



Informačný model pre zostavenie štruktúrovaných popisov štatistických údajov

2. časť

Ján Švolík
Národná banka Slovenska

V predchádzajúcej časti sme naznačili proces, ako efektívne popísať, čo znamenajú, resp. obsahujú štatistické údaje. Na tento účel sa vytvorilo viacero systémov, ako spájať údaje s ich popismi a ako s nimi ďalej pracovať. Ako príklad sme uviedli dva systémy na prenos údajov, GESMES/TS a XBRL. Tieto systémy vznikli z potreby efektívneho prenosu údajov medzi inštitúciami. Ich informačné modely (hlavne GESMES/TS) boli inšpiráciou pri tvorbe zásad informačného modelu NBS. V tejto časti sa sústredíme na popis princípov tohto modelu, načrtneme proces budovania dátového modelu NBS a zhodnotíme, čo a ako sa podarilo premietnuť do štatistického zberového portálu.

INFORMAČNÝ MODEL NA POPIS ŠTATISTICKÝCH ÚDAJOV V NBS

Pôvodný zámer zostavenia systému na identifikáciu údajov pomocou jednoduchých kódov sa ukázal ako neefektívny a postupne sa rozvinul do súčasnej podoby štruktúrovaného popisu. Za týmto popisom je informačný model, ktorý určuje, ako majú byť informácie o údajoch usporiadané, aké môžu nadobúdať hodnoty a ako sa má s nimi pracovať.

Ideový návrh modelu

Keď sa pozrieme na tab. 2 v predchádzajúcej časti, vidíme, že jednotlivé údaje v nej nie sú v prázdnom priestore, ale majú svojich susedov a navzájom spolu súvisia. Toto platí pre väčšinu údajov, ktoré sú predmetom nášho záujmu. Ukazuje sa, že údaje v jednej tabuľke často majú podobný obsah a zhodnú aspoň časť popisu. Ak s týmito údajmi chceme pracovať nezávisle, treba nájsť spôsob, ako ich efektívne uložiť a pripojiť k nim ich popisy. Spôsob popisovania údajov je vhodné navrhnúť tak, aby bol pre podobnú, resp. príbuznú skupinu údajov unifikovaný a súčasne umožnil zachytiť ich jedinečnosti a vzájomné súvislosti.

Skôr ako prejdeme k popisu informačného modelu pre štatistické údaje, ako aj pre údaje na účely dohľadu v NBS, vymedzíme si niektoré pojmy, s ktorými sa v tomto modeli budeme stretávať.

Predmet je nejaký nositeľ informácií (osoba, cenný papier, banka atď.), t. j. objekt, o ktorom sa zisťujú informácie.

Vlastnosť je pomenovaná charakteristika (objem, počet, splatnosť, dátum vyrovnania, názov, adresa atď.) alebo druh informácie použitej na identifikáciu predmetu (napr. meno, priezvisko, IČO, ISIN). Kvalitatívne a kvantitatívne vyjadrenie vlastnosti je jej konkrétnou hodnotou (číslo alebo text).

Pri práci s údajmi sa popisujú hodnoty vlastností jednotlivých samostatných predmetov a rôzne vlastnosti skupiny predmetov.

Individuálne údaje sú také údaje, ktorých hodnoty popisujú vlastnosti jedného predmetu (napr. cenný papier: názov, ISIN, menovitá hodnota).

Agregované údaje sú také údaje, pri ktorých sa hodnota vytvára (procesom agregácie) ako jediný údaj z individuálnych údajov určitej množiny predmetov. Agregované údaje reprezentujú nejakú vlastnosť celej množiny predmetov (napr. priemerný ročný výnos všetkých cenných papierov v držbe určitej banky alebo s určitou dobou splatnosti).

Aby sme nemuseli pre každý predmet definovať spôsob, ako s ním pracovať, zaviedli sme také usporiadanie (zoskupenie) predmetov, pri ktorom je možné s údajmi všetkých predmetov pracovať rovnakým spôsobom. **Triedou** budeme nazývať zoskupenie predmetov, pri ktorých je možné používať rovnaký spôsob popisu ich vlastností a vykonávať s nimi rovnaké operácie. Triedu teda tvoria také predmety, ktoré majú niektoré spoločné vlastnosti (podobnú štruktúru zisťovaných vlastností, ako napr. splatnosť, sektor emitenta), vstupujú do podobných vzťahov s inými predmetmi (vlastník) alebo sa s nimi spájajú rovnaké udalosti (nákup, predaj, splatnosť). Predmety, ktoré sa nachádzajú v triede, sa nazývajú objekty.

Individuálne triedy (prirodzené) sú také triedy, v ktorých objekty majú zodpovedajúce samostatné predmety v reálnom svete. V týchto triedach sa zaujímate o vlastnosti za každý predmet zvlášť. Každá vlastnosť objektu je vyjadrená svojou hodnotou, čo môže byť číselný údaj, text alebo dátum. Pri číselných vyjadreniach vlastností je nutné poznať aj jednotky, v ktorých sú vyjadrené. Vznik a zánik objektov v individuálnych triedach je



zviazaný so vznikom a zánikom predmetov, ktoré reprezentujú.

Často je potrebné zovšeobecniť informácie jednotlivých predmetov a študovať ich vplyv na väčšie celky. Vtedy nás nezaujímajú (a ani nie sú potrebné) detailné informácie o jednotlivých reálnych predmetoch, ale iba súhrnné informácie za určité zoskupenie predmetov. Na tento účel slúžia **agregované triedy**, ktorých objektmi sú zoskupenia predmetov. Jednotlivé objekty reprezentujú všetky predmety vyhovujúce kritériám vyplývajúcim z analytických potrieb. Objektom v týchto triedach je iba formálne združenie predmetov a ako celok nemusí mať zodpovedajúci obraz v reálnom svete. Vlastnosťami objektov sú agregované informácie, ktoré sú určené vždy za celé zoskupenie a nie za individuálne predmety tvoriace zoskupenie (hoci z nich vznikli).

Agregácia je proces, ktorým sa z vlastností individuálnych predmetov získavajú vlastnosti objektov agregovanej triedy. Objektom agregovanej triedy je zoskupenie individuálnych predmetov vyhovujúcich daným výberovým kritériám. Vlastnosti individuálnych predmetov, ktoré sa použijú ako výberové kritériá, sa transformujú na identifikačné informácie agregovaného objektu. Ďalšie vlastnosti sú tie, ktoré potrebujeme analyzovať. Pre tieto vlastnosti je nutné stanoviť agregáciu funkciu (spôsob agregácie, sumu, priemer, výber atď.) a po agregácii sa stávajú vlastnosťami objektov agregovanej triedy. Zvyšné vlastnosti nás buď nezaujímajú, alebo nie sú vhodné na agregáciu (napr. tie, ktoré obsahujú textové informácie), a tieto zvyčajne v procese agregácie zanikajú, teda nemajú žiadne vyjadrenie v agregovanej triede.

Dôvodom rozdelenia tried na individuálne a agregované je rozdielny spôsob práce s prvkami týchto tried. Individuálne triedy majú charakter registrov a žurnálu udalostí, kde sú hlavnými činnosťami vyhladávanie a študovanie vzájomných súvislostí vlastností jednotlivých objektov. Agregované triedy majú charakter tzv. hyperkociek, kde sa sledujú rôzne pohľady a spôsoby agregácie.

Účelom tried je efektívne získavanie a spracovanie informácií o jednotlivých objektoch. Na to, aby sme sa k nejakej informácii v triede dostali, potrebujeme vybrať objekt a vlastnosť, s ktorou budeme pracovať. Na efektívny a intuitívny výber informácie má veľký vplyv, akým spôsobom je organizovaný metadátový popis. Naša voľba padla na štruktúrovaný popis údajov. Štruktúrovaný v tom zmysle, že sa celý popis rozdelí na samostatné, relatívne nezávislé časti, ktoré budeme nazývať dimenzie a detaily. **Dimenzie** slúžia na identifikáciu predmetu a vlastností, ktorej sa daný údaj týka. **Detaily** dávajú doplnkové informácie o danom údaji (jednotka, násobiteľ, spôsob merania, povinnosť, dostupnosť atď.). Pre každú triedu je určené, aké dimenzie a aké detaily v nej možno používať. Pre každú dimenziu v triede je stanovený zoznam hodnôt, ktoré môže nadobúdať.

Zoznamy prípustných hodnôt sú za jednotlivé oblasti zoskupené do **číselníkov** alebo **regist-**

trov, kde sú im tiež priradené identifikačné kódy. Dimenzia môže nadobúdať hodnoty iba z jedného číselníka alebo jedného registra. V triedach sa nepoužívajú priamo prvky číselníkov a registrov, ale len im pridelené kódy. V ďalšom texte, keď budeme hovoriť o číselníkoch, dá sa to primerane uplatniť aj na registre, aj keď nie sú explicitne menované. Prvky číselníka môžu vyjadrovať hodnotu vlastnosti na rôznej úrovni rozlíšenia, resp. presnosti. Prvky číselníka, ktoré predstavujú najjemnejší detail (najvyššie rozlíšenie) hodnoty, budeme nazývať koncové prvky. Hodnota vlastnosti môže byť priradená maximálne jednému koncovému prvku. Je vhodné, aby všetky hodnoty vlastnosti boli pokryté koncovými prvkami.

Vzájomné súvislosti medzi jednotlivými prvkami číselníka vyjadrujeme pomocou hierarchií a zoskupení.

Zoskupenie je podmnožina prvkov číselníka. Je to jednoduchý zoznam, ktorý nemá žiadnu vnútornú štruktúru. So zoskupením sa navonok pracuje rovnako ako s prvkom číselníka. **Hierarchia** popisuje orientovaný vzťah podradenosti a nadradenosti medzi prvkami číselníka. Hierarchie môžu obsahovať mnoho úrovní týchto vzťahov. Hierarchia sa zobrazuje ako stromová štruktúra.

V agregovaných triedach v jednom čase pracujeme buď s rôznymi vlastnosťami toho istého objektu (podobne ako v prípade individuálnych objektov), alebo porovnávame tú istú vlastnosť za rôzne objekty, ktoré sa líšia jednou alebo dvoma dimenziami. Príklad objektov, ktoré sa líšia v dvoch dimenziách, je v tabuľke 6, kde sledujeme, ako sa mení určitá vlastnosť (v tomto prípade hodnota úverov) podľa jednotlivých sektorov a zároveň aj podľa meny.

Tabuľka 6 Ilustrácia dimenzií

		Dimenzia mena		
		SKK	EUR	OCM
Dimenzia sektor	SPOLU (r. 3+7+11+15+19+24 až 26)	19 371 842	3 659 545	152 880
	SPOLU (všetky sektory)	19 498 260	3 659 545	152 880
	S.11 Nefinančné spoločnosti	15 431 484	3 596 421	152 822
	S.11001 verejné	1 072 822	575 643	2

Aby sa zabezpečila jednotnosť a jednoznačnosť používaných pojmov, je vytvorený spoločný zoznam vlastností, spoločná množina číselníkov a registrov, z ktorých sa zostavujú jednotlivé triedy. **Definícia dátového modelu** (zjednodušene iba dátový model) je súhrn definícií všetkých vlastností, číselníkov s ich hierarchiami a štruktúrou konkrétnych tried.

Tvorba dátového modelu na popis údajov v NBS

Informačný model nám vytvoril rámec (pravidlá) pre budovanie dátového modelu, t. j. naplnenie zoznamov vlastností, zostavenie číselníkov s de-



finovaním ich vnútornej štruktúry pre jednotlivé oblasti a definovanie tried na základe konkrétnych predmetov skúmania, analytických potrieb NBS, ako aj existujúcich údajov, ktoré sa zisťovali v minulých obdobiach. Ideálne by bolo začínať na zelenej lúke a podľa potrieb zostaviť dátový model. Lenže realita je trochu iná a tvorba dátového modelu je prispôbená súčasnému stavu zbieraných údajov. Na to, aby sme zostavili dátový model, bola potrebná rozsiahla analýza štruktúry týchto údajov, zjednotenie metodík a názvoslovia. Snažili sme sa identifikovať, čo a ako, zo súčasného zberu údajov, sa zabuduje do dátového modelu. Po vyskúšaní rôznych spôsobov a cez mnohé slepé uličky sa našiel schodný postup, ktorým je možné premietnuť súčasný stav do funkčného dátového modelu. Ďalej si ukážeme postupnosť krokov pri budovaní dátového modelu NBS.

Vytvorenie zoznamu vlastností

Prvým krokom je vytvorenie zoznamu vlastností, t. j. všetkých možných charakteristík, ktoré zisťujeme na predmetoch skúmania alebo ktoré slúžia na ich identifikáciu. Existuje niekoľko zásad, ako zostavovať tento zoznam. Spomenieme niektoré z nich:

- *minimálnosť* – zoznam by mal obsahovať minimálny počet vlastností, ktoré sú nutné na popis údajov. Zlučovať, resp. zosúlaďovať vlastnosti, ktoré majú veľmi podobný obsah a líšia sa len veľmi jemnými rozdielmi. Tieto rozdiely riešiť jedinou vlastnosťou odkazujúcou na špecifiká;
- *atomárnosť* – každá vlastnosť by mala vyjadrovať atomárnu informáciu. Teda takú informáciu, ktorá už ďalej nie je vnútorne členená. Napr. výber adresy ako vlastnosti nie je vhodný, lebo obsahuje viacero atomárnych informácií, napr. PSČ, mesto, ulicu, číslo atď. Netreba si to myliť s hierarchiou, kde ide o tú istú informáciu na rôznej úrovni podrobnosti;
- *spôsob kódovania* – pridelovanie kódov musí zohľadňovať dve kritériá: kódy by mali byť čo najkratšie a súčasne určitým spôsobom čitateľné pre používateľov. Mali by preto obsahovať nejakú minimálnu mnemotechniku obmedzenú na niekoľko prvých písmen. Vo všeobecnosti by vlastnosti, ktoré sa používajú častejšie, mali mať kratšie kódy ako tie zriedkavejšie. Príbuzné vlastnosti, ktoré zachytávajú podobné vlastnosti, je vhodné zoskupovať tak, aby mali podobný kód;
- *samostatnosť* – vlastnosť by nemala obsahovať reláciu na iný predmet alebo vlastnosť, aby zbytočne nenarastal počet vlastností, ktoré prakticky označujú tú istú charakteristiku, len z rôznych pohľadov. Takúto vlastnosť treba rozdeliť na dve, kde prvá je samotná vlastnosť a druhá je relácia. Typickým príkladom je sektor, kde môžeme definovať jednotlivé sektory (napr. sektor emitenta, sektor držiteľa atď.) ako samostatné vlastnosti. Namiesto toho je vhodné zaviesť iba jedinou vlastnosť (sektor) a to, koho sa to týka, uložiť do inej vlastnosti;
- *stálosť* – zoznam vlastností tvorí chrbtovú kosť dátového modelu, a preto by sa mal tvoriť tak,

aby nevyžadoval časté zmeny definícií a nahrádzanie už existujúcich vlastností;

- *porovnatelnosť* – zisťované vlastnosti by mali byť porovnateľné, resp. harmonizované. Mali by zachytávať tie isté charakteristiky ako iné národné a medzinárodné inštitúcie. Vyžaduje si to odstránenie miestnych rozdielov. Pri definovaní zoznamu vlastností sa k ním pripájajú doplnkové informácie o tom, aké hodnoty môže tá-ktorá vlastnosť nadobúdať. Hodnoty môžu byť:
 - kvantitatívne – hodnota je vyjadrená číslom v definovaných jednotkách merania a pre ne je nutné definovať jednotky a formát čísla (patrí sem aj vyjadrenie dátumu a času);
 - popisné – hodnotu tvorí ľubovoľný text, ktorým sa definuje jej kvalita, t. j. špecifické uplatnenie pri danom predmete. Zvyčajne stačí určiť maximálny rozsah popisného textu;
 - kvalitatívne – hodnoty sa vyberajú z nejakej množiny preddefinovaných hodnôt na základe kvalitatívnych alebo kvantitatívnych kritérií. Pre hodnoty týchto vlastností sa používajú zoznamy prípustných hodnôt (číselníky).

Vytvorenie číselníkov

K vlastnostiam, ktoré nadobúdajú (alebo sa prevádzajú) konečný počet stavov, sa vytvorí zoznamy týchto stavov (obory hodnôt). Jednotlivé zoznamy, ktoré zachytávajú určitý charakter vlastností, sa nazývajú číselníky. **Číselník** predstavuje obor hodnôt pre tieto vlastnosti s definovaným počtom stavov. To znamená, že pre vlastnosti, ktoré majú rovnaký obor hodnôt, sa bude používať ten istý číselník.

Dobré zostavenie číselníka, aj keď sa to na prvý pohľad nezdá, je pomerne komplikovaná záležitosť. Pri tomto procese nás čaká množstvo nástrah, ktoré bude treba prekonať. Na definovanie rôznych prvkov číselníka sa používajú kvalitatívne kritériá, ktoré definujú rôzny stupeň detailnosti a presnosti, čo môže viesť k tomu, že niektoré prvky sa prekrývajú, jeden prvok úplne pohltí iný prvok alebo celá množina prvkov nepokrýva všetky možné stavy danej vlastnosti.

Typické situácie, ktoré treba riešiť pri zostavovaní číselníka:

- nevhodný výber vlastnosti
 - vlastnosť nie je atomárna – potom sa na výber stavov začne uplatňovať niekoľko rôznych kritérií, čím môže počet prvkov neprimerane narastať, duplikujú sa definície a hrozí vynechanie rôznych stavov,
 - vlastnosť nie je homogénna, pri rôznych predmetoch predstavuje rôznu kvalitu;
- duplicita a kvázi duplicita prvkov – postupne pri budovaní číselníka zistíme, že nová hodnota vyhovuje viacerým prvkom číselníka. Môže to byť dôsledkom hierarchie prvkov, prekrývania kritérií alebo zaradenia zoskupení medzi prvky číselníka;
- definícia hraníc v procese diskretizácie (prevod od spojitej, plynule sa meniacej hodnoty na konečnú množinu prvkov) kvantitatívnych vlast-



ností – hrozí, že hranica sa zaradí do dvoch prvkov alebo že sa nezaradí do žiadneho prvku;

- harmonizácia s číselníkmi externých inštitúcií;
- prepojenie s existujúcimi číselníkmi NBS.

Všetkým kvalitatívnym vlastnostiam používaným v informačnom systéme NBS je vhodné priradiť číselník. Pre každú samostatnú hodnotu vlastnosti vytvoríme prvok číselníka, ktorému priradíme kód, názov a definíciu jeho obsahu.

Vytvorenie tried

Keď máme vytvorený zoznam vlastností a číselníky, môžeme pristúpiť k definovaniu tried, do ktorých budeme zaradovať všetky predmety nášho skúmania. Pri definovaní triedy potrebujeme stanoviť:

- názov a identifikačný kód triedy,
- definíciu triedy, teda kritériá na zaradenie predmetov do triedy (aby sa stali objektmi triedy),
- zoznam dimenzií, teda vlastností, ktoré sa používajú na identifikáciu, resp. výber objektov triedy,
- určiť, ktoré dimenzie sú povinné (fixné, pevné) a ktoré nepovinné (variabilné), čiže dimenzie ktoré sa musia použiť pri identifikácii každého objektu triedy, a tie, ktoré sa použijú, iba ak daná dimenzia je nutná na identifikáciu príslušného objektu;
- pre vybrané dimenzie je nutné stanoviť ich implicitné hodnoty (hodnota, ktorá sa predpokladá, ak daná dimenzia nie je použitá; zvyčajne je to všetko alebo sa neuplatňuje);
- zoznam vlastností, ktoré môžeme skúmať na jednotlivých objektoch triedy.

V tomto procese sa stretáva niekoľko protichodných požiadaviek:

- minimalizovať počet tried z dôvodu ich prehľadnosti a jednoduchosti údržby,
- minimalizovať počet dimenzií použitých v triede na identifikáciu predmetu a vlastnosti,
- reprezentatívnosť – chceme aby trieda reprezentovala nejakú ekonomickú realitu, t. j. nejakú skupinu predmetov, ktorá má svoje bežne používané pomenovanie.

Názov novej triedy by mal vystihovať jej obsah. Malo by byť zrejmé, ktoré predmety do nej patria. Pre individuálne triedy je pomerne jednoduché nájsť také pomenovanie triedy, aby zaradenie predmetov do objektov triedy bolo rýchle a jednoznačné. Keďže objekty v agregovaných triedach reprezentujú len nejakú analytickú požiadavku, je potrebná omnoho väčšia obozretnosť pri konštrukcii triedy, ako aj pri vytváraní jej objektov.

Definovanie zoskupení a hierarchií

Nad číselníkmi môžeme následne definovať rôzne usporiadania prvkov typu zoskupenia a hierarchie. Hlavne pri zoskupeniach je tento proces pomerne flexibilný a zoskupenia sa môžu meniť podľa aktuálnych potrieb. Hierarchie zvyčajne odrážajú reálne usporiadanie prvkov, a preto sú omnoho stabilnejšie ako zoskupenia.

Zoskupenie je pomenovaná podmnožina prvkov číselníka. Zoskupenia sa vytvárajú iba nad

prvkami, ktoré samé nie sú zoskupením. Jednotlivé zoskupenia sú vo všeobecnosti navzájom nezávislé, t. j. zloženie prvkov jedného zoskupenia nijako neovplyvňuje zloženie prvkov iného zoskupenia. Zoskupenia ako také nie sú prvkami číselníka.

Hierarchia v tomto modeli predstavuje systém vzájomne usporiadaných prvkov, kde všetky hodnoty priradené prvkom na nižšej úrovni hierarchie musia byť obsiahnuté v prvkoch vyššej úrovne hierarchie. V praxi majú často jednotlivé úrovne hierarchie svoje vlastné pomenovanie (krajina, kraj, okres a obec, alebo rok, mesiac a deň). Pre pomenované vrstvy hierarchie je tendencia vytvárať samostatné číselníky, čo vedie k tomu, že v triedach im musíme priradiť rôzne vlastnosti. Tento krok má opodstatnenie iba vtedy, ak sa rôzne vrstvy nepoužívajú súčasne, alebo ak nižšie vrstvy majú rozdielnu štruktúru a sú dostupné iba pre niektoré prvky vyššej vrstvy. Každý prvok hierarchie je aj prvkom číselníka. Nad každým číselníkom môžeme definovať niekoľko vzájomne nezávislých hierarchií.

Využitie dátového modelu pri práci s údajmi

Dátový model sa buduje na uľahčenie práce s údajmi. Každá položka (typ údajov určený predmetom a vlastnosťou) je v tomto modeli identifikovaná tzv. kľúčom. Kľúč predstavuje identifikáciu objektu v triede a určenie vlastnosti, ktorej sa daný údaj týka. Celý kľúč sa zapisuje pomocou kódov definovaných dátovým modelom. Pomocou tohto kľúča sa zapisujú aj vzťahy medzi jednotlivými údajmi a tiež matematické operácie nad definovanou množinou údajov.

Existuje mnoho spôsobov, ako zapísať tento kľúč. V systéme GESMES kľúč, kde sú všetky dimenzie povinné s pevne stanoveným poradím, obsahuje iba hodnoty jednotlivých dimenzií. V našom modeli je kľúč tvorený premenlivým počtom dimenzií, a preto musíme pri jeho zápise uviesť aj kód dimenzie, ktorej je priradená hodnota. V dôsledku toho je zápis kódu dlhší, ale umožňuje oveľa väčšiu variabilitu, ako s ním pracovať. V štatistickom zberovom portáli kľúč reprezentuje celý časový rad danej položky za všetky subjekty. Výber konkrétnej hodnoty musí byť doplnený o parametre, ktoré určujú konkrétny subjekt a obdobie.

V nasledujúcom príklade si ukážeme, ako vyzerá GESMES kľúč vybranej položky z rodiny (*key family*) „Balance Sheet Items“ a ako rovnakú položku zakódujeme (pridelíme jej kľúč) v našom modeli.

Pre jednu vybranú položku GESMES kód, bez určenia dátumu a subjektu, vyzerá nasledovne:

BSI.M.SK.N.A.A20.A.1.U6.1000.Z01.E

V tabuľke 7 sú uvedené štruktúrne definície pre túto rodinu. Na účely tohto príkladu si definujeme triedu BSI. V tabuľke 7 sú kódy pre jednotlivé dimenzie a označené ich implicitné hodnoty. Pre



Tabuľka 7 Výber z definície kľúča BSI (ECB) na ilustráciu jeho transformácie do kľúča NBS

	Dimenzia/vlastnosť	Kód* dimenzie/vlastnosti	Hodnota	Význam hodnoty	Implicitná hodnota
1	Frequency	F	M*	Monthly	M
2	Reference area	ARE	SK*	Slovakia	SK
3	Adjustment indicator	ADJ	N*	Neither seasonally nor working day adjusted	N
4	BS reference sector breakdown	BSR	A	MFIs excluding ESCB	
5	Balance sheet item	BST	A20	Loans	
6	Original maturity	MAO	A	Total	
7	Data type	DT	1*	Outstanding amounts at the end of the period (stocks)	1
8	Counterpart area	ARC	U6	Domestic	
9	BS counterpart sector	BCS	1000	MFIs	
10	Currency of transaction	CUT	Z01	All currencies combined	
11	Balance sheet suffix	BSX	E*	Euro	E
12	Time period or range	T			
13	Time format code	TF			
14	Observation value				

* Kód a implicitná hodnota v dátovom modeli NBS.

názornosť je kódovanie hodnôt vlastností zhodné s GESMES kódovaním.

Potom môžeme vytvoriť zápis NBS tejto položky. Zápis NBS sa vždy začína kódom triedy (napr. BSI), za ktorým nasledujú v ľubovoľnom poradí kódy dimenzií s kódmi ich hodnôt. Pre zápis kódu je nutné definovať oddeľovače dimenzií a hodnôt. Ak oddeľovač dimenzií je dvojbodka a oddeľovač hodnôt bodka, úplný kód pre danú položku je nasledovný:

BSI:F.M:ARE.SK:ADJ.N:BSR.A:BST.A20:MAO.A:DT.1:ARC.U6:BCS.1000:CUT.Z01:BSX.E

Modrým sú vyznačené variabilné (nepovinné) dimenzie, ktoré súčasne nadobúdajú implicitné hodnoty, preto ich môžeme v kľuči vynechať. Po vynechaní nepovinných dimenzií bude kód danej položky vyzeráť takto:

BSI:BSR.A:BST.A20:MAO.A:ARC.U6:BCS.1000:CUT.Z01

V novom štatistickom zberovom portáli sa na popis položiek vo výrazoch zaviedli ako oddeľovače dimenzií čiarky a ako oddeľovače hodnôt dvojbodky a celý kód sa zvykne uzatvárať do hranatých zátvoriek:

[BSI, BSR:A, BST:A20, MAO:A, ARC:U6, BCS:1000, CUT:Z01]

V ďalšej časti si ukážeme, ako sa reálne podarilo premietnuť zásady informačného modelu do konkrétneho dátového modelu používaného v štatistickom zberovom portáli.

Dátový model v štatistickom zberovom portáli

Dátový model nevznikal ľahko. Aj keď sú už v súčasnosti vybudované jeho základy, zostáva mnoho práce na jeho kompletizácii. Uvedieme krátky pohľad na práce pri návrhu informačného modelu a na budovanie dátového modelu v kontexte so zavádzaním štatistického zberového portálu.

Pri budovaní štatistického zberového portálu bola snaha preniesť do jeho funkcionality zásady informačného modelu a efektívne naplniť dátový model. V čase vytvárania požiadaviek na funkcionality zberového portálu informačný model vôbec neexistoval a neobjavil sa ani v požiadavkách. V procese konzultácií funkcionality tohto portálu už boli známe hrubé rysy tohto modelu a tak sa ich s určitou mierou úspešnosti podarilo doplniť do funkčných požiadaviek na zberový portál. Postupne, tak ako sa uzatvárali funkčné požiadavky, finalizoval sa aj informačný model s prihladením na možnosti portálu.

Súbežne s tým prebiehali práce na dátovom modeli. Robili sa rozsiahle analýzy zbieraných údajov, kde bolo treba analyzovať niekoľko stotisíc položiek. Počas týchto analýz sa postupne zblížovali stanoviská rôznych odborov k obsahovej náplni jednotlivých výkazov. Prechádzala sa položka po položke, zisťovali sa duplicity v definíciách a názvosloví za výkazy z rôznych oblastí. Pokiaľ došlo ku zhode obsahu, hľadali sa vhodné jednoznačné pomenovania a postupne sa budoval zoznam vlastností, ako aj číselníky pokrývajúce možné hodnoty vybraných vlastností.

Niektoré problémy zjednocovania názvosloví a kódov vyriešil aj samotný štatistický zberový



Tabuľka 8 Vybraná časť z definície triedy Menová a finančná štatistika (kód triedy: MFS)

Typ	Kód	Názov	Dátový typ/ číselník	Povinnosť
D	APVN	aktíva/pasíva/náklady/výnosy	CK029	P
D	EP	predmet skúmania/ekonomická položka	CK065	P
D	EPT	predmet skúmania/druh ekonomickej položky	CK066	P
D	F	frekvencia	CK028	P
D	J	merná jednotka	CK073	P
D	TBN	brutto/Netto	CK039	P
...
D	VZA	zabezpečenie	CK030	
V	C	cena	NUM_PERC_2	
V	DXK	dátum – koniec	DATUM	
V	DXZ	dátum – začiatok	DATUM	
V	N	počet	NUM_15_2	
V	O	objem	NUM_15_2	

portál tým, že poskytol možnosť v zoskupeniach definovať tzv. aliasy kódov k jednotlivým hodnotám. Toto tiež zjednodušilo aj prenos údajov s rôznym kódovaním z existujúcich výkazov do portálu bez ich zložitej transformácie.

Na základe uvedených analytických činností sa tvorí zoznam vlastností, ktorých je v súčasnosti identifikovaných (bez EBA, EIOPA) okolo 500. Na výber hodnôt definovaných vlastností je zavedených okolo 135 číselníkov. Údaje môžu byť zaradené do 42 tried. Predpokladá sa, že časom bude počet individuálnych tried narastať, zatiaľ čo počet agregovaných tried by sa mal stabilizovať.

Práca s položkou v štatistickom zberovom portáli

Položka je bunka výkazu s definovaným kódom/kľúčom. Na zakódovanie položiek ilustračnej tabuľky 2 (v 1. časti príspevku) je možné použiť

triedu Menová a finančná štatistika (MFŠ), ktorej dimenzie a vlastnosti sú uvedené nižšie.

Pre vyznačenie položky v tabuľke 2 je potrebné zadať kód vlastnosti, hodnoty pre všetky povinné dimenzie a hodnoty pre relevantné nepovinné dimenzie, tak ako je uvedené v tabuľke 9.

Keď sa pozrieme na túto definíciu, chýba tam vykazujúci subjekt, obdobie a násobiteľ hodnoty. Prvé dve hodnoty implicitne dopĺňa systém a pri práci v zberovom portáli sa použijú aktuálne nastavené hodnoty. Násobiteľ má zmysel iba pri vyplňaní výkazu, vnútorne sa hodnoty ukladajú v jednotkovom tvare, a preto ho nie je nutné ďalej používať.

Kľúč pre danú položku je potom nasledovný:

[MFS, VLA:O, F:M, APVN:A, EP:M007, EPT:MCE, J:SKK, TBN:BT, TOC:C004, UT:U30, CUB:EUR, K:SK, SW3:S11, TOM:KIII]

Tabuľka 9 Metadátové informácie pre vybranú položku

Názov	Kód dimenzie	Popis hodnoty dimenzie	Kód hodnoty dimenzie	Povinnosť
Vlastnosť – objem	VLA	objem	O	
Frekvencia	F	mesačne	M	P
Aktíva/pasíva/náklady/výnosy	APVN	aktíva	A	P
Ekonomická položka	EP	pohľadávky/závazky z úverov	M007	P
Druh ekonomickej položky	EPT	spolu	MCE	P
Merná jednotka	J	slovenská koruna*	SKK	P
Brutto/Netto	TBN	brutto	BT	P
Typ ocenenia	TOC	menovitá hodnota	C004	P
Údajový typ	UT	stav ku koncu obdobia	U30	P
Mena – pôvodná (originálna)	CUB	euro	EUR	
Protistrana krajina	K	Slovenská republika	SK	
Protistrana – sektor	SW3	nefinančné korporácie	S11	
Kvalita I (druh ocenenia)*	TOM*	kategória III	KIII	

* Uvedený príklad pracuje so staršími údajmi, pre ktoré daná trieda nemá ešte definované všetky položky, preto sa chýbajúce položky doplnili ručne.



Význam klúča je v tom, že sa používa ako premená v výrazoch. Pri vyhodnocovaní výrazu sa klúč nahradí hodnotou zodpovedajúcej položky. Uvedený klúč je pomerne dlhý z dôvodu veľkého počtu povinných dimenzií a dĺžky jednotlivých čiastkových kódov. Časom sa predpokladá optimalizácia tried a kódov, čo by mohlo viesť k výrazne kratším klúčom. Dĺžka klúča nemá žiadny vplyv na funkcionality zberového portálu, výrazne však ovplyvňuje čitateľnosť klúčov a prehľadnosť výrazov, v ktorých sa tieto klúče používajú.

Prehľadnosť výrazov s dlhými klúčmi sa zlepšuje pomocou rozdelenia zápisu klúča na časti. Hodnoty dimenzií, ktoré sa vo výraze opakujú (nemenia), je možné vyňať pred výraz a v samotnom zápise výrazu sú už iba premenlivé časti klúča.

Zoznam s vyňatými dimenziami pre jednotlivé triedy zapisujeme pomocou príkazu „with“, ktorý má takýto syntax:

with(<dimenzie triedy1>,< dimenzie triedy2>,...),

kde sa dimenzie triedy zapíšu nasledovne:

[<kód triedy>, <kód dimenzie>:<kód prvku číselníka>,...]

Príklad zápisu vyňatých dimenzií pre dve triedy T1 a T2, s dimenziami D4-D7 a hodnotami Ha-Hq:

with([T1, D4:Ha, D5:Hb, D6:Hc], [T2, D7:Hp, D6:Hq])

Ak by sme chceli vypočítať hore uvedenú položku z položiek detailnejšieho sektorového členenia, kde S11 je súčet položiek za sektory S11001, S11002 a S11003, tak hodnoty dimenzií (okrem dimenzie sektor nadobúdateľ – SW3) sa nemenia a môžeme ich vyňať a výraz zapíšeme takto:

with([MFS, VLA:O, F:M, APVN:A, EP:M008, EPT:MCE, J:SKK, TBN:BT, TOC:C004, UT:U30, CUB:EUR, K:SK, TOM:KIII])
[MFS, SW3:S11] = [MFS, SW3:S11001] + [MFS, SW3:S11002] + [MFS, SW3:S11003]

Systém umožňuje zapísať univerzálny klúč, ktorý zodpovedá viacerým položkám tak, že za hodnotu dimenzie môžeme zadať množinu hodnôt alebo zoskupenie. Množinu hodnôt uzatvárame do zložených zátvoriek. Predchádzajúci výraz s pomocou funkcie SUM môžeme zapísať nasledovne:

with([MFS, VLA:O, F:M, APVN:A, EP:M008, EPT:MCE, J:SKK, TBN:BT, TOC:C004, UT:U30, CUB:EUR, K:SK, TOM:KIII])
[MFS, SW3:S11] = SUM [MFS, SW3:{S11001, S11002, S11003}]

Ak vytvoríme zoskupenie ZS11 = {S11001, S11002, S11003}, predchádzajúci výraz s pomocou využitia zoskupenia bude nasledovný:

with([MFS, VLA:O, F:M, APVN:A, EP:M008, EPT:MCE, J:SKK, TBN:BT, TOC:C004, UT:U30, CUB:EUR, K:SK, TOM:KIII])

[MFS, SW3:S11] = SUM [MFS, SW3:@ZS11]

Ak by sme chceli použiť všetky hodnoty, ktoré daná dimenzia môže nadobudnúť, na mieste hodnoty použijeme znak hviezdičku „*“. Potom by sa agregácia všetkých hodnôt za všetky sektory mohla zapísať takto:

with([MFS, VLA:O, F:M, APVN:A, EP:M008, EPT:MCE, J:SKK, TBN:BT, TOC:C004, UT:U30, CUB:EUR, K:SK, TOM:KIII])

[MFS] = SUM [MFS, SW3:(*)]

Formálne je tento výraz správny, avšak zmysluplné výsledky by sme získali iba vtedy, ak by každý úver bol zaradený iba do jedného sektora (pridelený jednému kódu).

Opakovaný výpočet pre rôzne hodnoty dimenzií sa vykonáva pomocou príkazu „for“, ktorého parametrom sú dimenzie so zoznamom hodnôt, pre ktoré sa vykonajú samostatné výčíslenia výrazu.

Príklad výpočtu objemu za sektor S11 pre rôzne meny denominácie:

with([MFS, VLA:O, F:M, APVN:A, EP:M008, EPT:MCE, J:SKK, TBN:BT, TOC:C004, UT:U30, K:SK, TOM:KIII])
for(CUB:{EUR, SKK, USD, JPY})
[MFS, SW3:S11] = [MFS, SW3:S11001] + [MFS, SW3:S11002] + [MFS, SW3:S11003]

Funkcia „for“ pracuje podobne ako vyňaté dimenzie s tým, že sa postupne prechádza cez všetky hodnoty dimenzií dané parametrami funkcie a vykoná sa výpočet výrazu tak, že sa aktuálna hodnota dimenzií dosadí do klúča položky.

Štatistický zberový portál obsahuje aj ďalšie typy konštrukcie výrazov spojené s grafickou prezentáciou výkazov. Sú navrhnuté rôzne konštrukcie pre prácu s položkami výkazov. Ďalej sú konštrukcie výrazov umožňujúce pracovať s premenlivými zoznamami hodnôt dynamických výkazov a ďalšie špeciálne konštrukcie, ktorých syntax a použitie možno nájsť v dokumentácii zberového portálu.

OČAKÁVANÝ VÝVOJ NA NAJBLIŽŠIE OBDOBIE

Štruktúrovaný popis údajov spracúvaných v NBS je v súčasnosti v stave prvotného nasadenia, kde sa postupne ukazujú jeho slabiny, ale aj možnosti. Budúce využitie a ďalší vývoj tohto nástroja možno sledovať z viacerých aspektov. Uvedieme si aspoň tie najdôležitejšie:

- rozvoj samotného informačného modelu,
 - optimalizácia dátového modelu,
 - prepojenie štatistického zberového portálu s užívateľským prostredím využívajúcim dátový model,
 - špecifikácia štandardu na výmenu informácií.
- V každom z týchto okruhov je viacero otvorených otázok, ktoré by bolo vhodné uspokojivo



vyriešiť. S postupom času sa niektoré z nich viac alebo menej dostávajú do popredia, niektoré sa stávajú bezvýznamnými a objavujú sa ďalšie, ktoré treba riešiť.

Rozvoj informačného modelu

V informačnom modeli zostala otvorená celá oblasť spojená s popisom procesov pre jednotlivé triedy. Je vhodné doplniť do modelu aj nástroje na popis vzájomnej interakcie a závislosti medzi objektmi jednej alebo viacerých tried (napr. kontroly, definícia drill down).

Ďalej je potrebné zaviesť všetky registre ako individuálne triedy a možnosť oboru hodnôt vybraných vlastností vyberať z objektov individuálnej triedy.

Treba tiež dopracovať pravidlá na definovanie hierarchií v číselníkoch a agregačné pravidlá medzi prvkami číselníka a stanoviť pravidlá na definovanie implicitných hodnôt, symboly a syntax na používanie zástupných hodnôt (tzv. žolíkov) so zohľadnením hierarchií a mechanizmov na vylúčenie dvojitého započítania tých istých objektov v procese agregácie.

Skúsenosť ukazuje, že tvorba dátového modelu, aj keď formálne vyhovuje informačnému modelu, podlieha rôznym individuálnym vplyvom, čo má za následok pomerne veľký počet a roztrieštenosť jednotlivých tried a tiež číselníkov. V triedach je definovaný príliš veľký počet povinných dimenzií, číselníky obsahujú množstvo iba marginálne sa líšiacich hodnôt a pod. Tieto nedostatky by bolo možné potlačiť pomocou záväznej metodiky na tvorbu dátového modelu, ktorá by stanovovala striktné pravidlá, ako vytvárať triedy, ako optimálne vytvárať dimenzie a vlastnosti pre objekty triedy, za akých podmienok môžeme dimenziu zaradiť ako povinnú a pod.

Optimalizácia dátového modelu

Aktuálny dátový model NBS vznikol v extrémne krátkom čase a predstavuje prvotné premietnutie súčasného výkazníctva v procese zavádzania štatistického zberového portálu, ako aj formalizácie informačného modelu. Tu bolo treba zohľadňovať množstvo technických požiadaviek a obmedzení, často aj na úkor vecných. V dôsledku toho sformovaný dátový model je síce funkčný, ale má ešte ďaleko od optimálneho stavu. Výber tried, číselníkov a ich prvkov do značnej miery odráža organizačné členenie NBS na úkor vecného členenia údajov. Triedy obsahujú mnoho tzv. exotických dimenzií, t. j. takých, ktoré umožňujú veľmi špecifický a malo používaný výber údajov, nezapadajú do jednotného systému zberu údajov a nevyskytujú sa v žiadnych medzinárodných triedeniach. Definované hodnoty dimenzií v číselní-

koch (odkazujúce na tú istú podstatu) sa opakujú z dôvodu drobných metodických odchýlok alebo z nemožnosti v krátkom čase posúdiť ich zhodu.

Užívateľské rozhrania k údajom zo zberového portálu

Pre priamych aj nepriamych používateľov štatistického zberového portálu je potrebné pripraviť nástroj na prístup k týmto údajom. Predpokladá to prípravu nástrojov na plnú podporu informačného modelu. Tieto nástroje by mali umožňovať pohodlné prezeranie dátového modelu interaktívnym spôsobom, prístup k zozbieraným štatistickým údajom pomocou výberov tried a hodnôt dimenzií. Ďalšie nástroje by mali umožňovať prácu s registrami ako so špecifickými triedami tohto dátového modelu.

Technický štandard na výmenu informácií

Pri prvotnom návrhu informačného modelu sa neriešili otázky prenosu údajov. Samozrejme, prenos údajov do štatistického zberového portálu je zabezpečený, ale tento prenos sleduje technické požiadavky portálu a nie informačného modelu. Preto by bolo vhodné definovať jednoduchý výmenný formát na prenos údajov vychádzajúci z informačného modelu. Prenášané údaje by v tomto formáte boli identifikované pomocou kľúčov platných pre jednotlivé triedy.

Tento základný formát na výmenu údajov by sa ďalej doplnil o formáty súborov určených na distribúciu dátového modelu, na distribúciu zmien dátového modelu, ako aj súborov na zadávanie požiadaviek na množinu údajov a súborov na definovanie kontrol.

Použité zdroje informácií

1. Interné zdroje NBS a dokumenty dostupné na www.nbs.sk
 - opatrenia NBS týkajúce sa predkladaní výkazov bankami
 - interný IS Status
 - informácie k štatistickému zberovému portálu (www.nbs.sk/sk/statisticke-udaje/statisticky-zberovy-portal)
2. Dokumenty týkajúce sa GESMES
 - štruktúrne definície ECB
 - GESMES/TS (Generic statistical message/time series) The time series data exchange message, User Guide
3. Dokumenty týkajúce sa datapoint modelu EBA (<http://www.eba.europa.eu/>)
 - Description of DPM formal model.pdf
 - EBA Architecture for XBRL representation of DPM.pdf
 - EBA CP 2014 03 (XBRL Taxonomy).pdf
 - DPM_Database_2300.accdb