

Řešení základního problému kvality dat IS AČR

V článku nabízím systém, jehož funkcí je organizace a technika zajišťování logické a formátové integrity datové základny rozsáhlého informačního systému. Tento problém je snad v současné době nejobtížnější řešitelný z těch, které doprovázejí zajišťování kvality datové základny informačního systému Armády České republiky (IS AČR) jako funkčního celku. Podotýkám, že jde o to, aby tento informační systém, budovaný postupně v čase různými dodavateli s využitím různých softwarových technologií, poskytoval do systému velení a řízení včasnou, pravdivou a konzistentní informaci. Článek je rozdělen do dvou částí. První část článku má úvodní charakter a je věnována datovým sborníkům a možnosti jejich využití v rozsáhlém informačním systému při řešení kvality jeho datové základny. Koncept vlastního systému organizace a techniky zajišťování logické a formátové integrity datové základny informačního systému je náplní druhé části článku (bude uveřejněn v některém z příštích čísel této publikace) pod názvem „Podstatná součást řízení a správy dat IS AČR“.

* * *

Pro využívání počítačů je již po mnoho let charakteristická integrační tendence, za jejíhož nositele lze jednoznačně považovat výrobce hardware. Neuvěřitelně se zvyšují rychlosti zpracování dat jednotlivých typů počítačů, provozují se síťové komplety, v provozu je celosvětová počítačová síť INTERNET. Jsou vytvářeny rozsáhlé informační systémy. Odvozeně od integrační tendence v nasazování a využívání počítačů a souvisejících komunikací se na mezinárodní úrovni realizují práce směřující k logickému a formátovému zabezpečení elektronické výměny dat. Jsou vytvářeny datové normy a sborníky datových prvků ISO (mezinárodní standardizační organizace). V České republice jsou tvořeny a vydávány datové normy a datové sborníky Českým normalizačním institutem.

Odráz těchto prací je v rezortu obrany, bohužel, zatím nedostatečný; nepodařilo se dosud udělat ani první krok, tj. ustavit zde orgán pro normalizaci (standardizaci) datových prvků a datových struktur. Toto konstatování potvrzuje nově vydaný RMO č. 3/96 k organizačnímu zabezpečení standardizace v rezortu Ministerstva obrany, podle něhož datová standardizace (prakticky je lhostejné, zda použijeme termín normalizace nebo standardizace dat) vůbec neexistuje nebo je zasunuta na povaze problému neodpovídajícím místě, viz odstavec 3. Standardizaci dělíme na: A přitom role interních normativních aktů při výstavbě a provozu tak rozsáhlého informačního systému, jakým je a bude IS AČR, je nezastupitelná. Datovou standardizaci (přikláním se spíše k používání pojmu normalizace dat) nelze v armádě provádět, aniž by se její organizace a průběh neopíral nejlépe o Rozkaz ministra obrany. Důvod je prostý: je žádoucí, aby normalizace dat zahrnovala celý rezort a byla řízena centrálně, to na jedné straně, a na druhé straně proto,

že činnosti s ní spojené jdou napříč systémem velitelské podřízenosti. RMO (nestačí N NGŠ) také proto, že informační systém správní části MO je zdrojem dat (informací) pro informační systém AČR.

Oblast normalizace dat je dle mého soudu neprávem stranou zájmu při výstavbě IS AČR, neboť kvalitu datové základny lze stěží postavit na dodatečné integraci dat, tj. dodatečně zabezpečované logické a formátové integrity použitých dat v provozovaných programových celcích. Tuto stránku kvality datové základny je nejlépe zajišťovat co nejdříve již v průběhu výstavby tak rozsáhlého informačního systému, jakým je IS AČR. Můžeme si právem položit otázku: „Kdy začít s dodatečnou nebo jinak řečeno s následnou integrací dat?“ Čím později, tím rozsáhlejší programové celky by podléhaly změnám, tím hlubší logická a formátová nejednotnost by byla do IS AČR zabudována.

V článku „Informační systém AČR, dělba práce - dělba odpovědnosti“, uveřejněném v čísle 6/95 Vojenských rozhledů, jsem vyjmenoval několik normativních aktů, které je podle mého názoru potřebné vydat k normativnímu zabezpečení kvality datové základny IS AČR. Nyní se vracím k problematice naposled jmenovaného normativního aktu na str. 73, k problematice **organizace a techniky zajišťování logické a formátové integrity datové základny IS AČR**, kde stěžejní roli v navrhovaném řešení sehrávají datové sborníky.

Co to je datový sborník

Mohlo by se na první pohled zdát, že jde o nějaká reálná data. Svým způsobem skutečně o reálná data jde, ale nejsou to data jako zdroj informací pro systém velení a řízení, jsou to specifikace datových prvků - DP, čili jde o jakási metadata - data o datech.

Datový prvek je jednotka dat v určitém kontextu považovaná za nedělitelnou a určená prostřednictvím specifikačních atributů (popisných charakteristik). Datovým prvkem je například vojenská odbornost nebo akademický titul.

Hlavní částí datového sborníku jsou specifikace jednotlivých datových prvků (nebo datových objektů - DO, tj. číselníků, informačních vět, informačních souborů a dalších projektových entit). Ostatní části datového sborníku se vztahují k formě vydání a k vydavateli datového sborníku. V článku se zabývám pouze hlavní částí datového sborníku.

Sborník datových prvků není nic jiného než jakási kolekce normativních specifikací jednotlivých datových prvků. Sborníky datových prvků se na úrovni ISO i ČSN tvoří a vydávají pro hospodářské či správní oblasti, jako je doprava, obchod, zemědělství, práce, mzdy a sociální věci, atd. V rezortu obrany by mohly být tvořeny a vydávány datové sborníky pro oblasti vojenské činnosti s větším počtem vlastních předmětných pojmů, dále nazývané „registrační oblasti“.

Normativní specifikace datového prvku

Tvorba jednotlivých normativních specifikací DP podléhá pravidlům. Jsou stanovena pravidla sestavování názvů a definic, odkazují buď na normu ISO 11179 ve všech jejích částech nebo na připravovanou normu ČSN 97 4001-1, která ovšem není s normou ISO obsahově totožná. Naše datová norma je stavěna pouze pro popis datových prvků jako metodická pomůcka při tvorbě hlavní části datových sborníků na národní úrovni. Pro

potřebu tvorby a využití datových sborníků při vývoji a provozu rozsáhlého informačního systému bude vhodné využít obě uvedené normy jako metodická doporučení.

Pro ilustraci uvádím z normativní specifikace datového prvku ty atributy, které jsou z pohledu logické a formátové integrity dat a rovněž pro projektanta a programátora při vývoji programového vybavení informačního systému nejdůležitější. Jsou to:

- | | |
|----------------------|----------------------------------|
| * definice, | * datový typ, |
| * název s kontextem, | * minimální a maximální hodnota, |
| * verze specifikace, | * přípustné hodnoty. |

Pro zájemce je ve VTCI AČR k dispozici v první verzi návrh metodiky popisu datové základny IS AČR, který obsahuje jak normativní specifikace datových prvků a datových objektů, tak projektové specifikace téhož. Projektová specifikace je podstatně užší, má méně atributů a je určena především k rozlišení logické a formátové shodnosti datových prvků a objektů použitých v různých dílčích IS a APV v IS AČR.

Pod pojmem „aplikační programové vybavení (APV)“ zde rozumíme relativně uzavřený funkční celek (programy + data), který je účelné využívat ve prospěch velení a řízení AČR.

Situace v tvorbě a využívání datových sborníků

Pro oblast výměny dat již některé datové sborníky existují. Například Sborník datových prvků pro obchod, Sborník datových prvků pro oblast práce, mzdy a sociální věci - část odměňování a některé další. Je rozpracován Sborník datových prvků pro železnici. Pro zajišťování logické a formátové integrity dat v rámci rozsáhlého informačního systému bude jistě možné existující specifikace datových prvků použít, ale většinu specifikací datových prvků bude nutno vytvořit v rámci projekční činnosti. V případě IS AČR to jistě nikoho neudiví. Převážná část datových prvků, domén, číselníků a výše zmíněných datových objektů, které jsou a budou v IS AČR použity, odráží jednotlivé vojenské předmětné oblasti. Může se ovšem vyskytnout názor, že bychom měli datové sborníky převzít z NATO. Do jaké míry je takový popud realizovatelný, čas ukáže. V této souvislosti existuje více problémů, především by však NATO datové sborníky muselo mít, průběžně je aktualizovat a využívat k udržování logické a formátové integrity všech dat v rámci životního cyklu svého informačního systému. Nejspíše lze očekávat, že NATO používá datovou normalizaci pouze pro zabezpečení výměny dat mezi informačními systémy armád členských zemí. Samozřejmě, že to není málo, ale tento článek je zaměřen na podstatně širší využití datové normalizace a datových sborníků. Datové sborníky sestavené pro zabezpečení výměny dat nepostačí jako nástroj zabezpečování logické a formátové integrity datové základny IS, viz schéma. Pro elektronickou výměnu dat potřebujeme normalizovat pouze DP/DO, které se vyskytují v přenášených dokumentech, a pravidla syntaxe přenosu dokumentů. To je však již problematika, kterou se článek nezabývá.

Logická a formátová integrita datové základny

Zabezpečit plnou funkčnost IS AČR vyžaduje, mimo jiné, pracovat s aktuálními a konzistentními daty. Prvním, nikoli však jediným, předpokladem dosažení tohoto stavu je logicky a formátově integritní (LFI) datová základna informačního systému AČR. Pojem formát zahrnuje datový typ, maximální velikost hodnoty a minimální velikost hodnoty. Následující charakteristika LFI datové základny odpovídá procesu vytváření normativní



Schéma tvorby a obsahu datových sborníků v závislosti na jejich využití

specifikace DP/DO, tj. procesu tvorby datových sborníků v průběhu životního cyklu IS AČR.

Logicky a formátově integritní bude datová základna tehdy, jestliže se v ní pro logicky shodné datové prvky a datové objekty nevyskytnou jejich duplicitní specifikace. Tohoto stavu lze dosáhnout za předpokladu, že se při vývoji a provozu IS budou používat pro logicky shodné datové prvky a datové objekty po celý životní cyklus systému stejné specifikace, které se mohou lišit pouze verzí, a ta musí být v dokumentaci systému zaevidována. Není nic jednoduššího, pokud ovšem potřebné specifikace DP/DO projektant má k dispozici tehdy, kdy je potřebuje.

Místo datových sborníků ve výstavbě a provozu IS AČR

Lze uvažovat dvě oblasti využití datových sborníků. První oblastí je zabezpečování výměny dat a druhou, neméně významnou, je zabezpečování logické a formátové integrity datové základny rozsáhlého informačního systému. Tvorba a obsah datových sborníků pro tyto dvě oblasti se liší, viz schéma. Přejdu k problému tvorby a využívání datových sborníků ve druhé oblasti.

Může vůbec projektant programového vybavení informačního systému mít k dispozici před zahájením projekčních prací specifikace všech potřebných DP/DO? Není to možné, protože použité DP/DO jsou odvozovány od funkcí, které má programové vybavení plnit. Avšak projektant může mít k dispozici mnoho specifikací DP/DO, které jsou v široké míře používány: rodné číslo občana, jméno občana, datum narození, státní občanství, variabilní symbol a mnoho dalších i z výše jmenovaných datových sborníků vydaných ČSNI. Projektant může mít rovněž k dispozici specifikace DP/DO, které již byly v rámci výstavby informačního systému dříve použity. Takové specifikace jsou prakticky využitelné, budou-li průběžně soustřeďovány do datových sborníků. Jestliže by ovšem datové sborníky vznikaly jako pouhý souhrn použitých

specifikací DP/DO, nedaly by se použít k zabezpečování LFi dat. Takto sestavené datové sborníky by zcela jistě obsahovaly i duplicitní specifikace logicky shodných DP/DO.

K zajišťování logické a formátové integrity datové základny rozsáhlého informačního systému mohou sloužit datové sborníky normativních specifikací DP/DO, které vznikají a jsou využívány průběžně v rámci životního cyklu informačního systému jako produkt datové normalizace.

Je ovšem třeba pamatovat na pokročilé SW technologie používané pro vývoj programového vybavení IS AČR. Tyto technologie zahrnují prostředky pro zajištění logické a formátové integrity dat (například na bázi systémového katalogu v INFORMIXu atd.). Jejich prostředky pro zajištění logické a formátové integrity dat jsou využitelné pouze v rámci uplatnění SW technologie. Avšak IS AČR tvoří neukončená řada dílčích IS a samostatných APV vyvíjených nikoli jednou SW technologií. Navíc vývoj součástí IS AČR probíhá postupně v čase různými dodavateli většinou z civilního sektoru.

V nejlepším případě lze uvažovat logickou a formátovou integritu dat v rámci použití jedné nebo propojených SW technologií (jedná se o propojení katalogů dat při vývoji programového vybavení) i bez využití datových sborníků normativních specifikací DP/DO. Tento závěr je, bohužel, slabou útěchou, neboť neposkytuje záruku LFi DZ IS AČR jako celku. **Je třeba, aby nad využívanými SW technologiemi působily vhodné, vůči nim externí procesy, v jejichž „režii“ budou, mimo jiné, tvořeny sborníky normativních specifikací DP/DO jako stěžejní nástroj zajišťování logické a formátové integrity DZ IS AČR.**

* * *

Problematika zabezpečování kvality datové základny informačního systému spadá do oblasti řízení a správy dat (ŘaSD), která je v současnosti, na škodu věci, mnoha odborníky z oblasti informatiky chápána jako problém řešený softwarovými technologiemi. Ve vztahu ke specifikám výstavby IS AČR a zabezpečování základu logické a formátové integrity jeho datové základny jsou možnosti SW technologií, jak jsem ukázal výše, omezené na dílčí informační systémy a aplikační programová vybavení.

Externí procesy působící nad použitými SW technologiemi, které by byly schopné zabezpečovat základ kvality datové základny IS AČR v průběhu jeho životního cyklu, lze skloubit v systém, jehož koncept je obsahem druhé části článku jmenovaného v úvodu.