



ČESKÁ SPRÁVA
SOCIÁLNÍHO ZABEZPEČENÍ

10 kroků k publikaci otevřených dat v ČSSZ

21. ročník konference ISSS
Hradec Králové, 9. 4. 2018

Ing. Jiří Šunka

kurátor otevřených dat ČSSZ





Deset kroků k publikaci:

1. Podpora
2. Organizační příprava
3. Technologie
4. Analýza zdrojů dat
5. Transformace dat
6. Dokumentace datové sady
7. Generování obsahu portálu
8. Vizualizace nad RDF databází
9. Rutinní řízení procesu publikace
10. Propagace



1. Podpora

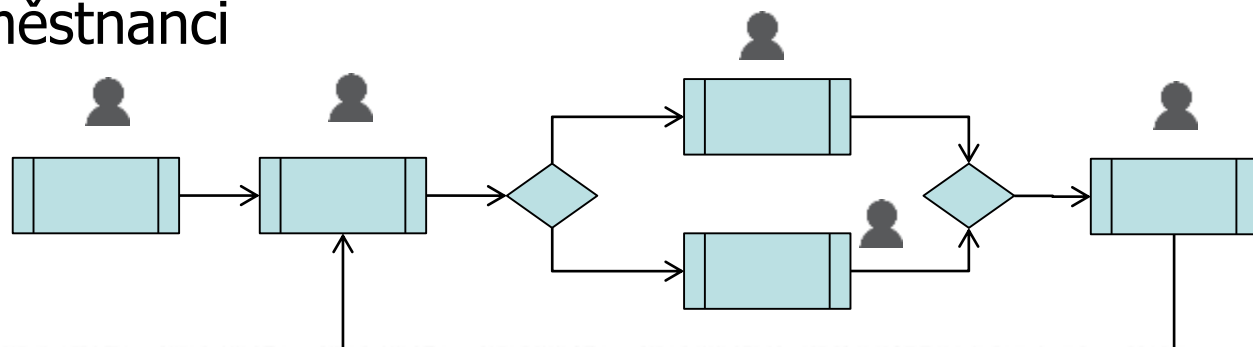
=> **kritický krok**

- Podpora a zájem vedení organizace
- Expertní podpora
- Součinnost s informatiky (IKT/ICT)
- Získat věcné útvary
- Kvalifikovaní zaměstnanci

2. Organizační příprava

=> **předpoklad úspěšného fungování** - vytvoření dobré směrnice

- Definice rolí
- Povinnosti účastníků procesu publikace
- Vlastníci dat × poskytovatelé dat
- Zapojení odboru komunikace





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
- Automatizované přenosy mezi přípravou a portálem
- Schvalovací workflow
- Optimalizace procesů

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

Katalogové záznamy datových sad

Filtrovat podle klíčového slova: [call centrum \(1\)](#) [číselník \(10\)](#) [diagnózy \(2\)](#) [doba řízení \(1\)](#) [důchodová ročenka \(17\)](#) [důchodové pojištění \(23\)](#) [exekuce \(1\)](#) [exekuční srážka \(1\)](#) [fluktuační \(1\)](#) [informativní osobní list důchodového pojištění \(1\)](#) [insolvenční řízení \(1\)](#) [invalidita \(4\)](#) [kontrola \(1\)](#) [kraje \(6\)](#) [lékařská posudková služba \(1\)](#) [námitky \(2\)](#) [nemocenské pojištění \(6\)](#) [okresy \(4\)](#) [osobní účet \(1\)](#) [osvc \(4\)](#) [podání \(1\)](#) [pohledávky \(2\)](#) [pojištěnci \(3\)](#) [pracovní neschopnost \(2\)](#) [průměrný plat \(1\)](#) [příslušnost k českým právním předpisům \(1\)](#) [sazba pojištění \(1\)](#) [sociální zabezpečení \(1\)](#) [statistika \(51\)](#) [svobodný přístup k informacím \(1\)](#) [systemizovaná místa \(1\)](#) [věk \(4\)](#) [výdaje \(1\)](#) [vyměřovací základ \(3\)](#) [vyplacené dávky \(7\)](#) [výplata důchodů \(4\)](#) [výše dávek \(5\)](#) [zaměstnanci \(5\)](#) [zaměstnavatelé \(4\)](#) [zánik výplaty \(2\)](#)

Číslo v závorce udává počet publikovaných datových sad, které obsahují dané klíčové slovo.

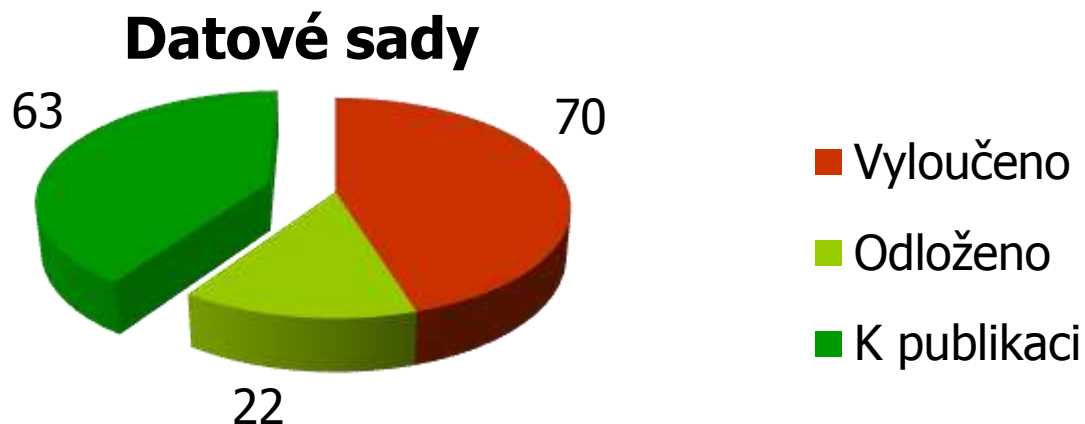
Datová sada	Vizuálizace
Číselník datových typů	<input type="checkbox"/> Číselník datových typů
Číselník druhů akcí v tiskopise Oznámení o nástupu do zaměstnání	<input type="checkbox"/> Číselník druhů akcí v tiskopise Oznámení o nástupu do zaměstnání
Číselník druhů pracovních činností	<input type="checkbox"/> Číselník druhů pracovních činností
Číselník druhů zaměstnání pro e-Neschopenku	<input type="checkbox"/> Číselník druhů zaměstnání pro e-Neschopenku
Číselník důvodů k ukončení pracovní neschopnosti	<input type="checkbox"/> Číselník důvodů k ukončení pracovní neschopnosti
Číselník okresů	<input type="checkbox"/> Číselník okresů
Číselník pohlaví	<input type="checkbox"/> Číselník pohlaví
Číselník specifikace cizozemského nositele pojištění	<input type="checkbox"/> Číselník specifikace cizozemského nositele pojištění
Číselník zdravotních pojišťoven	<input type="checkbox"/> Číselník zdravotních pojišťoven
Doba řízení o námitkách	<input checked="" type="checkbox"/> Doba řízení
Kontroly plnění povinností zaměstnavatelů v oblasti sociálního zabezpečení	<input checked="" type="checkbox"/> Počty kontrol plnění povinností zaměstnavatelů



Prvotní výběr dat k publikaci v roce 2015:

„Bezpečná“ data – již publikována na stránkách ČSSZ:	155 datových sad
Po vyloučení nevhodných sad k publikaci	: 85 datových sad
Po posouzení přínosů a rizik a odložení k řešení	: 22 datových sad

Do prvního plánu publikace zařazeno : 63 datových sad





Předzpracování vstupních dat:

- Vstupní data nemají strukturu jednoduché tabulky, vyskytují se na více listech nebo ve více vstupních souborech
- Převod vstupních dat do lépe zpracovatelné struktury (např. v MS Access)

UnifiedViews:

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



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. Fakta a dimenze tvořící 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



← Výdaje na důchody

Výdaje na důchody podle roku a druhu důchodu

Periodicita aktualizace	Jednou ročně
URI souvisejícího geografického území	http://ruian.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-duxody-v-cr
Dotčené časové období od-do	01.01.2008 - 31.12.2016
Název a URL poskytovatele dat	Česká správa sociálního zabezpečení

Datové zdroje

Data	Schéma	Podmínky užití	Dotčené časové období
Data (CSV) 4,6 kB	Schéma (JSON) 884 B	podmínky užití	01.01.2008 - 31.12.2016
Data (RDF/TriG) 46,4 kB	Schéma (RDF/TriG) 16,6 kB	podmínky užití	01.01.2008 - 31.12.2016

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.

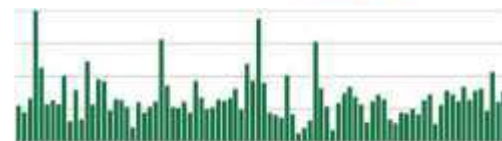
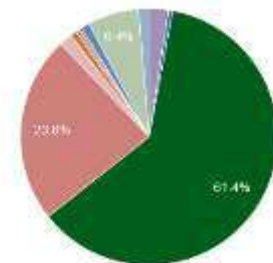
Metadata o datové sadě v RDF (TriG) 6,8 kB

Vizualizace

Výdaje na důchody

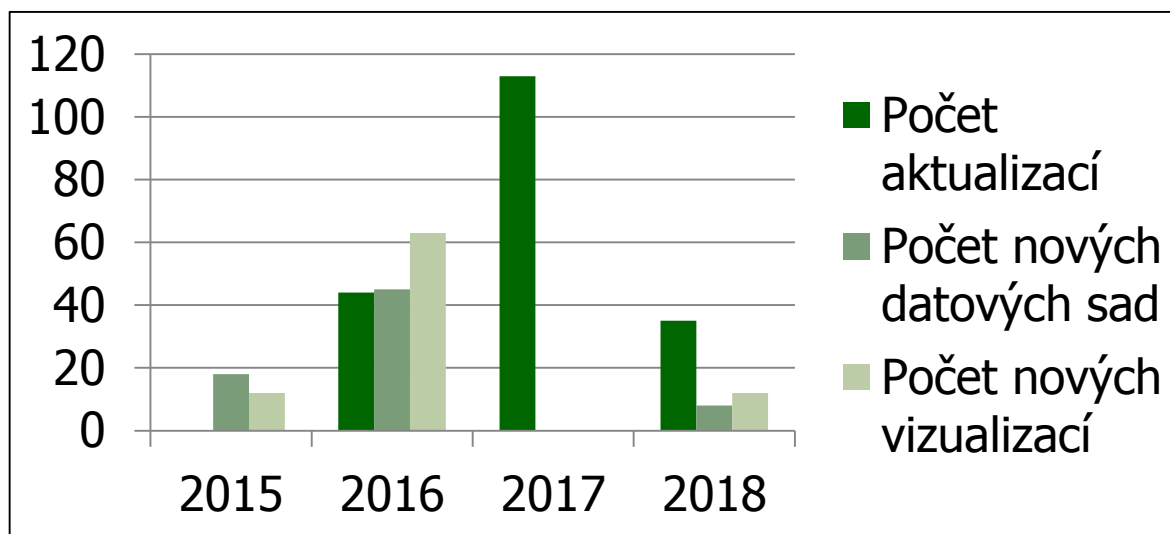
Interaktivní vizualizace:

- Základní typy grafů – spojnicový, sloupcový, koláčový, tabulka, mapy (kraj, okres, svět)
- 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





Rok	Počet aktualizací	Počet nových datových sad	Počet nových vizualizací
2015	0	18	12
2016	44	45	63
2017	113	0	0
2018	35 (95)	(8)	(12)
Celkem	192 (287)	63 (71)	75 (87)



Údaje v závorce jsou předpokládané počty pro rok 2018.



- **Informovanost veřejnosti**
 - Využití v tiskových zprávách
 - Účast na akcích k otevřeným datům (ISSS, Open Data Expo, hackathon)
- **Podněty uživatelů**
 - Granularita publikovaných datových sad
 - Četnost aktualizace
 - Podrobná data × problematika ochrany osobních údajů
- **Omezení**
 - Přístup ke zdrojovým datům
 - Konzistence s jinými formami publikace
- **A co dál?**
 - Nové datové sady



Otevřená data:

Nežádejte o ně, vezměte si je!





ČESKÁ SPRÁVA
SOCIÁLNÍHO ZABEZPEČENÍ

10 kroků k publikaci otevřených dat v ČSSZ

Děkuji za pozornost.

Ing. Jiří Šunka

kurátor otevřených dat ČSSZ

e-mail: opendata@cssz.cz

