7.24.4 Konfiguračné parametre

Meno	Popis
Zip súbor cesta/názov (s príponou)	Špecifikuje cestu s názvom výstupného zip súboru. Cesta s názvom musí byť relatívna, napr. /subor1.zip, /adresar1/subor2.zip. Absolútna cesta ako napr. C:/ nemôže byť použitá. V prípade systémov UNIX /adresar1/subor2.zip je braná ako relatívna cesta

7.24.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	FilesDataUnit	Zoznam súborov na zbalenie do zipu	áno
output	výstup	FilesDataUnit	Zip súbor	áno

7.25 Stiahnutie súboru z katalógu

7.25.1 Identifikátor

e-filesFromCKAN

7.25.2 Typ

Stiahnutie dát.

7.25.3 Popis

Stiahne súbor zo zdrojov CKAN.

7.25.4 Konfiguračné parametre

Meno	Popis
------	-------

Meno	Popis
Iba vlastné datasety	Ak je zaškrtnuté, budú zobrazené len (verejné aj privátne) datasety mojej organizácie. Inak sa zobrazia aj všetky verejné datasety iných organizácií
Rozbaliť/Zbaliť všetko	Tlačidlá, ktoré spôsobujú rozbalenie/zbalenie stromu
Vyhľadajte dataset/súbor	Filtrovanie uzlov stromu podľa názvu datasetu alebo súboru
Dátový zdroj na stiahnutie	Strom obsahuje uzly: Názov organizácie (1. level), Názov datasetu (2. level), Názov zdroja (3. level)
Premenovať stiahnutý súbor na	Názov súboru, na ktorý má byť premenovaný stiahnutý súbor. Ak nie je zadany, použije sa ID zdroja CKAN

7.25.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
output	výstup	FilesDataUnit	Stiahnutý súbor zo zdrojov CKAN	

7.26 Stiahnutie tabuľkových dát z katalógu

7.26.1 Identifikátor

e-relationalFromCKAN

7.26.2 Typ

Stiahnutie dát.

7.26.3 Popis

Stiahne relačné dáta zo zdroja CKAN.

7.26.4 Konfiguračné parametre

Meno	Popis
Iba vlastné datasety	Ak je zaškrtnuté, budú zobrazené len (verejné aj privátne) datasety mojej organizácie. Inak sa zobrazia aj všetky verejné datasety iných organizácií
Ignoruj typy databázových stĺpcov (použi reťazce)	Ak je zaškrtnuté, ako typy databázových stĺpcov sa vo výslednej tabuľke použijú reťazce
Rozbaliť/Zbaliť všetko	Tlačidlá, ktoré spôsobujú rozbalenie/zbalenie stromu

Meno	Popis
Vyhľadajte dataset/relačný zdroj	Filtrovanie uzlov stromu podľa názvu datasetu alebo relačného zdroja
Relačný zdroj na stiahnutie	Strom obsahujúci uzly: Názov organizácie (1. level), Názov datasetu (2. level), Názov zdroja (3. level)
Premenovať názov tabuľky na	Názov tabuľky, na ktorý má byť premenovaný relačný zdroj z CKAN. Ak nie je zadáný, použije sa ID zdroja CKAN

7.26.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
output	výstup	FilesDataUnit	Stiahnuté relačné dáta zo zdrojov CKAN	

7.27 Uloženie súboru do katalógu

7.27.1 Identifikátor

l-filesToCKAN

7.27.2 Typ

Uloženie dát.

7.27.3 Popis

Nahrá súbory do zadanej inštancie CKAN.

7.27.4 Konfiguračné parametre

Meno	Popis
Názov zdroja CKAN	Názov zdroja vytvoreného v CKAN, má prednosť pred vstupom z e-distributionMetadata, a aj v prípade, ak nie je zadáný, použije virtuálnu cestu alebo symbolické meno ako názov zdroja
Použije názov súboru ako CKAN meno	Ak je checkbox aktívny, názov súboru je použitý ako meno zdroja v CKAN. Musí byť aktívny v prípade viacerých súborov na vstupe, inak DPU zlyhá
Prepísať existujúce zdroje	Ak je checkbox aktívny, existujúce zdroje sú prepísané

7.27.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
filesInput	vstup	FilesDataUnit	Súbor/súbory, ktoré sa majú nahráť do CKAN-u	áno
distributionInput	vstup	RDFDataUnit	Distribution metadata produced by e-distributionMetadata	

7.28 Uloženie RDF do katalógu

7.28.1 Identifikátor

l-rdfToVirtuosoAndCkan

7.28.2 Typ

Uloženie dát.

7.28.3 Popis

Nahrá RDF dáta na Virtuoso použitím L-RdfToVirtuoso a vytvorí CKAN zdroje použitím L-RdfToCkan.

7.28.4 Konfiguračné parametre

Meno	Popis
Vyčistiť cieľový graf pred nahratím (checkbox)	Ak je checkbox aktívny, pred nahratím nového RDF vyčistí všetky predtým nahrané dáta (do príslušného grafu). Inak budú pôvodné a nové RDF dáta zjednotené ako prienik RDF trojíc (RDF trojice nemôžu byť duplicitné)

7.28.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
rdfInput	vstup	RDFDataUnit	RDF data nahrané na Virtuoso a zadané inštancie CKAN	áno
distributionInput	vstup	RDFDataUnit	Distribučné metadáta vytvorené z e-distributionMetadata	

7.29 Aktualizácia tabuľky v katalógu

7.29.1 Identifikátor

l-relationalDiffToCKAN

7.29.2 Typ

Uloženie dát.

7.29.3 Popis

Aktualizuje dáta v úložisku CKAN z relačných dát na vstupe.

7.29.4 Konfiguračné parametre

Meno	Popis
Názov zdroja CKAN	Názov zdroja CKAN, ktorý má byť vytvorený, má prednosť pred vstupom z e-distributionMetadata a ak nie je zadáný, použije sa virtuálna cesta alebo symbolické meno ako meno zdroja

7.29.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
tablesInput	vstup	RelationalDataUnit	Tabuľky, ktoré sa majú aktualizovať v úložisku CKAN	áno
distributionInput	vstup	RDFDataUnit	Distribučné metadáta vytvorené z e-distributionMetadata	

7.30 Uloženie tabuľky do katalógu

7.30.1 Identifikátor

l-relationalToCKAN

7.30.2 Typ

Uloženie dát.

7.30.3 Popis

Nahrá tabuľky internej relačnej databázy na úložisko CKAN.

7.30.4 Konfiguračné parametre

Meno	Popis
Názov zdroja CKAN	Názov zdroja CKAN, ktorý má byť vytvorený, má prednosť pred vstupom z e-distributionMetadata a ak nie je zadáný, použije sa virtuálna cesta alebo symbolické meno ako meno zdroja
Prepísať existujúce	Ak zdroj už existuje, bude prepísaný

Meno	Popis
zdroje	

7.30.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
tablesInput	vstup	RelationalDataUnit	Tabuľky, ktoré sa majú uložiť na úložisko CKAN	áno
distributionInput	vstup	RDFDataUnit	Distribučné metadáta vytvorené z e-distributionMetadata	

7.31 Publikácia dát na elektronický podpis

7.31.1 Identifikátor

l-dataToXMLSignature

7.31.2 Typ

Uloženie dát.

7.31.3 Popis

Publikuje XML dokumenty na webservis, ktorý následne umožňuje získanie podpísaných XML dát.

Pri publikovaní musí byť poskytnutá správna URI XSD schémy pre podpisovaný XML dokument - XSD URI je jednoznačný identifikátor do podpisovacieho servisu.

Poskytnutý XPATH výraz musí vracať XML, ktoré je validné voči XSD schéme pre podpisovaný dokument.

Na vstupe DPU musí byť 1 súbor - dátový zdroj stiahnutý z CKAN-u a musí obsahovať ID, ktoré je jeho jedinečný identifikátor v CKAN-e aj v podpisovacom webservice.

Výsledok DPU je publikovaný XML v servise, z ktorého je možné získať podpísané XML dáta volaním jednoduchého REST servisu

* k registrovaným záznamom potom možno pristupovať cez metódu `get_signed_data`, ktorá má 2 parametre:

* `resource_id`: ID, pod ktorým bol registrovaný XML dokument

* `param` (nepovinné): Hodnota tohto parametru bude použitá v XPATH výraze pre „`param`” premennú

Príklad volania:

http://WEB_SERVICE_HOST:8080/xml/get_signed_data?resource_id=b6210126-2d7c-4dcb-bc1a-6998310d1471?param=12589

Príklad odpovede:

```
<PodpisOdpoved> <Podpis> <xzep:DataEnvelope>....</xzep:DataEnvelope> </Podpis>
</PodpisOdpoved>
```

7.31.4 Konfiguračné parametre

Meno	Popis
URI XSD schémy	URI XSD schémy pre podpisovaný dokument - jedinečný identifikátor pre podpisovací WS
XPATH výraz	XPATH výraz pre prístup k elementom XML súboru na základe vstupného parametra
	XPATH výraz môže obsahovať iba jednu premennú, ktorá musí byť nazvaná „param“
Prepísať existujúci dátový zdroj	Ak dátový zdroj s daným ID už existuje a je zvolená táto možnosť, zdroj sa prepíše
	Ak zdroj už existuje a nie je zvolené, DPU zlyhá a zdroj nebude aktualizovaný

7.31.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
filesInput	vstup	FilesDataUnit	Publikované XML - dátový zdroj z CKAN-u	áno

7.32 Extrakcia metadát zo súboru

7.32.1 Identifikátor

t-filesMetadata

7.32.2 Typ

Transformácia dát.

7.32.3 Popis

Extrahuje/odvodzuje metadáta pre vstupné textové súbory s využitím rozhrania PoolParty komponentu.

Metadáta sú uložené a prenášané medzi krokmi v Resource objekte, ktorý sa používa na

publikovanie metadát do CKAN katalógu v krokoch
pre publikáciu dát do CKAN-u.

7.32.4 Konfiguračné parametre

Meno	Popis
Projekt	ID PoolParty projektu
Jazyk	Jazyk dokumentu

7.32.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
filesInput	vstup	FilesDataUnit	Vstupné dokumenty, pre ktoré sa vykoná extrakcia metadát	áno
filesOutput	výstup	FilesDataUnit	Spracované súbory s pripojenými metadátami	áno

7.33 Extrakcia textu zo súboru (OCR)

7.33.1 Identifikátor

t-ocr

7.33.2 Typ

Transformácia dát.

7.33.3 Popis

Vykonáva optické rozoznávanie znakov (OCR) na vstupných dokumentoch a konvertuje ich tak na textové súbory jedného z podporovaných formátov použitím ABBYY FREngine.

Aktuálne podporované vstupné formáty:

- * BMP
- * DCX
- * GIF
- * JBIG2
- * JPEG
- * JPEG 2000
- * PCX
- * PDF
- * PNG
- * TIFF

7.33.4 Konfiguračné parametre

Meno	Popis
Výber OCR profilu	Výber OCR profilu na základe typu vstupných dát a požadovaného výsledku
Formát výstupných súborov	Formát výsledných textových dokumentov (e.g. RTF, TXT, HTML, etc.)
Modifikátor názvu	Prípona pridaná k názvom výstupných súborov, pred príponou typu súboru (súbor. prípona -> súbor{modifikátor_názvu}.prípona_cieľového_formátu)
Správanie pri chybe	Stratégia ošetrovania chýb pri OCR procese

7.33.5 Vstupy a výstupy

Meno	Typ	Dátová hrana	Popis	Povinné
input	vstup	FilesDataUnit	Vstupné súbory pre OCR proces	áno
output	výstup	FilesDataUnit	Výsledné textové dokumenty	áno

8 Súvisiace dokumenty

Referenčný kľúč	Názov a popis dokumentu
[Katalog]	DOV-1-5-Prirucka_pouzivatela_MOD_Katalog_otvorených_dat-v1_0.docx Príručka používateľa pre webovú aplikáciu Katalóg otvorených dát