Publikační databáze

Září 2014

Autoři

Autor	Organizace
Dušan Chlapek	Vysoká škola ekonomická v Praze
Tomáš Knap	Vysoká škola ekonomická v Praze
Jan Kučera	Vysoká škola ekonomická v Praze
Luboš Marek	Vysoká škola ekonomická v Praze
Petr Mazouch	Vysoká škola ekonomická v Praze
Martin Nečaský	Vysoká škola ekonomická v Praze
Jiří Makalouš	KOMIX s.r.o.
Tomáš Vahalík	KOMIX s.r.o
Jan Vrána	KOMIX s.r.o

Popis výstupu

Publikační databáze obsahuje data důchodové statistiky (data statistických ročenek z oblasti důchodového pojištění) České správy sociálního zabezpečení (ČSSZ) ve formátu RDF za roky 2008-2012. Tento dokument pak představuje dokumentaci této databáze.

Poděkování

Projekt "*Publikace dat statistických ročenek ve standardu otevřených dat*" (TD020121) je spolufinancován Technologickou agenturou České republiky.

Obsah

1 Úvod

Publikační databáze obsahuje data důchodové statistiky (data statistických ročenek z oblasti důchodového pojištění) České správy sociálního zabezpečení (ČSSZ) ve formátu RDF za roky 2008-2012. Publikační databáze slouží ke zpřístupnění těchto dat v podobě otevřených propojitelných dat. Aplikací, která využívá data z publikační databáze, je např. webová prezentační aplikace vyvinutá taktéž v rámci tohoto projektu. Tato aplikace je dostupná na adrese: <https://opendata.vse.cz/duchodova-statistika/>.

2 Přístup k publikační databázi

Data jsou uložena v úložišti OpenLink Software Virtuoso verze 07.10.321 a data jsou zpřístupněna prostřednictvím SPARQL endpointu. Údaje pro přístup k databázi jsou následující:

- **Adresa SPARQL endpointu**: <http://opendata.vse.cz:8890/sparql>.
- **Graf s daty**: <http://linked.opendata.cz/resource/dataset/cssz/pensions>.
- **Graf s definicemi kostek**: <http://linked.opendata.cz/resource/dataset/cssz/pensions/def>.

Pro využívání výše uvedeného SPARQL endpointu nejsou třeba žádné přihlašovací údaje.

Definice datových kostek a další ontologie jako je např. ontologie druhů důchodů jsou dostupné také z repositáře na adrese: <https://code.google.com/p/cssz-pensions/>.

Jednotlivé datové sady, které jsou obsaženy v publikační databázi, jsou uvedeny v datovém katalogu, který je dostupný na adrese: <opendata.vse.cz/catalog>. Pro jednotlivé datové sady jsou uvedena následující metadata:

- **Název** – název datové sady
- **Popis** – stručný popis datové sady
- **URI datasetu** – identifikátor datové sady v podobě HTTP URI
- **Vydavatel** – je uveden subjekt, který poskytuje datovou sadu; kromě tohoto subjektu může být dále uveden subjekt, který poskytl data, ze které datová sada vznikla
- **Licence** – licence, pod kterou je datová sada poskytována
- **Zdroje** – odkaz na původní data, ze kterých datová sada vznikla
- **Autoři** – osoba či subjekt, resp. osoby či subjekty, které datovou sadu vytvořily
- **Naposledy změněno** – datum poslední změny datové sady
- **Příklady entit** – odkazy na datové entity, které slouží jako příklady, pokud se uživatel chce seznámit s obsahem a strukturou datové sady
- **SPARQL endpoint** – odkaz na SPAQL endpoint pomocí které lze provádět dotazování nad datovou sadou pomocí jazyka SPARQL
- **Dump** – odkaz na stažitelný soubor, který obsahuje celý obsah datové sady

3 Podmínky užití dat

Definice datových kostek a další ontologie vytvořené v rámci projektu, metadata a data důchodové statistiky ve formátu RDF jsou zpřístupněny pod licencí Creative Commons Attribution 4.0 International Public License (CC BY 4.0).

Data obsažená v datových kostkách uvedených v tabulce 3 vznikla transformací dat publikovaných Českým statistickým úřadem (ČSÚ) do formátu RDF dle definovaných ontologií a nebyla významově upravena. Užití zdrojových dat poskytovaných ČSÚ se řídí Podmínkami pro využívání a další zveřejňování statistických údajů ČSÚ. Data ČSÚ ve formátu RDF jsou taktéž poskytována pod licencí CC BY 4.0.

Při dalším šíření a užití dat publikační databáze (včetně ontologií a definic datových kostek) uveďte následující poznámku:

„Publikační databáze obsahující data důchodové statistiky České správy sociálního zabezpečení (ČSSZ) za roky 2008-2012 a vybrané statistiky Českého statistického úřadu ve formátu RDF (dále jen publikační databáze) vznikla v rámci projektu TD020121 Publikace dat statistických ročenek ve standardu otevřených dat spolufinancovaného Technologickou agenturou České republiky. Pořizovateli databáze jsou Vysoká škola ekonomická v Praze a KOMIX s.r.o. Databáze podléhá licenci Creative Commons Attribution 4.0 International Public License. Databáze byla vytvořena na základě dat poskytnutých Českou správou sociálního zabezpečení a na základě dat publikovaných Českým statistickým úřadem na webových stránkách <http://www.czso.cz>. Užití dat poskytovaných ČSÚ se řídí Podmínkami pro využívání a další zveřejňování statistických údajů ČSÚ."

4 Struktura publikační databáze

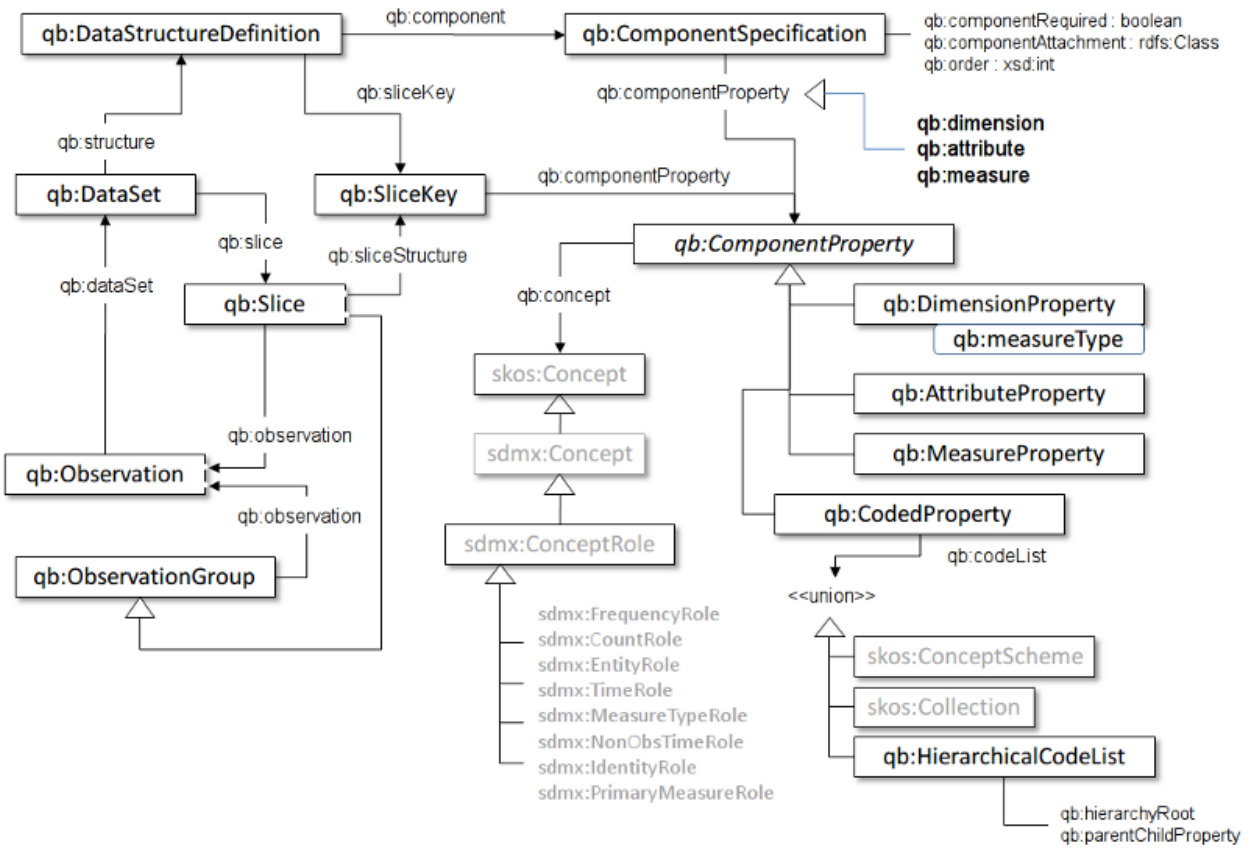
Data důchodové statistiky jsou již ve své zdrojové podobě organizována jako fakty (např. počet důchodců, počet vyplácených důchodů) a k nim přiřazené dimenze, pomocí kterých jsou jednotlivé fakty klasifikovány (např. pohlaví, rok či platnost údajů k určitému datu, druh důchod atd.). Zdrojovým formátem dat je formát MS Excel. Ačkoli data v MS Excel jsou organizována primárně jako tabulka, kde lze pro přiřazení dimenzí faktům využití záhlaví sloupce a řádků, zdrojové soubory důchodové statistiky využívají např. členění na listy, příp. umístění více tabulek na jeden list souboru k tomu, aby faktům obsaženým ve zdrojových souborech byly přiřazeny zpravidla více než dvě dimenze. Soubory jsou dostupné za jednotlivé roky, díky čemuž mají zdrojová data i časovou dimenzi.

Díky multidimenzionálnímu charakteru dat jsou data důchodové statistiky reprezentována ve formátu RDF pomocí ontologie The RDF Data Cube Vocabulary[1] (dále jen Data Cube). Data Cube je doporučení W3C a vychází z mezinárodního standardu SDMX[2]. Způsob reprezentace dat pomocí Data Cube je stručně popsán v následující části. Pro podrobnější seznámení s touto ontologií doporučujeme prostudování příslušného doporučení W3C.

4.1 The RDF Data Cube Vocabulary

Obrázek 1 znázorňuje schéma ontologie Data Cube.

[1] http://www.w3.org/TR/vocab-data-cube/
[2] http://sdmx.org/



Obrázek 1: Schéma ontologie Data Cube, zdroj: [Cyganiak & Reynolds, 2014]

Data reprezentovaná pomocí Data Cube představují tzv. datové kostky (`qb:DataSet`)[3]. Datové kostky jsou tvořeny pozorováními (`qb:Observation`). Pozorování pak dále obsahují fakt, jemu přiřazené dimenze a případně také mohou obsahovat i atribut faktu, např. jednotku, ve které je hodnota faktu vyjádřena.

Každá datová kostka pak má svoji definici struktury (`qb:DataStructureDefinition`), která udává, jaké fakty (`qb:MeasureProperty`), dimenze (`qb:DimensionProperty`) a příp. atributy (`qb:AttributeProperty`) jsou v kostce obsaženy.

Data Cube umožňuje definovat i tzv. řezy (`qb:Slice`), tj. předdefinované pohledy na data vzniklé volbou hodnot všech dimenzí s výjimkou jedné. Např. pokud bychom chtěli definovat řez v datové kostce obsahující *počet vyplácených důchodů* jako fakt, u kterého jsou sledovány dimenze *pohlaví*, *druh důchodu* a *stav údajů k určitému datu*, mohli bychom definovat řez tak, že bychom pevně stanovili hodnotu dimenze *pohlaví* na "ženy" a hodnotu *druhu důchodu* na "starobní důchody vyplácené sólo". Dimenze *stav k* by zůstala nezafixovaná a řez by nám tak

[3] Ztotožnění pojmu datová kostka a třídy `qb:DataSet` není zcela přesné vzhledem k terminologii používané v <http://www.w3.org/TR/vocab-data-cube/>. Protože ale za datovou sadu považujeme tzv. pojmenovaný graf, který obsahuje všechny datové kostky dat důchodové statistiky, jsou v tomto dokumentu označovány instance třídy `qb:DataSet` jako datové kostky, aby pojem datová sada nebyl používán v rámci dokumentu ve dvou různých významech.

poskytoval data o počtu starobních důchodů vyplácených sólo ženám za jednotlivá období, ke kterým jsou data vykazována.

Pohledy na data lze nicméně vytvářet i dynamicky pomocí dotazů v jazyce SPARQL[4]. V rámci projektu je vytvářena webová aplikace demonstrující využitelnost dat v publikační databázi pro tvorbu interaktivních výstupů[5]. Data z publikační databáze jsou tak zpřístupněna pomocí SPARQL endpointu a předdefinované řezy nejsou v publikační databázi využity.

4.2 Zdrojová data

Zdrojová data důchodové statistiky ČSSZ jsou uložena v sadě souborů MS Excel. Zdrojové soubory s daty pro ročenky 2008-2012 uvádí tabulka 1.

Tabulka 1: Přehled zdrojových souborů s daty pro ročenky 2008-2012

2008	2009	2010	2011	2012
05 Demografie 2008.xls	05 Demografie 2009.xls	05 Demografie 2010.xls	05 Demografie 2011.xls	05 Demografie 2012.xls
06 a 07 Agenda pro ročenku 2008.xls	06 a 07 Agenda pro ročenku 2009.xls	06 a 07 Agenda pro ročenku 2010.xls	06 a 07 Agenda pro ročenku 2011.xls	06 a 07 Agenda pro ročenku 2012.xls
08.01 Počet důchodců podle krajů.xls	08.01 Počet důchodců podle krajů.xls	08.01 Počet důchodců podle krajů.xls	08.01 Počet důchodců podle krajů.xls	08.01 Počet důchodců podle krajů.xls
08.02 Počet důchodců podle věku.xls	08.02 Počet důchodců podle věku.xls	08.02 Počet důchodců podle věku.xls	08.02 Počet důchodců podle věku.xls	08.02 Počet důchodců podle věku.xls
08.03 Počet důchodců podle výše důchodu.xls	08.03 Počet důchodců podle výše důchodu.xls	08.03 Počet důchodců podle výše důchodu.xls	08.03 Počet důchodců podle výše důchodu.xls	08.03 Počet důchodců podle výše důchodu.xls
09.01 Nově přiznané důchody dle věku důchodce.xls	09.01 Nově přiznané důchody dle věku důchodce.xls	09.01 Nově přiznané důchody dle věku důchodce.xls	09.01 Nově přiznané důchody dle věku důchodce.xls	09.01 Nově přiznané důchody dle věku důchodce.xls
09.02 Nově přiznané důchody dle výše důchodu.xls	09.02 Nově přiznané důchody dle výše důchodu.xls	09.02 Nově přiznané důchody dle výše důchodu.xls	09.02 Nově přiznané důchody dle výše důchodu.xls	09.02 Nově přiznané důchody dle výše důchodu.xls
09.03 Nově přiznané důchody dle OVZ.xls	09.03 Nově přiznané důchody dle OVZ.xls	09.03 Nově přiznané důchody dle OVZ.xls	09.03 Nově přiznané důchody dle OVZ.xls	09.03 Nově přiznané důchody dle OVZ.xls
10 Zaniklé důchody 2008.xls	10 Zaniklé důchody 2009.xls	10 Zaniklé důchody 2010.xls	10 Zaniklé důchody.xls	10 Zaniklé důchody.xls
11 Invalidita.xls	11 Invalidita.xls	11 Invalidita.xls	11 Invalidita.xls	11 Invalidita.xls
		12 Změny mezi stupni invalidního důchodu.xls	12 Změny mezi stupni invalidního důchodu.xls	12 Změny mezi stupni invalidního důchodu.xls
12 Data pro grafy 2008.xls	12 Data pro grafy 2009.xls	13 Data pro grafy 2010.xls	13 Data pro grafy 2011.xls	13 Data pro grafy 2012.xls
	2009_PREHEDOPOCTUDUCHODCUPOOKRESECH.xls	2010_prehledopoctuducchodcupookresech_1.xls	duchodcipookresechakrajichk31122011.xls	duchodcipookresechakrajichk31122012.xls

[4] http://www.w3.org/TR/sparql11-query/
[5] https://opendata.vse.cz/duchodova-statistika/

4.3 Datové kostky

Struktura publikační databáze je tvořena jednotlivými datovými kostkami. Jak bylo uvedeno v rámci představení ontologie Data Cube, každé datové kostce pak náleží definice její struktury. Tabulka 2 uvádí přehled vytvořených datových kostek pro data důchodové statistiky ČSSZ včetně definic jejich struktury. Ontologie definující strukturu těchto datových kostek je dostupná na adrese: https://code.google.com/p/cssz-pensions/source/browse/dc-definition/cssz-pensions-dc-definition.ttl .

Tabulka 2: Přehled datových kostek důchodové statistiky ČSSZ

Datová kostka	Název datové kostky	Definice struktury datové kostky	Název definice struktury datové kostky
http://linked.cssz.cz/dataset/penze/duchodci-prehled-cr	Celkový počet důchodců, průměrná výše důchodu a průměrný věk důchodců v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#duchodci-prehled-cr	Definice struktury datové kostky s počty důchodců v ČR.
http://linked.cssz.cz/dataset/penze/duchodci-v-krajich	Celkový počet důchodců v krajích České republiky	http://linked.cssz.cz/ontology/dataset-definitions/penze#duchodci-v-krajich	Definice struktury datové kostky s počty důchodců v krajích ČR.
http://linked.cssz.cz/dataset/penze/duchodci-v-cr-krajich-okresech	Celkový počet důchodců v krajích a okresech České republiky	http://linked.cssz.cz/ontology/dataset-definitions/penze#duchodci-v-cr-krajich-okresech	Definice struktury datové kostky s počty důchodců v ČR, krajích a okresech.
http://linked.cssz.cz/dataset/penze/casove-rady-grafy-1-6	Důchody a důchodci v České republice - různé statistiky	http://linked.cssz.cz/ontology/dataset-definitions/penze#casove-rady-grafy-1-6	Definice struktury datové kostky pro data časových řad obsažená v přílohách statistické ročenky (grafy 1-6).
http://linked.cssz.cz/dataset/penze/rozlozeni-souboru-duchodcu-podle-vyse-duchodu-v-kvantilovem-vyjadreni	Měsíční výše důchodů	http://linked.cssz.cz/ontology/dataset-definitions/penze#rozlozeni-souboru-duchodcu-podle-vyse-duchodu-v-kvantilovem-vyjadreni	Definice struktury datové kostky přehledu o rozložení souboru důchodců podle výše důchodů v kvantilovém vyjádření.
http://linked.cssz.cz/dataset/penze/duchodci-v-cr	Počet důchodců v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#duchodci-v-cr	Definice struktury datové kostky s počty důchodců v České republice.
http://linked.cssz.cz/dataset/penze/nove-priznane-duchody-v-cr	Počet nově přiznaných důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#nove-priznane-duchody-v-cr	Definice struktury datové kostky s počty nově přiznaných důchodů v České republice.
http://linked.cssz.cz/dataset/penze/invalidita	Počet nově přiznaných invalidních důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#invalidita	Definice struktury datové kostky počtu nově přiznaných důchodů podle skupin diagnóz definovaných v MKN-10.
http://linked.cssz.cz/dataset/penze/nove-priznane-duchody-dle-vyse-duchodu	Počet nově přiznaných důchodů v České republice dle měsíční výše důchodu	http://linked.cssz.cz/ontology/dataset-definitions/penze#nove-priznane-duchody-dle-vyse-duchodu	Definice struktury datové kostky s počty nově přiznaných důchodů v ČR dle měsíční výše důchodu.

Datová kostka	Název datové kostky	Definice struktury datové kostky	Název definice struktury datové kostky
http://linked.cssz.cz/dataset/penze/nove-priznane-duchody-dle-veku	Počet nově přiznaných důchodů v České republice dle věkové kategorie	http://linked.cssz.cz/ontology/dataset-definitions/penze#nove-priznane-duchody-dle-veku	Definice struktury datové kostky s počty nově přiznaných důchodů v ČR dle věku důchodce.
http://linked.cssz.cz/dataset/penze/nove-priznane-duchody-dle-osobniho-vymerovaciho-zakladu	Počet nově přiznaných důchodů v České republice dle osobního vyměřovacího základu	http://linked.cssz.cz/ontology/dataset-definitions/penze#nove-priznane-duchody-dle-osobniho-vymerovaciho-zakladu	Definice struktury datové kostky s počty nově přiznaných důchodů v ČR dle osobního vyměřovacího základu.
http://linked.cssz.cz/dataset/penze/obyvatelstvo-podle-kraju	Počet obyvatel v krajích České republiky	http://linked.cssz.cz/ontology/dataset-definitions/penze#obyvatelstvo-podle-kraju	Definice struktury datové kostky obyvatelstva České republiky v členění dle krajů.
http://linked.cssz.cz/dataset/penze/obyvatelstvo-podle-veku	Počet obyvatel ve věkových skupinách	http://linked.cssz.cz/ontology/dataset-definitions/penze#obyvatelstvo-podle-veku	Definice struktury datové kostky obyvatelstva České republiky v členění dle věku.
http://linked.cssz.cz/dataset/penze/vyplacene-duchody-v-cr	Počet vyplacených důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#vyplacene-duchody-v-cr	Definice struktury datové kostky s počty vyplácených důchodů v České republice.
http://linked.cssz.cz/dataset/penze/vyplacene-duchody-dle-vyse	Počet vyplacených důchodů v České republice dle měsíční výše důchodu	http://linked.cssz.cz/ontology/dataset-definitions/penze#vyplacene-duchody-dle-vyse	Definice struktury datové kostky s počty vyplácených důchodů v ČR dle měsíční výše důchodu.
http://linked.cssz.cz/dataset/penze/duchody-dle-veku	Počet vyplacených důchodů v České republice dle věkové kategorie	http://linked.cssz.cz/ontology/dataset-definitions/penze#duchody-dle-veku	Definice struktury datové kostky s počty důchodů podle věku důchodce.
http://linked.cssz.cz/dataset/penze/zmeny-mezi-stupni-invalidniho-duchodu	Počet vyplácených invalidních důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#zmeny-mezi-stupni-invalidniho-duchodu	Definice struktury datové kostky počtu invalidních důchodů po změně stupně invalidity.
http://linked.cssz.cz/dataset/penze/casove-rady-grafy-7-8	Počet vyplácených předčasných starobních důchodů, výdaje na předčasné starobní důchody v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#casove-rady-grafy-7-8	Definice struktury datové kostky pro data časových řad obsažená v přílohách statistické ročenky (grafy 7-8).
http://linked.cssz.cz/dataset/penze/zanikle-duchody	Počet zaniklých důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#zanikle-duchody	Definice struktury datové kostky počtu zaniklých důchodů dle důvodu zániku důchodu.
http://linked.cssz.cz/dataset/penze/prum-delka-pobirani-s-duchodu	Průměrná délka pobírání starobního důchodu	http://linked.cssz.cz/ontology/dataset-definitions/penze#prum-delka-pobirani-s-duchodu	Definice struktury datové kostky s daty o průměrné délce vyplácení starobního důchodu.
http://linked.cssz.cz/dataset/penze/prum-vyse-duchodu-u-nove-priznanych-duchodu-podle-druhu-duchodu	Průměrná výše důchodů v Kč u nově přiznávaných důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#prum-vyse-duchodu-u-nove-priznanych-duchodu-podle-druhu-duchodu	Definice struktury datové kostky s daty o průměrné výši důchodu u nově přiznaných důchodů dle druhu důchodu.

Datová kostka	Název datové kostky	Definice struktury datové kostky	Název definice struktury datové kostky
http://linked.cssz.cz/dataset/penze/prum-vyse-duchodu-podle-druhu-duchodu	Průměrná výše důchodu v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#prum-vyse-duchodu-podle-druhu-duchodu	Definice struktury datové kostky udávající průměrnou měsíční výši důchodu dle druhu důchodu.
http://linked.cssz.cz/dataset/penze/prum-vyse-osobniho-vymerovaciho-zakladu-u-nove-priznanych-duchodu-podle-druhu-duchodu	Průměrná výše osobního vyměřovacího základu u nově přiznávaných důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#prum-vyse-osobniho-vymerovaciho-zakladu-u-nove-priznanych-duchodu-podle-druhu-duchodu	Definice struktury datové kostky průměrné výše osobního vyměřovacího základu u nově přiznaných důchodů dle druhu důchodu.
http://linked.cssz.cz/dataset/penze/prum-vek-duchodce-dle-druhu-duchodu	Průměrný věk důchodce v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#prum-vek-duchodce-dle-druhu-duchodu	Definice struktury datové kostky s průměrným věkem důchodců dle druhu důchodu.
http://linked.cssz.cz/dataset/penze/prum-vek-u-nove-priznanych-duchodu-dle-druhu	Průměrný věk u nově přiznaných důchodů v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#prum-vek-u-nove-priznanych-duchodu-dle-druhu	Definice struktury datové kostky pro průměrný věk důchodce u nově přiznaných důchodů podle druhu důchodu.
http://linked.cssz.cz/dataset/penze/relat-rozlozeni-populace	Relativní rozložení populace	http://linked.cssz.cz/ontology/dataset-definitions/penze#relat-rozlozeni-populace	Definice struktury datové kostky popisující rozložení populace České republiky mezi definované skupiny.
http://linked.cssz.cz/dataset/penze/srovnani-vekoveho-slozeni-obyvatel-v-letech	Srovnání počtu obyvatel ve věkových skupinách	http://linked.cssz.cz/ontology/dataset-definitions/penze#srovnani-vekoveho-slozeni-obyvatel-v-letech	Definice struktury datové kostky porovnávající počet obyvatel České republiky v jednotlivých věkových skupinách s predikcí pro rok 2030.
http://linked.cssz.cz/dataset/penze/vydaje-na-duchody-v-cr	Výdaje na důchody	http://linked.cssz.cz/ontology/dataset-definitions/penze#vydaje-na-duchody-v-cr	Definice struktury datové kostky s celkovými výdaji na důchody v České republice v členění dle druhů důchodu.
http://linked.cssz.cz/dataset/penze/vydaje-na-duchody-od-roku-1996	Výdaje na důchody od roku 1996	http://linked.cssz.cz/ontology/dataset-definitions/penze#vydaje-na-duchody-od-roku-1996	Definice struktury datové kostky s celkovými výdaji na důchody v České republice v členění dle let.
http://linked.cssz.cz/dataset/penze/casove-rady-grafy-9	Výdaje na starobní důchody v České republice	http://linked.cssz.cz/ontology/dataset-definitions/penze#casove-rady-grafy-9	Definice struktury datové kostky pro data časových řad obsažená v přílohách statistické ročenky (graf 9).

Publikační databáze neobsahuje pouze data důchodové statistiky ČSSZ, ale pro demonstraci výhod propojování dat a také za účelem umožnit výpočet zajímavých ukazatelů s využitím dat důchodové statistiky v kombinaci s daty z jiných datových zdrojů, obsahuje publikační databáze také datové kostky s daty poskytovanými Českým statistickým úřadem. Tyto datové kostky jsou vedeny v tabulce 3. Datové kostky jsou taktéž reprezentovány v podobě Data Cube. Reprezentaci jsme vytvořili na základě zdrojových dat Českého statistického úřadu poskytovaných prostřednictvím Veřejné databáze[6] Českého statistického úřadu ve formátu XLS.

[6] http://vdb.czso.cz/

Pro přístup k datovým kostkám prostřednictvím serveru VŠE platí stejná pravidla jako výše uvedená pro datové kostky České správy sociálního zabezpečení.

Tabulka 3: Přehled ostatních datových kostek

Datová kostka	Název datové kostky	Definice struktury datové kostky	Název definice struktury datové kostky
http://data.czso.cz/resource/dataset/demography	Počty obyvatel v regionech ČR	http://data.czso.cz/ontology/dataset-definition/DemographyDefinition	Definice datové kostky obsahující demografická data o regionech České republiky
http://data.czso.cz/resource/dataset/social-service-facilities	Statistická data o zařízeních sociálních služeb a domů s pečovatelskou službou v okresech ČR	http://data.czso.cz/ontology/dataset-definition/SocialServiceFacilitiesDefinition	Definice datové kostky obsahující statistická data o zařízeních sociálních služeb a domů s pečovatelskou službou.
http://data.czso.cz/resource/dataset/average-salaries	Průměrná mzda v krajích České republiky	http://data.czso.cz/ontology/dataset-definition/AverageSalariesDefinition	Definice datové kostky s průměrnou mzdou v krajích České republiky
http://data.czso.cz/resource/dataset/selected-indicators-of-public-health	Vybrané ukazatele zdravotního stavu v regionech České republiky	http://data.czso.cz/ontology/dataset-definition/SelectedIndicatorsOfPublicHealthDefinition	Definice datové kostky s vybranými ukazateli zdravotního stavu v regionech České republiky
http://data.czso.cz/resource/dataset/job-applicants	Statistická data o neumístěných uchazečích o zaměstnání v regionech ČR	http://data.czso.cz/ontology/dataset-definition/JobApplicantsDefinition	Definice datové kostky se statistikou neumístěných uchazečů o zaměstnání v regionech České republiky
http://data.czso.cz/resource/dataset/unemployment-rate	Statistická data o míře registrované nezaměstnanosti v regionech ČR	http://data.czso.cz/ontology/dataset-definition/UnemploymentRateDefinition	Definice datové kostky se statistikou míry registrované nezaměstnanosti v regionech České republiky
http://data.czso.cz/resource/dataset/job-applicants-and-unemployment-rate	Statistická data o uchazečích o zaměstnání a podílu nezaměstnaných osob v regionech ČR	http://data.czso.cz/ontology/dataset-definition/JobApplicantsAndUnemploymentRateDefinition	Definice datové kostky se statistikou neumístěných uchazečů o zaměstnání a mírou nezaměstnanosti v regionech České republiky
http://data.czso.cz/resource/dataset/deaths-by-selected-causes-of-death	Statistická data s mrtvými dle vybraných příčin úmrtí v regionech ČR	http://data.czso.cz/ontology/dataset-definition/DeathsBySelectedCausesOfDeathDefinition	Definice datové kostky s mrtvými dle vybraných příčin úmrtí v regionech České republiky

5 Propojení

V tabulce 3 v předcházející části jsou uvedeny datové kostky obsahující data nikoli důchodové statistiky ČSSZ, ale data publikovaná Českým statistickým úřadem, která byla taktéž převedena do formátu RDF dle ontologie Data Cube. Datové kostky s daty důchodové statistiky a datové kostky s daty o počtu obyvatel, zařízeních sociálních služeb, průměrných mzdách, zdravotním stavu obyvatelstva, příčinách úmrtí a nezaměstnanosti v regionech používají ve svých dimenzích shodné entity. Zejména se jedná o entity představující kraje, pohlaví, roky, či věková pásma. Prostřednictvím těchto entit jsou data z jednotlivých datových kostek propojena.

Propojení pak umožňuje analyzovat data z různých datových kostek ve vzájemných souvislostech a lze definovat ukazatele, které např. vztahují hodnoty faktů v datech důchodové statistiky k populaci příslušného kraje či okresu. Využití propojení a ověřování, zda datové kostky skutečně ve svých dimenzích odkazují na shodné entity je blíže popsáno v kapitole "*Způsob práce s daty*".

6 Způsob práce s daty

Dle principů propojitelných dat je každá datová kostka jednoznačně identifikována svým URL (uvedené v tabulce v prvním sloupci). URL není pouze identifikátorem, ale také tzv. lokátorem. Tj. prostřednictvím standardního webového protokolu HTTP je možné k datové kostce přímo přistoupit, např. s pomocí webového prohlížeče. URL datové kostky má následující tvar:

http://linked.cssz.cz/dataset/penze/[ID-DATOVE-KOSTKY]

tj. URL je v doméně České správy sociálního zabezpečení. Data jsou však v této fázi projektu dostupná pouze prostřednictvím serveru provozovaného Vysokou školou ekonomickou (VŠE) a datové kostky tedy nejsou přístupné přímým přístupem na jejich URL. Je nutné přistoupit na server VŠE pomocí URL v následujícím tvaru:

http://opendata.vse.cz:8890/describe/?url=[URL-DATOVE-KOSTKY]

Např. pro přístup k datové kostce s URL

http://linked.cssz.cz/dataset/penze/duchodci-v-krajich

lze tedy využít URL

http://opendata.vse.cz:8890/describe/?url=http://linked.cssz.cz/dataset/penze/duchodci-v-krajich

Jelikož jsou definice a obsah datových kostky uloženy v RDF databázi a zpřístupněna prostřednictvím SPARQL endpointu, je možné s kostkami vzdáleně pracovat pomocí dotazovacího jazyka SPARQL. Nejedná se samozřejmě o využití koncovým uživatelem (pro něj je určena v projektu vyvíjená webová aplikace), ale o využití expertním uživatelem se znalostí dotazovacího jazyka SPARQL[7]. Pro takového expertního uživatele uvádíme jeden ze scénářů demonstrující práci s datovými kostkami ČSSZ reprezentované v podobě RDF propojené na datové kostky ČSÚ taktéž reprezentované v podobě RDF.

Východisko scénáře: Zajímáme se, jaké statistické údaje jsou měřeny pro daný okres a v jakých kostkách ČSSZ, resp. ČSÚ. Pokud nalezneme dvě kostky, jednu z ČSSZ a jednu z ČSÚ, které jsou porovnatelné, snažíme se je srovnat a zkombinovat do jedné datové kostky ČSÚ publikuje demografické a jiné údaje z okresů a krajů České republiky. ČSSZ publikuje statistiky důchodů z okresů a krajů České republiky.

Následuje popis jednotlivých kroků scénáře. Každý krok definuje několik SPARQL dotazů, které je možné vyzkoušet ve výše zmiňovaném SPARQL endpointu

[7] Pro seznámení s jazykem SPARQL doporučujeme specifikaci jazyka na stránkách konsorcia W3 : http://www.w3.org/TR/sparql11-query/

http://opendata.vse.cz:8890/sparql.

Krok č. 1) Začneme SPARQL dotazem, zda existují nějaká fakta týkající se okresu Semily. Pro identifikaci okresu Semily používám URL dané systémem RUIAN[8] (resp. jeho Linked Data reprezentací vytvořenou iniciativou OpenData.cz):
http://ruian.linked.opendata.cz/resource/okresy/3608

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?fakt
WHERE {
  ?fakt a qb:Observation ;
    ?dimenzeNeboMetrika <http://ruian.linked.opendata.cz/resource/okresy/3608> .
}
```

Krok č. 2) Dotazem jsme zjistili, že nějaká fakta existují. Zeptáme se na kostky, do kterých fakta patří a na jaké dimenzi či metrice mají okres Semily umístěn.

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?kostka ?nazevKostky ?dimenzeNeboMetrika ?nazevDimenzeNeboMetriky ?dimenzeNeboMetrikaTyp
WHERE {
  ?kostka a qb:DataSet ;
    dcterms:title ?nazevKostky .
  FILTER( lang(?nazevKostky) = "cs" )
  ?fakt a qb:Observation ;
    qb:dataSet ?kostka ;
    ?dimenzeNeboMetrika <http://ruian.linked.opendata.cz/resource/okresy/3608> .
  ?dimenzeNeboMetrika a ?dimenzeNeboMetrikaTyp ;
    rdfs:label ?nazevDimenzeNeboMetriky .
  FILTER( lang(?nazevDimenzeNeboMetriky) = "cs" )
  FILTER (?dimenzeNeboMetrikaTyp = qb:MeasureProperty OR ?dimenzeNeboMetrikaTyp = qb:DimensionProperty)
}
```

Krok č. 3) Zjistili jsme, že existuje 8 datových kostek. Jedna publikovaná ČSSZ, a zbylé publikované ČSÚ. ČSSZ publikuje informace o celkovém počtu důchodců v krajích a okresech České republiky. ČSÚ publikuje informace o celkové populaci v okresech České republiky, informace o zařízeních sociálních služeb a další datové kostky. Nabízí se tedy kombinovat kostky dohromady a získat zajímavé statistické údaje. Abychom ale mohli kostky srovnat, musíme ověřit, zda mají srovnatelnou strukturu. (Dále ve scénáři pracujeme jen s datovými kostkami ČSÚ obsahující demografické údaje a informace o zařízeních sociálních služeb. Zbylé kostky pro jednoduchost opomíjíme.)

Ověříme nejprve strukturu kostky s demografií publikované ČSÚ, tj. jaké dimenze a metriky obsahuje:

[8] Registr územních identifikátorů, adres a nemovitostí. http://vdp.cuzk.cz/

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?dimensionOrMeasure ?dimensionOrMeasureType
WHERE {
  ?observation a qb:Observation ;
    ?dimensionOrMeasure ?value ;
    qb:dataSet <http://data.czso.cz/resource/dataset/demography> .
  ?dimensionOrMeasure a ?dimensionOrMeasureType .
  FILTER (?dimensionOrMeasureType = qb:MeasureProperty OR ?dimensionOrMeasureType = qb:DimensionProperty)
}
```

Výsledek jsou následující dimenze a metriky:

dimensionOrMeasure	dimensionOrMeasureType
http://data.czso.cz/ontology/populace	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/refArea	http://purl.org/linked-data/cube#DimensionProperty
http://data.czso.cz/ontology/refPeriod	http://purl.org/linked-data/cube#DimensionProperty
http://data.czso.cz/ontology/pohlavi	http://purl.org/linked-data/cube#DimensionProperty
http://data.czso.cz/ontology/vekovakategorie	http://purl.org/linked-data/cube#DimensionProperty

A strukturu kostky se zařízeními sociálních služeb publikované ČSÚ

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?dimensionOrMeasure ?dimensionOrMeasureType
WHERE {
  ?observation a qb:Observation ;
    ?dimensionOrMeasure ?value ;
    qb:dataSet <http://data.czso.cz/resource/dataset/social-service-facilities> .
  ?dimensionOrMeasure a ?dimensionOrMeasureType .
  FILTER (?dimensionOrMeasureType = qb:MeasureProperty OR ?dimensionOrMeasureType = qb:DimensionProperty)
}
```

Výsledek jsou následující dimenze a metriky:

dimensionOrMeasure	dimensionOrMeasureType
http://data.czso.cz/ontology/pocetAzylovychDomu	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/pocetDomovuProDuchodce	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/pocetDomovuProOsobySeZdravotnimPostizenim	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/pocetDomuSPecovatelskouSluzbou	http://purl.org/linked-data/cube#MeasureProperty

dimensionOrMeasure	dimensionOrMeasureType
http://data.czso.cz/ontology/pocetMistVAzylovychDomech	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/pocetMistVDomechSPecovatelskouSluzbou	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/pocetMistVDomovechProDuchodce	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/pocetMistVDomovechProOsobySeZdravotnimPostizenim	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/pocetSocialnichZarizeni	http://purl.org/linked-data/cube#MeasureProperty
http://data.czso.cz/ontology/refArea	http://purl.org/linked-data/cube#DimensionProperty
http://data.czso.cz/ontology/refPeriod	http://purl.org/linked-data/cube#DimensionProperty

A poté ČSSZ

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?dimensionOrMeasure ?dimensionOrMeasureType
WHERE {
  ?observation a qb:Observation ;
    ?dimensionOrMeasure ?value ;
    qb:dataSet <http://linked.cssz.cz/dataset/penze/duchodci-v-cr-krajich-okresech> .
  ?dimensionOrMeasure a ?dimensionOrMeasureType .
  FILTER (?dimensionOrMeasureType = qb:MeasureProperty OR ?dimensionOrMeasureType = qb:DimensionProperty)
}
```

Výsledek jsou dimenze a metriky:

dimensionOrMeasure	dimensionOrMeasureType
http://linked.cssz.cz/ontology/dataset-definitions/penze#pocet-duchodcu	http://purl.org/linked-data/cube#MeasureProperty
http://linked.cssz.cz/ontology/dataset-definitions/penze#druh-duchodu	http://purl.org/linked-data/cube#DimensionProperty
http://linked.cssz.cz/ontology/dataset-definitions/penze#pohlavi	http://purl.org/linked-data/cube#DimensionProperty
http://linked.cssz.cz/ontology/dataset-definitions/penze#refArea	http://purl.org/linked-data/cube#DimensionProperty
http://linked.cssz.cz/ontology/dataset-definitions/penze#refPeriod	http://purl.org/linked-data/cube#DimensionProperty

dimensionOrMeasure	dimensionOrMeasureType
http://linked.cssz.cz/ontology/dataset-definitions/penze#prumerny-vek	http://purl.org/linked-data/cube#MeasureProperty
http://linked.cssz.cz/ontology/dataset-definitions/penze#prumerna-vyse-duchodu-v-kc	http://purl.org/linked-data/cube#MeasureProperty

Krok č. 4) Pomocí dotazů jsme zjistili, že lze srovnávat následující kostky:

Demografie ČSÚ	Důchodová statistika ČSSZ
Statistiky sociálních zařízení ČSÚ	Důchodová statistika ČSSZ

Kostky mohou být porovnatelné, neboť se zdá, že v prvním případě se shodují na dimenzích pohlaví a referenční období a zřejmě i na dimenzi určující místo. V druhém případě se zdá, že se shodují na dimenzích referenční období a místo. Bohužel, ale shoda není přesná, tj. nepoužívají úplně stejné dimenze. Jen se zdá, podle jejich URL identifikátorů, že by mohly znamenat to samé. Pro ověření musíme provést další ověření - zda se kostky shodují i na hodnotách v dimenzích.

Začneme pohlavím.

V kostce s demografií ČSÚ (kostku se sociálními zařízeními neprozkoumáváme, neboť dimenzi pohlaví nepoužívá):

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?hodnota
WHERE {
  ?observation a qb:Observation ;
    <http://data.czso.cz/ontology/pohlavi> ?hodnota ;
    qb:dataSet <http://data.czso.cz/resource/dataset/demography> .
}
ORDER BY ?hodnota
```

Výsledek jsou následující hodnoty:

hodnota
http://purl.org/linked-data/sdmx/2009/code#sex-F
http://purl.org/linked-data/sdmx/2009/code#sex-M

V kostce ČSSZ:

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?hodnota
WHERE {
  ?observation a qb:Observation ;
    <http://linked.cssz.cz/ontology/dataset-definitions/penze#pohlavi> ?hodnota ;
    qb:dataSet <http://linked.cssz.cz/dataset/penze/duchodci-v-cr-krajich-okresech> .
}
ORDER BY ?hodnota
```

Výsledek jsou následující hodnoty:

hodnota
http://purl.org/linked-data/sdmx/2009/code#sex-F
http://purl.org/linked-data/sdmx/2009/code#sex-M
http://purl.org/linked-data/sdmx/2009/code#sex-T

Vidíme, že se hodnoty shodují, a na dimenzích pohlaví tak jsou kostky plně srovnatelné. Podobně ověříme referenční období a místo. U dimenze určující místo je situace trochu komplikovanější neboť dotazem zjistíme, že kostka ČSÚ pokrývá pouze okresy. Naproti tomu kostka ČSSZ pokrývá okresy, kraje i městské části Prahy.

U obou kostek ČSÚ zjistíme na dimenzí určující místo, že se na ní vyskytují okresy reprezentované pomocí URI přiřazené iniciativou OpenData.cz okresům dle informačního systému RÚIAN.

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?hodnota
WHERE {
  ?observation a qb:Observation ;
    <http://data.czso.cz/ontology/refArea> ?hodnota ;
    qb:dataSet <http://data.czso.cz/resource/dataset/demography> .
}
ORDER BY ?hodnota
```

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?hodnota
WHERE {
  ?observation a qb:Observation ;
    <http://data.czso.cz/ontology/refArea> ?hodnota ;
    qb:dataSet <http://data.czso.cz/resource/dataset/social-service-facilities> .
}
ORDER BY ?hodnota
```

U ČSSZ zjistíme, že se na dimenzi určující místa vyskytují nejen okresy, ale i kraje a celá ČR. Reprezentace ale používá stejná URI, jako je tomu v případě ČSÚ. Kostky jsou tedy v dimenzi

určující místo srovnatelné.

```
PREFIX qb: <http://purl.org/linked-data/cube#>

SELECT DISTINCT ?hodnota
WHERE {
  ?observation a qb:Observation ;
    <http://linked.cssz.cz/ontology/dataset-definitions/penze#refArea> ?hodnota ;
    qb:dataSet <http://linked.cssz.cz/dataset/penze/duchodci-v-cr-krajich-okresech> .
}
ORDER BY ?hodnota
```

Krok č. 5) Z předchozích dotazů jsme zjistili, že kostka s demografií ČSÚ a důchodovou statistikou ČSSZ jsou porovnatelné na pohlaví, letech a okresech. Pokud kostky v ostatních dimenzích agregujeme, kostky můžeme zajímavě kombinovat pomocí vhodných SPARQL dotazů. V následujícím textu uvádíme ukázku 3 takových SPARQL dotazů.

- Dotaz č. 1) Kombinace ČSÚ (věkové kategorie 0, 1-4, 5-9, 10-14, 15-19) + ČSSZ (druh důchodu D - sirotčí důchody) a jejich agregace (pouze rok 2012, z ČSSZ odstraníme pohlaví T (= M+F), **tj. srovnání počtu dětí vs. počty sirotčích důchodů.** Vše setříděno dle podílu sirotčích důchodů ku populaci dětí.

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX czso: <http://data.czso.cz/ontology/>
PREFIX cssz: <http://linked.cssz.cz/ontology/dataset-definitions/penze#>
PREFIX vek: <http://linked.opendata.cz/generated/resource/age/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX druh-duchodu: <http://linked.cssz.cz/resource/pension-kind/>

SELECT ?okres ?okresNazev ?pohlavi ?rok (SUM(?czsoPopulace) AS ?czsoDetskaPopulaceCelkem) (SUM(?csszPocetDuchodcu)
AS ?csszPocetSirotcichDuchodcuCelkem) ((SUM(?csszPocetDuchodcu)/SUM(?czsoPopulace)) AS
?podilSirotkuNaDetskePopulaci)
WHERE {
SELECT DISTINCT ?okres ?okresNazev ?pohlavi ?rok ?czsoPopulace ?csszPocetDuchodcu
WHERE {
{
  ?czsoObservation a qb:Observation ;
    qb:dataSet <http://data.czso.cz/resource/dataset/demography> ;
    czso:refArea ?okres ;
    czso:pohlavi ?pohlavi ;
    czso:refPeriod ?rok .
  ?czsoObservation czso:populace ?czsoPopulace .
  ?okres skos:prefLabel ?okresNazev .
  ?czsoObservation czso:vekovakategorie ?vek .
  FILTER (?vek = vek:Y0 || ?vek = vek:Y1T4 || ?vek = vek:Y5T9 || ?vek = vek:Y10T14 || ?vek = vek:Y15T19)
}
UNION
{
  ?csszObservation a qb:Observation ;
    qb:dataSet <http://linked.cssz.cz/dataset/penze/duchodci-v-cr-krajich-okresech> ;
    cssz:refArea ?okres ;
    cssz:pohlavi ?pohlavi ;
    cssz:refPeriod ?rok ;
    cssz:pocet-duchodcu ?csszPocetDuchodcu ;
    cssz:druh-duchodu ?druhDuchodu .
```

```
   FILTER (?pohlavi != <http://purl.org/linked-data/sdmx/2009/code#sex-T>)
   FILTER (?druhDuchodu = druh-duchodu:PK_D_2008 || ?druhDuchodu = druh-duchodu:PK_D_2010)
  ?okres skos:prefLabel ?okresNazev .
}
}
}
GROUP BY ?okres ?okresNazev ?pohlavi ?rok
HAVING ( SUM(?czsoPopulace) > 0 && SUM(?csszPocetDuchodcu) > 0)
ORDER BY ?podilSirotkuNaDetskePopulaci
```

- Dotaz č. 2) Kombinace ČSÚ (věkové kategorie 20-24, …, 65-69) + ČSSZ (druh důchodu IP, ID, IT - invalidní důchody 1., 2. a 3. st.) a jejich agregace (pouze rok 2012, s ČSSZ odstraníme pohlaví T (= M+F), **tj. srovnání počtu obyvatel v produktivním věku vs. počty invalidních důchodů.** Vše setříděno dle podílu invalidních důchodů ku produktivní populaci.

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX czso: <http://data.czso.cz/ontology/>
PREFIX cssz: <http://linked.cssz.cz/ontology/dataset-definitions/penze#>
PREFIX vek: <http://linked.opendata.cz/generated/resource/age/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX druh-duchodu: <http://linked.cssz.cz/resource/pension-kind/>

SELECT ?okres ?okresNazev ?pohlavi ?rok (SUM(?czsoPopulace) AS ?czsoAktivniPopulaceCelkem) (SUM(?csszPocetDuchodcu) AS
?csszPocetInvalidnichDuchodcuCelkem) (((SUM(?csszPocetDuchodcu)/SUM(?czsoPopulace))) AS ?podilInvaliduNaAktivniPopulaci)
WHERE {
 SELECT DISTINCT ?okres ?okresNazev ?pohlavi ?rok ?czsoPopulace ?csszPocetDuchodcu
 WHERE {
 {
   ?czsoObservation a qb:Observation ;
     qb:dataSet <http://data.czso.cz/resource/dataset/demography> ;
     czso:refArea ?okres ;
     czso:pohlavi ?pohlavi ;
     czso:refPeriod ?rok .
   ?czsoObservation czso:populace ?czsoPopulace .
   ?okres skos:prefLabel ?okresNazev .
   ?czsoObservation czso:vekovakategorie ?vek .
   FILTER (?vek = vek:Y20T24 || ?vek = vek:Y25T29 || ?vek = vek:Y30T34 || ?vek = vek:Y35T39 || ?vek = vek:Y40T44 || ?vek = vek:Y45T49
|| ?vek = vek:Y50T54 || ?vek = vek:Y55T59 || ?vek = vek:Y60T64 || ?vek = vek:Y65T69)
 } UNION {
   ?csszObservation a qb:Observation ;
     qb:dataSet <http://linked.cssz.cz/dataset/penze/duchodci-v-cr-krajich-okresech> ;
     cssz:refArea ?okres ;
     cssz:pohlavi ?pohlavi ;
     cssz:refPeriod ?rok ;
     cssz:pocet-duchodcu ?csszPocetDuchodcu ;
     cssz:druh-duchodu ?druhDuchodu .
   FILTER (?pohlavi != <http://purl.org/linked-data/sdmx/2009/code#sex-T>)
   FILTER (?druhDuchodu = druh-duchodu:PK_IP_2008 || ?druhDuchodu = druh-duchodu:PK_ID_2008 || ?druhDuchodu = druh-
duchodu:PK_IT_2008 || ?druhDuchodu = druh-duchodu:PK_I_2008 || ?druhDuchodu = druh-duchodu:PK_IC_2008 || ?druhDuchodu = druh-
duchodu:PK_IP_2010 || ?druhDuchodu = druh-duchodu:PK_ID_2010 || ?druhDuchodu = druh-duchodu:PK_IT_2010 || ?druhDuchodu =
druh-duchodu:PK_I_2010 || ?druhDuchodu = druh-duchodu:PK_IC_2010)
   ?okres skos:prefLabel ?okresNazev .
 } }
}
GROUP BY ?okres ?okresNazev ?pohlavi ?rok
HAVING ( SUM(?czsoPopulace) > 0 && SUM(?csszPocetDuchodcu) > 0)
ORDER BY ?podilInvaliduNaAktivniPopulaci
```

- Dotaz č. 3) Kombinace ČSÚ sociální zařízení (počty míst v domovech pro důchodce) +

ČSSZ (druh důchodu SD) a jejich agregace (z ČSSZ odstraníme pohlaví T (= M+F), **tj. srovnání počtu míst v domovech důchodců a počtu starobních důchodů.**

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX czso: <http://data.czso.cz/ontology/>
PREFIX cssz: <http://linked.cssz.cz/ontology/dataset-definitions/penze#>
PREFIX vek: <http://linked.opendata.cz/generated/resource/age/>
PREFIX skos:    <http://www.w3.org/2004/02/skos/core#>
PREFIX druh-duchodu: <http://linked.cssz.cz/resource/pension-kind/>

SELECT ?okres ?okresNazev ?rok (SUM(?czsoPocetMist) AS ?czsoPocetMistCelkem) (SUM(?csszPocetDuchodcu) AS
?csszPocetDuchodcuCelkem) (((SUM(?czsoPocetMist)/SUM(?csszPocetDuchodcu))) AS
?podilPoctuMistVDomovechProDuchodceNaPoctuStarobnichDuchodcu)
WHERE {
 SELECT DISTINCT ?okres ?okresNazev ?rok ?czsoPocetMist ?csszPocetDuchodcu
 WHERE {
 {
  ?czsoObservation a qb:Observation ;
   qb:dataSet <http://data.czso.cz/resource/dataset/social-service-facilities> ;
   czso:refArea ?okres ;
   czso:refPeriod ?rok .
  ?okres skos:prefLabel ?okresNazev .
  ?czsoObservation czso:pocetMistVDomovechProDuchodce ?czsoPocetMist .
 }
 UNION
 {
  ?csszObservation a qb:Observation ;
   qb:dataSet <http://linked.cssz.cz/dataset/penze/duchodci-v-cr-krajich-okresech> ;
   cssz:refArea ?okres ;
   cssz:pohlavi ?pohlavi ;
   cssz:refPeriod ?rok ;
   cssz:pocet-duchodcu ?csszPocetDuchodcu ;
   cssz:druh-duchodu ?druhDuchodu .
  FILTER (?pohlavi != <http://purl.org/linked-data/sdmx/2009/code#sex-T>)
  FILTER (?druhDuchodu = druh-duchodu:PK_SD_2008 || ?druhDuchodu = druh-duchodu:PK_SD_2010)
  ?okres skos:prefLabel ?okresNazev .
 }}
}
GROUP BY ?okres ?okresNazev ?rok
HAVING ( SUM(?czsoPocetMist) > 0 && SUM(?csszPocetDuchodcu) > 0)
ORDER BY ?podilPoctuMistVDomovechProDuchodceNaPoctuStarobnichDuchodcu
```

7 Validace dat

Vytvořené datové kostky ve formátu RDF byly validovány. Výsledkem validace je zjištění, zda se ve vytvořené reprezentaci vyskytují nějaké chyby vzniklé chybným převodem zdrojových XLS dat. V rámci řešení projektu jsme validaci provedli a zajistili, že výsledná data chyby neobsahují. Popsané postupy validace umožňují validaci kdykoliv v budoucnu opakovat v případě, že bude zasahováno do ETL procedur, nebo v případě převodu dat v dalších letech.

7.1 Vizuální validace

Vizuální validací rozumíme ruční kontrolu obsahu datových kostek pohledem na různé vizualizace (sloupcové grafy apod.), pomocí které je možné odhalit a poté opravit podezřelé hodnoty vzniklé chybou při převodu (např. překlepem v šabloně pro převod dat z primární XLS reprezentace do CSV reprezentace). Pro vizuální validaci využíváme webové aplikace

vytvářené v rámci řešení projektu. Ta umožňuje zobrazovat různé řezy v dané kostce v podobě sloupcového grafu. Vizuální validace konkrétně probíhá tak, že validátor postupně prochází jednotlivé řezy jednotlivými kostkami a pohledem vyhledává podezřelé výchylky v hodnotách, např. příliš nízká nebo příliš vysoká hodnota v porovnání s okolními hodnotami.

7.2 Validace pomocí SPARQL dotazů

Dalším typem validace je automatizované testování RDF reprezentace datových kostek pomocí SPARQL dotazů. Pro účely projektu jsme definovali několik typů testů:

- **Test úplnosti definic datových kostek**: Sada SPARQL dotazů, které kontrolují, že ontologie definující strukturu jednotlivých datových kostek jsou úplné.
- **Testování pokrytí číselníků definujících možné hodnoty dimenzí datovými kostkami**: Sada SPARQL dotazů, které kontrolují, že pro danou množinu hodnot je využita každá hodnota v alespoň jednom faktu nějaké datové kostky.
- **Testování součtů**: Sada SPARQL dotazů, které kontrolují, že fakt umístěný v dimenzi s hierarchickou strukturou na nelistový uzel odpovídá součtu fakt umístěných v té samé dimenzi na nižších uzlech v hierarchii. (Pozn.: Lze aplikovat pouze pro měření, jejichž hodnoty se dají sčítat).

Veškeré připravené testovací SPARQL dotazy jsou seřazeny v podobě pipeline v nástroji UnifiedViews, který používáme i pro samotnou přípravu dat. Tuto pipeline nazýváme validační pipeline. Kdykoliv je možné ji spustit a provést tak validaci dat pomocí automatizovaných testovacích SPARQL dotazů. Export pipeline je pak přílohou k tomuto dokumentu (Příloha 1).

V projektu jsme zajistili, že vytvořená RDF reprezentace datových kostek ČSSZ je z pohledu vytvořené testovací pipeline bezchybná. Jinými slovy, testovací pipeline nevrací žádná chybová hlášení.

7.2.1 Testování úplnosti definic datových kostek

Struktura obsahu datových kostek je definována ontologií vytvořenou dle standardu Data Cube Vocabulary stručně popsaného na začátku tohoto dokumentu. V rámci automatizovaného testování kontrolujeme správnost a úplnost této definiční ontologie. Např. kontrolujeme, že je dán název a popis každé datové kostky, jejích dimenzí a měření a že každá datové kostky má alespoň jednu dimenzi a alespoň jedno měření.

Zavádíme např. následující testovací SPARQL dotazy (plný výčet by obsahoval celkem 11 testovacích dotazů, ale pro jednoduchost uvádíme jen vybrané). Poznamenejme, že všechny využívají stejné prefixy:

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX report: <http://data.unifiedviews.eu/ontology/quality-report/>
```

Definice struktury datové kostky by	CONSTRUCT { ?report a report:Report ; foaf:primaryTopic ?cubeStructure ;

měla mít název.	report:warning _:m .
	
	_:m a report:Message ;
	dcterms:description ?warningCubeStructureLabel ;
	report:missingStatement _:ms .
	
	_:ms a rdf:Statement ;
	rdf:subject ?cubeStructure ;
	rdf:predicate rdfs:label ;
	rdf:object _:mso .
	
	_:mso a rdfs:Literal .
	
	} WHERE {
	?cubeStructure a qb:DataStructureDefinition .
	FILTER NOT EXISTS {
	?cubeStructure rdfs:label ?cubeStructureLabel .
	FILTER(isLiteral(?cubeStructureLabel) = true)
	}
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(STR(?cubeStructure)))) AS ?report)
	BIND("The data structure definition should have a label. The label should be assigned by rdfs:label. The label should be a rdfs:Literal."^^xsd:string AS ?warningCubeStructureLabel)
	
	}

Definice struktury datové kostky by měla definovat alespoň jednu dimenzi.	CONSTRUCT {
	?report a report:Report ;
	foaf:primaryTopic ?cubeStructure ;
	report:warning _:m .
	
	_:m a report:Message ;
	dcterms:description ?warningCubeStructureDimension ;
	report:missingStatement _:ms1 ;
	report:missingStatement _:ms2 .
	
	_:ms1 a rdf:Statement ;
	rdf:subject ?cubeStructure ;
	rdf:predicate qb:component ;
	rdf:object _:ms1o .
	
	_:ms2 a rdf:Statement ;
	rdf:subject _:ms1o ;
	rdf:predicate qb:dimension ;
	rdf:object _:ms2o .
	
	} WHERE {
	?cubeStructure a qb:DataStructureDefinition .
	FILTER EXISTS {
	?cubeStructure qb:component ?somecomponent .
	}
	FILTER NOT EXISTS {
	?cubeStructure qb:component ?anycomponent .
	FILTER EXISTS {
	?anycomponent qb:dimension ?dimension .
	}
	}
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(STR(?cubeStructure)))) AS ?report)
	BIND("The data structure definition should have at least one component which is a dimension. The dimension should be associated with the component by qb:dimension property."^^xsd:string AS ?warningCubeStructureDimension)
	}

Definice struktury datové kostky by měla definovat alespoň jedno měření.	CONSTRUCT {
	?report a report:Report ;
	foaf:primaryTopic ?cubeStructure ;
	report:warning _:m .
	
	_:m a report:Message ;
	dcterms:description ?warning ;
	report:missingStatement _:ms1 ;
	report:missingStatement _:ms2 .
	
	_:ms1 a rdf:Statement ;
	rdf:subject ?cubeStructure ;
	rdf:predicate qb:component ;
	rdf:object _:ms1o .
	
	_:ms2 a rdf:Statement ;
	rdf:subject _:ms1o ;
	rdf:predicate qb:measure ;
	rdf:object _:ms2o .
	
	} WHERE {
	?cubeStructure a qb:DataStructureDefinition .
	FILTER EXISTS {
	?cubeStructure qb:component ?somecomponent .
	}
	FILTER NOT EXISTS {
	?cubeStructure qb:component ?anycomponent .
	FILTER EXISTS {
	?anycomponent qb:measure ?dimension .
	}
	}
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(STR(?cubeStructure)))) AS ?report)
	BIND("The data structure definition should have at least one component which is a measure. The measure should be associated with the component by qb:measure property."^^xsd:string AS ?warning)
	}

Každá dimenze a měření definované v definici struktury kostky by měla být navázána na alespoň jeden koncept popisující její sémantiku.	CONSTRUCT {
	?report a report:Report ;
	foaf:primaryTopic ?dimensionOrMeasure ;
	report:warning _:m .
	
	_:m a report:Message ;
	dcterms:description ?warning ;
	report:missingStatement _:ms1, _:ms2 .
	
	_:ms1 a rdf:Statement ;
	rdf:subject ?dimensionOrMeasure ;
	rdf:predicate qb:concept .
	
	_:ms2 a rdf:Statement ;
	rdf:subject ?dimensionOrMeasure ;
	rdf:predicate rdfs:subPropertyOf .
	
	} WHERE {
	VALUES ?type { qb:DimensionProperty qb:MeasureProperty }
	?dimensionOrMeasure a ?type .
	FILTER NOT EXISTS {
	{ ?dimensionOrMeasure qb:concept ?concept . }
	UNION
	{ ?dimensionOrMeasure rdfs:subPropertyOf ?parent . }
	}
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(STR(?dimensionOrMeasure)))) AS ?report)

	BIND(IF(?type = qb:DimensionProperty, "The dimension should have at least one concept or it should be a subproperty of another property which specifies its semantics. The concept should be assigned by qb:concept. The parent property shoulb be assigned by rdfs:subPropertyOf."^^xsd:string, "The dimension should have at least one concept or it should be a subproperty of another property which specifies its semantics. The concept should be assigned by qb:concept. The parent property shoulb be assigned by rdfs:subPropertyOf."^^xsd:string) AS ?warning) }

7.2.2 Testování pokrytí číselníků

Ontologie modelované jako SKOS Concept Scheme představují číselníky. Bylo by vhodné formulovat SPARQL dotaz, který ukáže, ve kterých datových kostkách nejsou uplatněny všechny položky číselníků, které jsou v dané datové kostce použity. Výsledky je pak třeba prověřit a určit, kde skutečně došlo k chybě při převodu dat, a kdy není celý číselník využit již v podkladových datech. Např. číselník druhů důchodů obsahuje mimo jiné jednotlivé kombinace vdovského či vdoveckého důchodu s jednotlivými druhy starobního důchodu, nicméně nemusí být ve všech případech jednotlivé kombinace uvedeny v podkladových datech.

Následující SPARQL dotaz vrací trojice (datová kostka K, dimenze D, hodnota dimenze H) takové, že datová kostka K obsahuje měření namapované na dimenzi D, ale neexistuje v ní jediný fakt, který by v dimenzi D měl hodnotu H.

```
PREFIX qb: <http://purl.org/linked-data/cube#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?kostka ?dimenze ?hodnotaDimenze
FROM <http://linked.opendata.cz/resource/dataset/cssz/pensions/def>
FROM <http://linked.opendata.cz/resource/dataset/cssz/pensions>
WHERE {
  ?kostka a qb:DataSet ;
       qb:structure ?definiceKostky .

  ?definiceKostky qb:component/qb:dimension ?dimenze .

  ?dimenze rdfs:range ?typHodnotDimenze.

  ?hodnotaDimenze a ?typHodnotDimenze .

  FILTER NOT EXISTS {

    ?pozorovani a qb:Observation ;
      qb:dataSet ?kostka ;
      ?dimenze ?hodnotaDimenze .
  }

}
```

Trojice vrácené výše uvedeným SPARQL dotazem říkají, že v datové kostce K se nevyskytuje fakt namapovaný v dimenzi D na hodnotu H. Např. trojice (K, okresy, Semily) by znamenala, že

v datové kostce K neexistuje fakt pro okres Semily. Tzn., že okres Semily je v datové kostce K opominut a je nutné zkontrolovat, zda je to záměr nebo chyba.

7.2.3 Testování součtů

V následujících ontologií reprezentovaných jako SKOS Concept Scheme je definována hierarchie konceptů:

- ČSSZ
  - druhy důchodů - `pen-onto:PensionKindScheme_2008` a `pen-onto:PensionKindScheme_2010`
    - `pen-onto:PensionKindScheme` hierarchii definovanou nemá, ale obsahuje definici jednotlivých druhů důchodu nezávislou na roku. Druhy důchodů ve schématech `pen-onto:PensionKindScheme_2008` a `pen-onto:PensionKindScheme_2010` obsahují hierarchii, která platila v daných letech. Přes `skos:exactMatch` jsou druhy důchodů závislé na letech napojeny na druhy důchodů, jejichž vymezení na letech závislé není. Např. `pen:PK_S` představuje starobní důchod typu S podle příslušných paragrafů zákona. Protože byl vyplácen jak v letech 2008 a 2009, tak i v letech 2010-2012, jsou v příslušných schématech definovány druhy důchodů `pen:PK_S_2008` a `pen:PK_S_2010`. Oba tyto druhy důchodů jsou napojeny na druh důchodu `pen:PK_S` pomocí predikátu `skos:exactMatch`.
  - kapitoly MKN10 - `icd10-onto:ICD10ChaptersScheme`
    - pouze 2 úrovně - `icd10:C_T` (všechny kapitoly) je nadřazený koncept všech jednotlivých kapitol

Hierarchie definované mezi koncepty lze využít pro kontrolu dat. Podkladová data zpravidla obsahují jak data pro koncepty na nejnižší úrovni hierarchie (např. jednotlivé druhy starobních důchodů), tak i data součtová (např. součet přes všechny druhy starobních důchodů, součet za muže a ženy apod.). Díky tomu lze provést následující kontrolu:

- spočítat součet přes jednotlivé uzly na nižší úrovni hierarchie a porovnat s uloženou hodnotou, která představuje součet přes tyto uzly. Např. součet přes druh vdovský důchod kombinovaný se starobním důchodem (pro roky 2008 a 2009 druhy důchodu `pen:PK_SV_2008, pen:PK_SRNV_2008, pen:PK_STV_2008, pen:PK_SDV_2008, pen:PK_SRV_2008, pen:PK_IV_2008, pen:PK_ICV_2008`) a porovnat součet s hodnotou, kde hodnotou dimenze Druh důchodu je druh důchodu představující sumu za všechny kombinované vdovské důchody (`pen:PK_V-KOMB_2008`)

Pokud je vypočtený a uložený součet shodný, data jsou v pořádku. Pokud si součty neodpovídají, existuje chyba v převedených datech nebo je chyba obsažena již v podkladových datech.

Konkrétní postup v tomto případě testu uvedeme na příkladu datové kostky http://linked.cssz.cz/dataset/penze/duchodci-v-krajich, dimenze

http://linked.cssz.cz/ontology/dataset-definitions/penze#druh-duchodu a měření http://linked.cssz.cz/ontology/dataset-definitions/penze#pocet-duchodcu. Kontrolujeme, zda počty důchodců uvedené pro typy důchodů, které jsou součtem počtů důchodců v jiných podřazených typech důchodů, tímto součtem opravdu jsou. Bohužel se ukázalo, že je velmi obtížné až nemožné vytvořit jeden SPARQL dotaz provádějící takový test (z důvodů omezení současných úložišť RDF dat). Proto jsme přistoupili k rozepsání do několika SPARQL dotazů, které řadíme v testovací pipeline v nástroji UnifiedViews do sekvence. Sekvence dotazů je následující:

Krok 1: Získání faktů z testované datové kostky a hodnot testovaného měření a jejich seřazení do hierarchie dané testovanou dimenzí. Výsledkem je pro každý agregovaný fakt množina faktů, které jsou agregovány.	PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
	PREFIX qb: <http://purl.org/linked-data/cube#>
	PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
	
	PREFIX aux: <http://data.unifiedviews.eu/ontology/auxiliary/>
	
	PREFIX tested-dimension: <http://linked.cssz.cz/ontology/dataset-definitions/penze#druh-duchodu>
	PREFIX tested-measure: <http://linked.cssz.cz/ontology/dataset-definitions/penze#pocet-duchodcu>
	PREFIX cube: <http://linked.cssz.cz/dataset/penze/duchodci-v-krajich>
	
	PREFIX cssz-dsd: <http://linked.cssz.cz/ontology/dataset-definitions/penze#>
	
	CONSTRUCT {
	?observationP qb:dataSet cube: ;
	tested-dimension: ?conceptP ;
	tested-measure: ?valuePD ;
	aux:tested-dimension tested-dimension: ;
	aux:tested-measure tested-measure: ;
	cssz-dsd:pohlavi ?pohlavi ;
	cssz-dsd:refArea ?refArea ;
	cssz-dsd:refPeriod ?refPeriod .
	
	?observationP aux:narrower ?observationD .
	
	?observationD qb:dataSet cube: ;
	tested-dimension: ?conceptD ;
	tested-measure: ?valueDD ;
	aux:tested-dimension tested-dimension: ;
	aux:tested-measure tested-measure: ;
	cssz-dsd:pohlavi ?pohlavi ;
	cssz-dsd:refArea ?refArea ;
	cssz-dsd:refPeriod ?refPeriod .
	
	cssz-dsd:pohlavi a aux:OtherDimension .
	cssz-dsd:refArea a aux:OtherDimension .
	cssz-dsd:refPeriod a aux:OtherDimension .
	}
	WHERE {
	
	?observationP qb:dataSet cube: ;
	tested-dimension: ?conceptP ;
	tested-measure: ?valueP ;
	cssz-dsd:pohlavi ?pohlavi ;
	cssz-dsd:refArea ?refArea ;
	cssz-dsd:refPeriod ?refPeriod .
	
	?conceptP skos:narrower+ ?conceptD .
	
	?observationD qb:dataSet cube: ;

	tested-dimension: ?conceptD ; tested-measure: ?valueD ; cssz-dsd:pohlavi ?pohlavi ; cssz-dsd:refArea ?refArea ; cssz-dsd:refPeriod ?refPeriod . BIND(xsd:decimal(?valueP) AS ?valuePD) BIND(xsd:decimal(?valueD) AS ?valueDD) }

Krok 2: Pro každý agregovaný fakt vyjmeme z množiny faktů tvořících agregaci nadbytečná fakta. Fakt je nadbytečný, pokud na cestě k agregovanému faktu leží jiný fakt. Nadbytečný fakt tak nemůže přispívat do součtu, neboť již jednou přispívá prostřednictvím nalezeného jiného faktu. Proto je nutné jej vyjmout.	PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
	
	PREFIX aux: <http://data.unifiedviews.eu/ontology/auxiliary/>
	
	DELETE {
	?observationA aux:narrower ?observationD .
	} WHERE {
	
	?observationA aux:narrower ?observationD ;
	aux:tested-dimension ?testedDimension .
	
	?observationA ?testedDimension ?conceptA .
	?observationD ?testedDimension ?conceptD .
	
	?observationI ?testedDimension ?conceptI .
	?conceptA skos:narrower+ ?conceptI .
	?conceptI skos:narrower+ ?conceptD .
	
	}

Krok 3: Spočítání hodnot agregovaných faktů agregací množiny faktů tvořících agregaci.	PREFIX qb: <http://purl.org/linked-data/cube#>
	
	PREFIX aux: <http://data.unifiedviews.eu/ontology/auxiliary/>
	
	CONSTRUCT {
	?observationA aux:value ?valueA ;
	aux:aggValue ?aggValue ;
	qb:dataSet ?cube ;
	aux:concept ?conceptA ;
	aux:tested-measure ?testedMeasure ;
	aux:tested-dimension ?testedDimension ;
	?otherDimensionA ?otherDimensionValueA .
	?otherDimensionA a aux:OtherDimension .
	} WHERE {
	
	?observationA qb:dataSet ?cube ;
	aux:tested-measure ?testedMeasure ;
	aux:tested-dimension ?testedDimension ;
	?testedDimension ?conceptA ;
	?testedMeasure ?valueA ;
	?otherDimensionA ?otherDimensionValueA .
	
	?otherDimensionA a aux:OtherDimension .
	
	{
	SELECT ?observationA ?testedMeasure ?testedDimension (SUM(?valueD) AS ?aggValue)
	WHERE {
	?observationA aux:narrower ?observationD .
	?observationD ?testedMeasure ?valueD ;
	?testedDimension ?conceptD .

	} GROUP BY ?observationA ?testedMeasure ?testedDimension } }

Krok 4: Porovnání spočítaných hodnot agregovaných faktů se skutečně uvedenými hodnotami.	PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
	PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
	PREFIX foaf: <http://xmlns.com/foaf/>
	PREFIX dcterms: <http://purl.org/dc/terms/>
	PREFIX report: <http://data.unifiedviews.eu/ontology/quality-report/>
	
	PREFIX qb: <http://purl.org/linked-data/cube#>
	
	PREFIX aux: <http://data.unifiedviews.eu/ontology/auxiliary/>
	
	CONSTRUCT {
	?report a report:Report ;
	foaf:primaryTopic ?cube, ?testedMeasure, ?testedDimension ;
	report:warning ?message .
	
	?message a report:Message ;
	foaf:primaryTopic ?observation ;
	dcterms:description ?warning ;
	report:problemStatement ?ps ;
	report:missingStatement ?ms1 .
	
	?ps a rdf:Statement ;
	rdf:subject ?observation ;
	rdf:predicate ?testedMeasure ;
	rdf:object ?value .
	
	?ms1 a rdf:Statement ;
	rdf:subject ?observation ;
	rdf:predicate ?testedMeasure ;
	rdf:object ?aggValue .
	
	} WHERE {
	?observation aux:value ?value ;
	aux:aggValue ?aggValue ;
	qb:dataSet ?cube ;
	aux:tested-measure ?testedMeasure ;
	aux:tested-dimension ?testedDimension ;
	?otherDimension ?otherDimensionValue .
	
	?otherDimension a aux:OtherDimension .
	
	FILTER (?value != ?aggValue)
	
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(CONCAT(STR(?cube), STR(?testedMeasure), STR(?testedDimension))))) AS ?report)
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(CONCAT(STR(?cube), STR(?testedMeasure), STR(?testedDimension))), "/", MD5(STR(?observation)), "/test-of-hierarchical-measured-values")) AS ?message)
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(CONCAT(STR(?cube), STR(?testedMeasure), STR(?testedDimension))), "/", MD5(STR(?observation)), "/test-of-hierarchical-measured-values/problem-statement")) AS ?ps)
	BIND(IRI(CONCAT("http://data.unifiedviews.eu/resource/quality-report/", MD5(CONCAT(STR(?cube), STR(?testedMeasure), STR(?testedDimension))), "/", MD5(STR(?observation)), "/test-of-hierarchical-measured-values/missing-statement-1")) AS ?ms1)
	}

8 Zdroje

Cyganiak, Richard, Reynolds, Dave, 2014. The RDF Data Cube Vocabulary. In: *W3C* [online]. 16 January 2014 [cit. 2014-09-25]. Dostupné z: http://www.w3.org/TR/vocab-data-cube/.

Harris, Steve, Seaborne, Andy, 2013. SPARQL 1.1 Query Language. In: *W3C* [online]. 16 January 2014 [cit. 2013-03-21]. Dostupné z: http://www.w3.org/TR/sparql11-query/.

9 Příloha 1 - Validační pipeline

K tomuto dokumentu náleží následující příloha: **Validační pipeline (export)**. Příloha je ke stažení na adrese: http://opendata.vse.cz/cssz/cssz-validacni-pipeline.zip. Užití validační pipeline podléhá licenci GNU GPL v3.