

Otevřená data ČSSZ:

**Přehledné informace dostupné všem,
snadno a zdarma**

ISSS
Hradec Králové, 5. 4. 2016

Ing. Jiří Šunka
Ing. Michaela Hendrychová



Obsah

1. Představení ČSSZ
2. Proces publikace otevřených dat ČSSZ
3. Ukázka řešení katalogu otevřených dat ČSSZ a jejich vizualizace
4. Shrnutí

Představení ČSSZ - Profil ČSSZ

- organizační složka státu v rezortu Ministerstva práce a sociálních věcí ČR
- působnost v oblasti sociálního zabezpečení (důchodového a nemocenského pojištění) a lékařské posudkové služby
- celkový objem příjmů a výdajů v částce 804 miliard Kč za rok 2015
- objem vybraného pojistného => 1/3 státního rozpočtu ČR
- 8,6 milionu klientů (počet obyvatel ČR - 10,5 milionu)
 - 2,9 milionu důchodců
- 8 690 zaměstnanců



Proces publikace otevřených dat ČSSZ

Hlavní kroky:

- Spolupráce ČSSZ s VŠE na Projektu TD020121 „Publikace dat statistických ročenek ve standardu otevřených dat“ 01.2014 – 06.2015
- Smluvní zajištění expertní podpory 05.2015
- Příprava a vydání organizační směrnice 05.2015 – 10.2015
- Výběr dat k publikaci 05.2015 – 07.2015
- Tvorba Publikačního plánu 06.2015 – 10.2015
- Návrh a příprava technického prostředí 05.2015 – 09.2015
- Vyhodnocení bezpečnostních hledisek 08.2015
- Publikace dat 2.11.2015
- Interaktivní vizualizace 22.12.2015

Smluvní zajištění expertní podpory v rámci „Zvýšení otevřenosti a efektivnosti dat ČSSZ“ pro oblasti:

- Metodiky
- Organizace přípravy a procesu publikace otevřených dat, vč. návrhu Publikačního plánu
- Publikace a katalogizace vybraných datových sad
- Technického řešení
- Interaktivních vizualizací

Proces publikace – organizační zajištění

Organizační zajištění:

- **Příprava organizační směrnice**
Nutnost organizačního ukotvení procesů v rámci činnosti organizace
- **Definice rolí**
Stanovení rolí, jejich kompetencí a odpovědností v procesu publikace otevřených dat
- **Zapojení věcných útvarů**
Aktivní spolupráce s odbornými útvary – vlastníky dat
- **Připomínky oddělení komunikace**
Spolupráce na tvorbě vzhledu internetových stránek a vizualizace dat

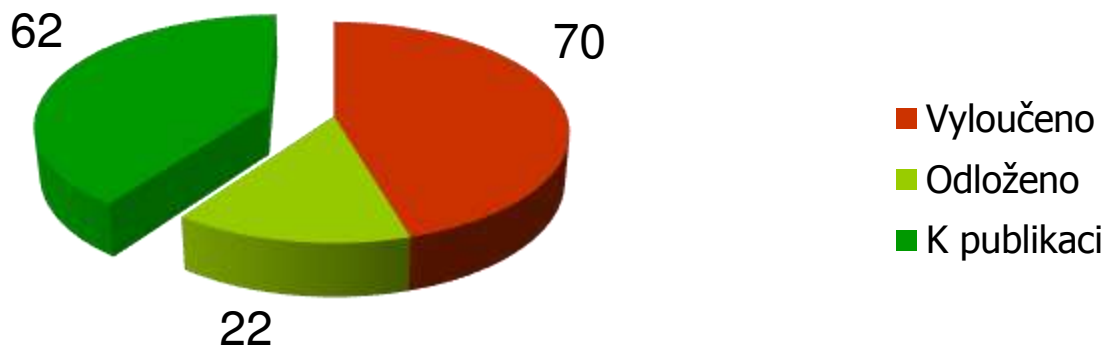
Proces publikace – výběr dat

Výběr dat k publikaci:

„Bezpečná“ data – již publikována na webu ČSSZ	:	154 datových sad
Po vyloučení nevhodných sad k publikaci	:	84 datových sad
Po posouzení přínosů a rizik a odložení k řešení	:	62 datových sad

Do plánu publikace zařazeno : 62 datových sad

Datové sady



Proces publikace - Vyhodnocení přínosů a rizik

Přínosy

- (P1) Posílení transparentnosti veřejné správy
- (P2) Podpora opětovného použití dat
- (P3) Zlepšení vnímání veřejné správy veřejností
- (P4) Zlepšení procesů a dat veřejné správy, zlepšení komunikace a spolupráce veřejné správy
- (P5) Snížení počtu dotazů dle zákona č. 106/1999 Sb.

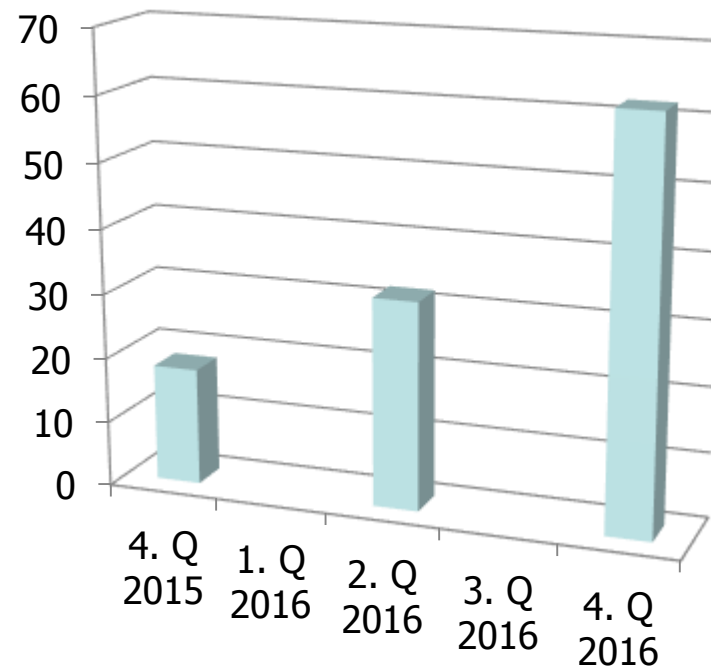
Rizika

- (R1) Zveřejnění dat v rozporu se zákonem
- (R2) Ohrožení bezpečnosti státu / majetku / osob
- (R3) Riziko porušení ochrany osobních údajů a dobrého jména
- (R4) Zveřejnění nevhodných dat či informací
- (R5) Překrývání dat

Proces publikace – publikace dat

Publikační plán:

- 2. 11. 2015 ⇒ 18 datových sad
 - statistické ročenky z oblasti důchodového pojištění
 - statistika informací poskytovaných podle zákona č. 106/1999 Sb.
- 04/2016 ⇒ 14 datových sad
 - oblast nemocenského pojištění
 - oblast LPS
 - oblast OSVČ
 - oblast správního řízení
- 12/2016 ⇒ 30 datových sad
 - různé okruhy dat



Proces publikace – technické prostředí

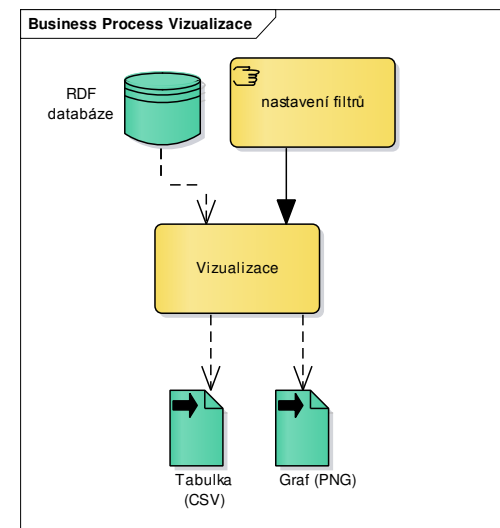
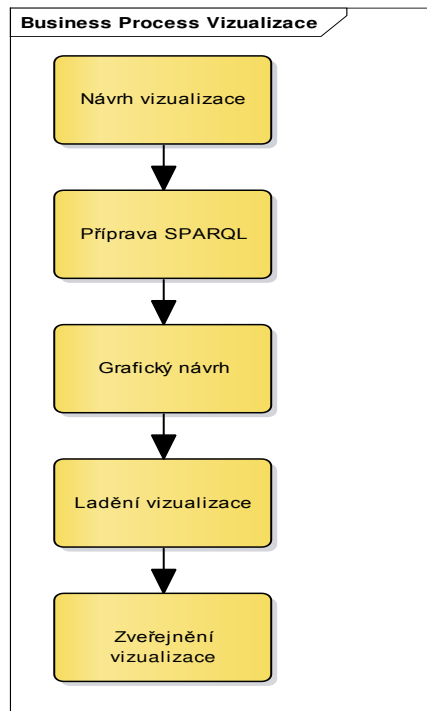
Technické prostředí:

- Výchozí návrh vycházel z nástrojů – UnifiedViews, CKAN, Openlink Virtuoso
- Modifikace podle standardů IIS ČSSZ
- Publikovaný formát – primárně propojená data v RDF, doplňkově v CSV
- Finální řešení
 - Dvě hlavní prostředí – testovací a produkční
 - Příprava dat – UnifiedViews, CKAN, Openlink Virtuoso
 - Publikace – LifeRay, databáze Apache Jena Fuseki, vlastní řešení katalogu dat, Google Charts pro vizualizaci
 - Umožněno automatické předávání na NKOD
- Další rozvoj
 - Automatizované přenosy mezi přípravou a portálem
 - Schvalovací workflow
 - Optimalizace procesů

Proces publikace – interaktivní vizualizace

Interaktivní vizualizace:

- Základní typy grafů – spojnicový, sloupcový, koláčový, mapy (kraj, okres)
- Vstupní data z databáze RDF
- Výběrové filtry na data
- Zobrazení grafu nebo tabulky dat
- Export dat grafu do CSV
- Export grafu do PNG
- Upřesňující komentáře
- Komponenta Google Charts



Otevřená data ČSSZ – <https://data.cssz.cz/>



ČESKÁ SPRÁVA SOCIÁLNÍHO ZABEZPEČENÍ Otevřená data

[Základní informace](#)[Katalog otevřených dat](#)[Podmínky užití](#)

Vítejte na portálu Otevřená data České správy sociálního zabezpečení

Portál je určen pro publikaci a vizualizaci otevřených dat, která vlastní a spravuje Česká správa sociálního zabezpečení (ČSSZ). Veškerá data publikovaná na tomto portále jsou primárně dostupná ve formátech **RDF(TriG)** a **CSV**. Data každé publikované datové sady jsou dostupná také pomocí příslušného SPARQL endpointu, který umožňuje získat data, odpovídající kritériím zadaným formou SPARQL dotazu.

Otevřená data jsou publikována jako datové sady. Seznamy aktuálně publikovaných sad a odkazy pro jejich stažení jsou dostupné v záložce **Katalog otevřených dat**.

Tento portál slouží také jako lokální katalog, který je registrován v **Národním katalogu otevřených dat (NKOD)**.

Před jejich použitím se prosím seznamte s podmínkami užití v záložce **Podmínky užití**.

Pokud se chcete o otevřených datech dozvědět více, navštivte webové stránky **Standardy publikace a katalogizace otevřených dat veřejné správy ČR**.

[Prohlášení o přístupnosti](#)[Technická pomoc](#)[Provozovatel ČSSZ](#)



Základní informace

Katalog otevřených dat

Podmínky užití

Katalogové záznamy datových sad

Filtrovat podle klíčového slova: [důchodová ročenka \(17\)](#) [důchodové pojištění \(17\)](#) [kraje \(1\)](#) [okresy \(1\)](#) [statistika \(21\)](#) [svobodný přístup k informacím \(1\)](#)
[zák. č. 106 1999 sb. \(1\)](#)

Titul	Vizualizace dat - grafy
Měsíční výše důchodů	Měsíční výše důchodů podle roku, kvantilu a druhu důchodu
Počet nově přiznaných důchodů v České republice dle měsíční výše důchodu	Počet nově přiznaných důchodů dle měsíční výše důchodu
Počet nově přiznaných důchodů v České republice dle osobního vyměřovacího základu	Počet nově přiznaných důchodů dle osobního vyměřovacího základu
Počet nově přiznaných důchodů v České republice dle věkové kategorie	Počet nově přiznaných důchodů dle věkové kategorie
Počet nově přiznaných invalidních důchodů v České republice	Počet nově přiznaných invalidních důchodů v České republice Počet nově přiznaných invalidních důchodů v České republice - koláčový graf
Počet vyplacených důchodů v České republice dle měsíční výše důchodu	Počet vyplacených důchodů podle měsíční výše důchodu
Počet vyplacených důchodů v České republice dle věkové kategorie	Počet vyplacených důchodů podle věkové kategorie
Počet vyplacených invalidních důchodů v České republice dle skupin diagnóz	Počet vyplacených invalidních důchodů v České republice
Počet zaniklých důchodů v České republice	Počet zaniklých důchodů podle druhu zániku
Průměrná délka pobírání starobního důchodu	Průměrná délka pobírání starobního důchodu

Nejčastější dotazy

Prohlášení o přístupnosti

Technická pomoc

Provozovatel ČSSZ

Otevřená data ČSSZ – <https://data.cssz.cz/>

← Výdaje na důchody

Výdaje na důchody podle roku a druhu penze

Periodičita aktualizace	Jednou ročně
URI souvisejícího geografického území	http://nian.linked.opendata.cz/resource/staty/1
Související geografické území	Česká republika
Klasifikace dle EuroVoc	důchodové pojištění podmínky pro odchod do důchodu
Jméno a e-mail kurátora dat	Ing. Jiří Šunka
Odkaz na dokumentaci datové sady	https://data.cssz.cz/documentation/vydaje-na-duchody-v-cr
Dotčené časové období od-do	01.01.2008 - 31.12.2014
Název a URL poskytovatele dat	Česká správa sociálního zabezpečení

Datové zdroje

Data	Schéma	Podmínky užití	Dotčené časové období
Data (RDF/TrIG) 36.2 kB	Schéma (RDF/TrIG) 17.3 kB	podmínky užití	01.01.2008 - 31.12.2014
Data (CSV) 3.6 kB	Schéma (JSON) 884 B	podmínky užití	01.01.2008 - 31.12.2014

Data v RDF jsou přístupná i přes SPARQL endpoint, který je na URL <https://data.cssz.cz/sparql>

Pro získání výběrů dat z datových sad je připraven jednoduchý **formulář**, který umožňuje zadávání dotazů ve SPARQL jazyku.

Metadata o datové sadě v RDF (TrIG) 6.8 kB

Vizualizace dat - grafy

Výdaje na důchody v ČR

[důchodová ročenka](#) [důchodové pojištění](#) [statistika](#)



ČESKÁ SPRÁVA SOCIÁLNÍHO ZABEZPEČENÍ
Otevřená data

Dokumentace datové sady Výdaje na důchody

Struktura dat

Ve formátu **RDF** jsou data reprezentována pomocí slovníku **The RDF Data Cube Vocabulary**. S využitím tohoto slovníku byla vytvořena definice struktury datové kostky s výdaji na důchody podle roku a druhu důchodu. Fakty a dimenze tvoří strukturu této datové kostky jsou definovány v rámci jmenných prostorů, které uvádí následující tabulka.

Prefix	Jmenný prostor
cssz-dimension	https://data.cssz.cz/ontology/dimension/
cssz-measure	https://data.cssz.cz/ontology/measure/

Datová sada **Výdaje na důchody** má strukturu, kterou udává následující tabulka. Vedle názvu a popisu jednotlivých atributů je uveden i název sloupce, který obsahuje příslušný atribut v případě souborů ve formátu **CSV** a dále predikát použitý pro reprezentaci příslušného atributu ve formátu **RDF**.

Název atributu	Popis atributu	[CSV] Sloupec	[RDF] Predikát
Druh důchodu	Druh důchodu, pro který jsou výdaje na důchody uvedeny	druh_duchodu	cssz-dimension:druh-duchodu
Referenční období	Období, za které jsou výdaje na důchody uvedeny	referencni_obdobi	cssz-dimension:refPeriod
Výdaje na důchody opravené o zálohy v tisících Kč	Výdaje na důchody opravené o zálohy v tisících Kč	vydaje_na_duchody_opravene_o_zalohy_v_tis_kc	cssz-measure:vydaje-na-duchody-opravene-o-zalohy-v-tis-kc

Soubor ve formátu CSV obsahuje i kódy jednotlivých druhů důchodů (sloupec druh_duchodu_kod).

Použité číselníky a referenční údaje

V datové sadě jsou použity následující **pomocné číselníky**:

- Číselník druhů důchodů a jejich kombinací
- Číselník použitých let

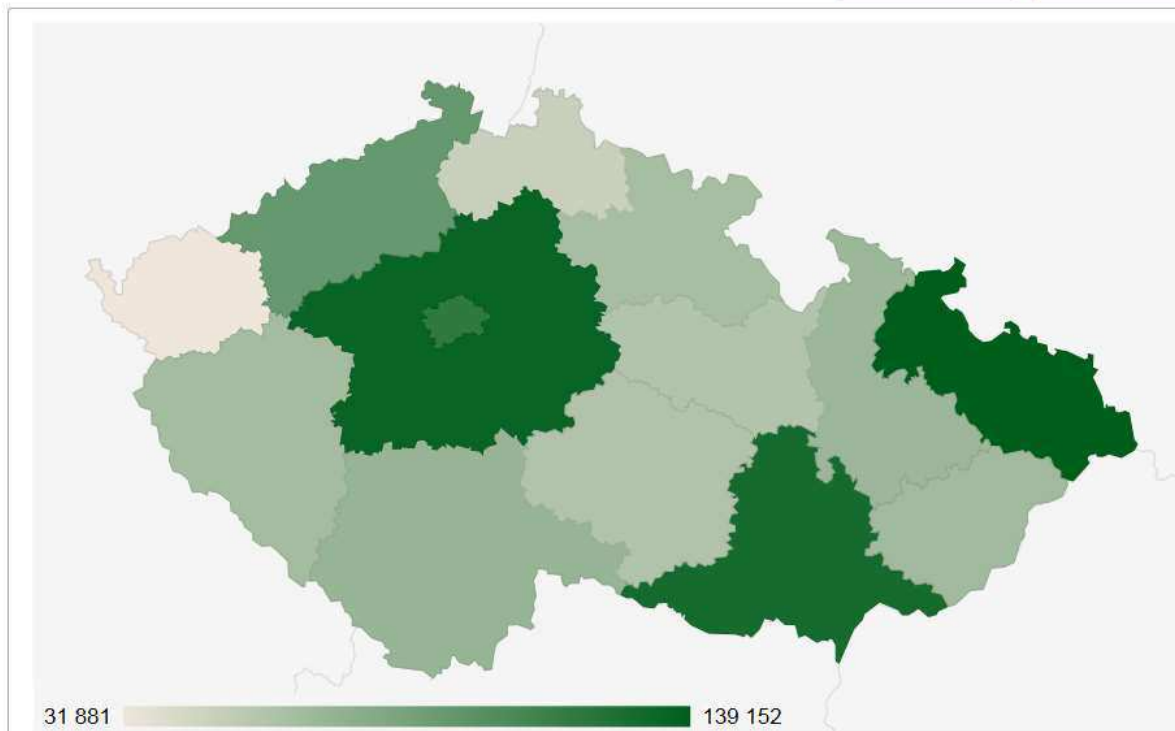
Počet důchodců v krajích ČR

Druh důchodu: Pohlaví: Rok:

Zobrazit graf

Zobrazit tabulku

Export



Počet důchodců v krajích ČR podle druhu důchodu, pohlaví a roku

Kódy krajů:

- CZ-JC - Jihočeský kraj
- CZ-JM - Jihomoravský kraj
- CZ-KA - Karlovarský kraj
- CZ-KR - Královéhradecký kraj
- CZ-LI - Liberecký kraj
- CZ-MO - Moravskoslezský kraj
- CZ-OL - Olomoucký kraj
- CZ-PA - Pardubický kraj
- CZ-PL - Plzeňský kraj
- CZ-PR - Hlavní město Praha
- CZ-ST - Středočeský kraj
- CZ-US - Ústecký kraj
- CZ-VY - Kraj Vysočina
- CZ-ZL - Zlínský kraj

Počet nově přiznaných invalidních důchodů v České republice

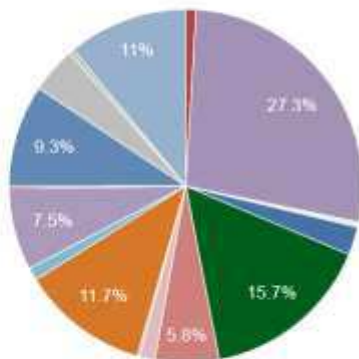
Druh důchodu: Pohlaví: Rok: Věková skupina:

Zobrazit graf

Zobrazit tabulku

Export

Počet nově přiznaných invalidních důchodů v České republice - koláčový graf



- 1. skupina
- 2. skupina
- 3. skupina
- 4. skupina
- 5. skupina
- 6. skupina
- 7. skupina
- 8. skupina
- 9. skupina
- 10. skupina
- 11. skupina
- 12. skupina
- 13. skupina
- 14. skupina
- 17. skupina
- 19. skupina

Počet nově přiznaných invalidních důchodů v jednotlivých skupinách diagnóz WHO z celkového počtu nově přiznaných invalidních důchodů

Skupiny diagnóz WHO:

1. skupina - Některé infekční a parazitární nemoci A00-B99
2. skupina - Novotvary C00-D48
3. skupina - Nemoci krve, krevtovorných orgánů a některé poruchy týkající se mechanismu imunity D50-D89
4. skupina - Nemoci endokrinní, výtvy a přeměny látek E00-E90
5. skupina - Duševní poruchy a poruchy chování F00-F99
6. skupina - Nemoci nervové soustavy G00-G99
7. skupina - Nemoci oka a očních adnex H00-H59
8. skupina - Nemoci ucha a bradavkového výběžku H60-H95
9. skupina - Nemoci oběhové soustavy I00-I99
10. skupina - Nemoci dýchací soustavy J00-J99
11. skupina - Nemoci trávicí soustavy K00-K93
12. skupina - Nemoci kůže a podkožního vaziva L00-L99
13. skupina - Nemoci svalové a kožní soustavy a pojivové tkáně M00-M99
14. skupina - Nemoci močové a pohlavní soustavy N00-N99
15. skupina - Těhotenství, porod a šestinedělí O00-O99
16. skupina - Některé stavy vzniklé v perinatálním období P00-P96
17. skupina - Vrozené vady, deformace a chromozomální abnormality Q00-Q99
18. skupina - Příznaky, znaky a abnormalní klinické a laboratorní nálezy nezařazené jinde R00-R99
19. skupina - Forezní, otravy a některé jiné následky vnějších příčin S00-T98

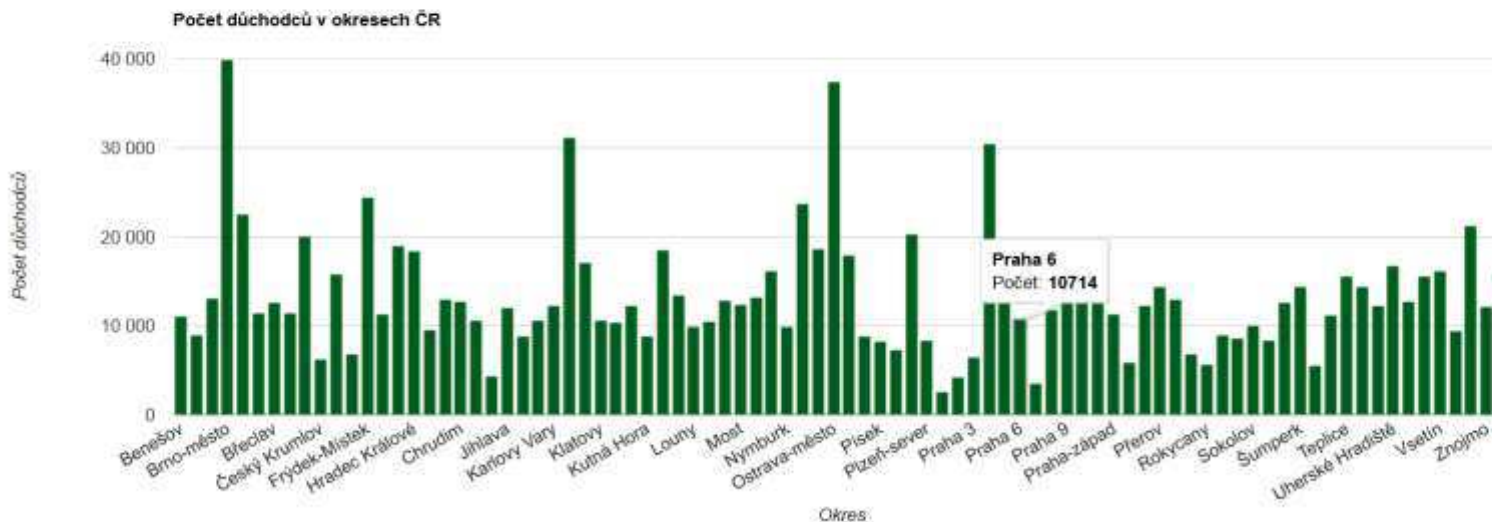
Počet důchodců v okresech ČR

Druh důchodu: Celkem bez netypických výplat Pohlaví: Muž Rok: 2014

Zobrazit graf

Zobrazit tabulku

CSV



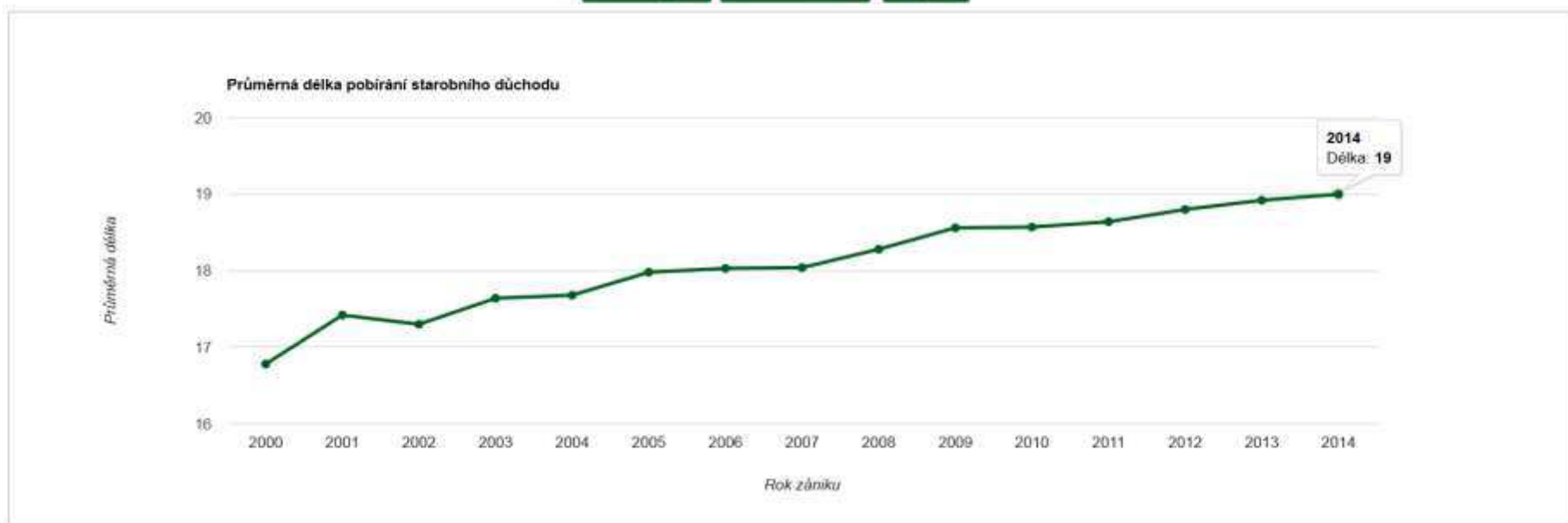
Průměrná délka pobírání starobního důchodu

Pohlaví: Muž

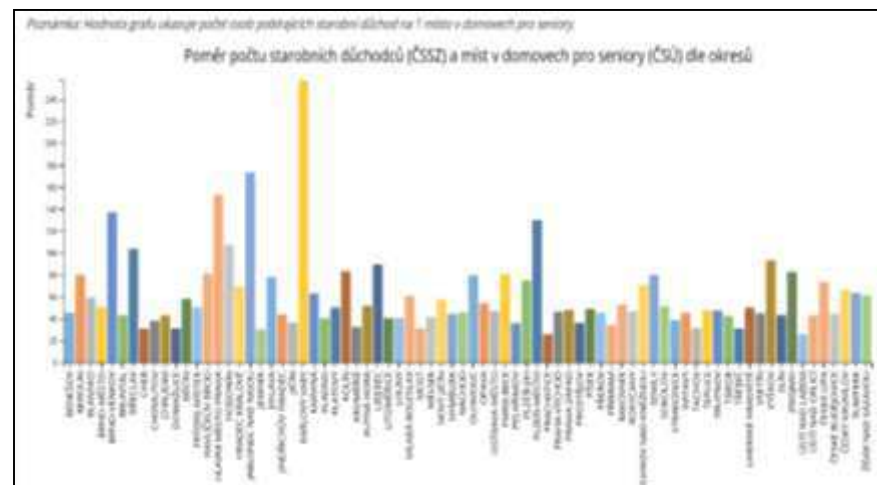
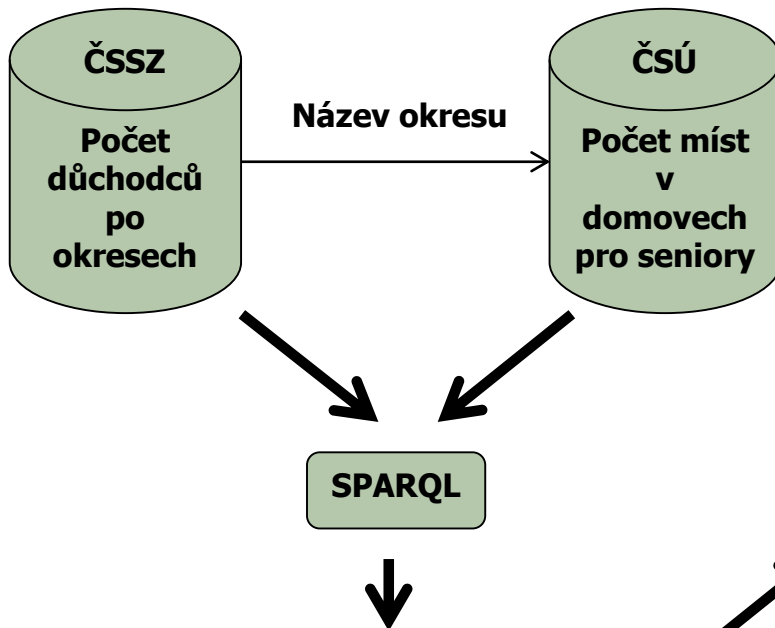
Zobrazit graf

Zobrazit tabulku

Export



Možnosti využití otevřených dat



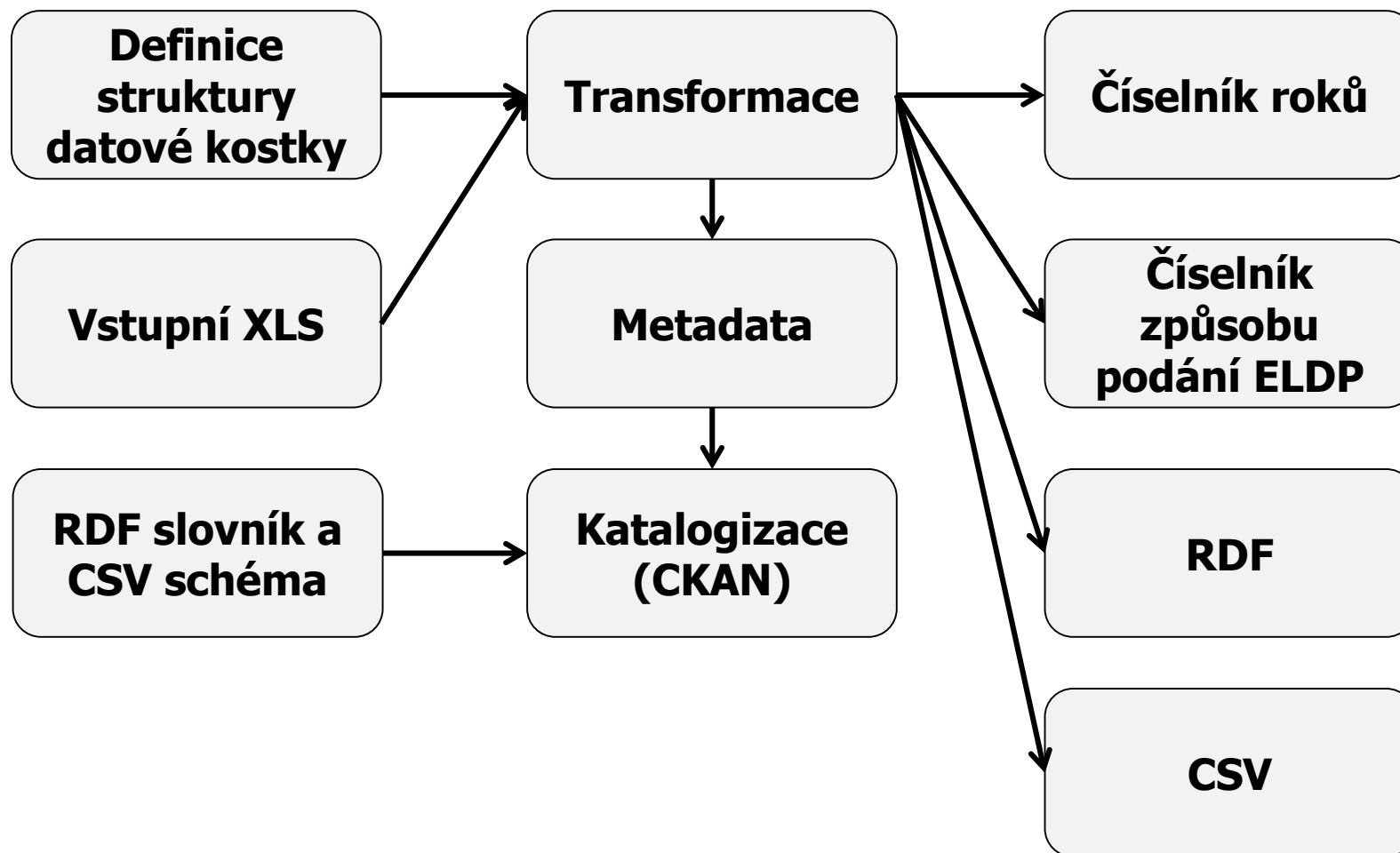
Rok	Okres	Poměr	Počet důchodců	Počet míst v domovech důchodců
2009	BENEŠOV	45,666	59868	1311
2009	BEROUN	80,354	49659	618
2009	BLANSKO	59,503	66405	1116
2009	BRNO-MĚSTO	50,76	245934	4845
2009	BRNO-VENKOV	137,636	113550	825
2009	BRUNTÁL	43,501	77084	1772
2009	BŘECLAV	104,132	68415	657
2009	CHEB	31,287	53031	1695

Zdroj: <https://opendata.vse.cz/duchodova-statistika>

Postup publikace

- Pohovor s vlastníkem dat
- Analýza zdrojových souborů
- Analýza datové sady
- Příprava vstupních dat pro transformaci
- Tvorba HTML dokumentace
- Sestavení transformačního procesu
- Vygenerování výstupních souborů
- Publikace dat

Datová sada – převod z CVS do RDF



Shrnutí – výhody automatizace

Výhody automatizace

- Snadný převod do formátu RDF i obtížných zdrojů (formátované XLS)
- Udržení pořádku při přípravě dat
- Možnost publikace dat ve více formátech – RDF, CSV
- Opakovatelnost spouštění při průběžných aktualizacích datových sad
- Dokumentování transformace dat
- Tvorba vzorů pro převody
- Jedno místo pořizování metadat

Shrnutí – kritické faktory úspěšnosti

Kritické faktory úspěšnosti procesu publikace otevřených dat:

- Podpora vedení, návaznost na Strategické cíle a Prioritní úkoly
- Stanovení odpovědného útvaru v rámci organizační struktury
- Expertní podpora
- Odborné zázemí v útvaru informatiky
- Úzká spolupráce s věcnými útvary
- Lidské zdroje

Způsob publikace otevřených dat ČSSZ může sloužit jako příklad dobré praxe, jak mají být otevřená data publikována a prezentována, a to jak pro Českou republiku, tak v mezinárodním kontextu.

Děkujeme za pozornost.

Ing. Michaela Hendrychová
michaela.hendrychova@cssz.cz

Ing. Jiří Šunka
jiri.sunka@cssz.cz

