

# Význam datových standardů pro automatizované sdílení dat

## Data Standards Meaning for the Management of Shareable Data

Jitka Štěpánová  
Katedra informačních technologií, PEF ČZU Praha  
165 21 Praha 6 - Suchdol

### **Anotace:**

Článek je uvedením do problematiky datových prvků a jejich standardizace. Zabývá se vysvětlením základních pojmů souvisejících s problematikou a konceptuálním modelováním obsahu registrů datových prvků. Vychází z platných nebo v současné době revidovaných standardů ISO, případně ANSI.

### **Summary:**

The ISO/IEC 11179, part 1 - 6 and ANSI X3.285 standards outline the information elements (attributes) associated with a data element concept that need to be available for determining the meaning of a data element to be shared between systems.

### **Klíčová slova:**

data, metadata, koncept, datový prvek, sborník (registr) datových prvků, formy prezentace datových prvků, atributy datových prvků, hodnotová doména

### **Key words:**

data, metadata, concept, data element, data element registry, data element representation, attributes, value domain

Pochopení zmíněných pojmů je nezbytnou podmínkou vytypování a standardizace datových prvků používaných v komunikaci. Nedostatek standardizovaných datových prvků brzdí globální implementaci EDI (Electronic Data Interchange, elektronická výměna dat), kde datové prvky tvoří obsah tzv. EDI zpráv.

Základní problémy, které je třeba v této oblasti řešit, jsou:

- jednotné směrnice pro identifikaci, definici a popis datových prvků (základem je šestidílná norma ISO/IEC 11179, Informační technologie - Specifikace a standardizace datových prvků, z níž článek vychází)
- jednoznačné globální identifikátory normalizovaných datových prvků, které dosud běžně neexistují

- dokumentace charakteristik datových prvků, která by podporovala automatizované sdílení dat včetně jejich vyhledávání, výběru a výměny
  - mechanismy umožňující globálně pořizovat a vyměňovat data, zejména data z různých aplikačních oblastí.

### **Datové standardy**

Při elektronické výměně dat mají význam **standardy**, které zajišťují, aby se přenos mezi různými platformami vůbec mohl uskutečnit, a aby se obsah informace, její struktura a uspořádání na obou stranách přenosu správně interpretovaly. **Standardy existují nejen pro hardware, software, komunikace, ale i pro data, tj. datové standardy.**

Datové standardy se definují ve formě **sborníků (registrů)**, které obsahují popisy, definice a další atributy tzv. **datových prvků** (data elements), případně složených datových prvků, segmentů, zpráv. Tyto datové sborníky bývají specifické pro určité oblasti použití, nebo odvětví a jejich význam vzrůstá s globální výměnou dat a dokumentů prostřednictvím sítí.

Standardy pro elektronickou výměnu obchodních dat souvisí s hierarchií obchodní dokumentace, kde různé úrovně, počínaje tou nejnižší, představují:

- datové prvky
- složené datové prvky (adresa)
- segmenty (popis zboží)
- zprávy (reprezentují obchodní dokument: faktura, objednávka, dopravní příkaz, celní deklarace)
- funkční skupiny (souhrn zpráv, např. objednávek).

**Datový sborník (registr) je místo, kde jsou uloženy charakteristiky (atributy) dat potřebné pro jasný popis, inventuru, analýzu a klasifikaci dat.** Jednotky sdílených dat mají ve sborníku přesné identifikátory, významy, struktury a hodnoty. Sborník je nástroj pro správu sdílených dat. Neobsahuje data samotná, ale poskytuje informaci o definici, původu, zdroji a umístění dat. Popisy sdílených dat se nevztahují k žádné určité aplikaci, ani k aplikačnímu prostředí kde se datový prvek bude používat, k funkci datového prvku ani k úrovni abstrakce jeho významu.

Některé logické **datové prvky** jsou pokládány za generické. Jsou to prvky, které mají ustálenou, dobře definovanou hodnotovou doménu a mohou být sdílené více systémy. Jsou to např. datové prvky **Název země a Kód země**. Standard ISO 3166 Kódy pro prezentaci názvů zemí, představuje takový sdílený referenční seznam názvů a kódů zemí. Představuje hned několik datových prvků a jejich hodnotových domén:

- krátký název země v angličtině
- oficiální název země v angličtině
- dvouznakový abecední kód
- tříznakový abecední kód
- tříznakový numerický kód
- krátký název země ve francouzštině
- oficiální název země ve francouzštině

### **Atributy datových prvků podle standardu ISO/IEC 11179**

Standard ISO/IEC 11179, Informační technologie - Specifikace a standardizace datových prvků, část 3 se soustřeďuje na základní atributy datových prvků, které by měly být shromážděny ve sborníku, viz následující tabulka.

#### **Přehled atributů datových prvků**

<b>Kategorie atributu</b>	<b>Název atributu</b>	<b>Závaznost</b>
<i>Identifikační atributy</i>		
	Název	M
	Identifikátor	C
	Verze	C
	Registrační úřad	C
	Synonymum	O
	Kontext	C
<i>Definiční atributy</i>		
	Definice	M
<i>Relační atributy</i>		
	Klasifikační schéma	O
	Klíčové slovo	O
	Odkaz na související data	O
	Typ vztahu	C
<i>Prezentační atributy</i>		
	Kategorie prezentace	M
	Forma prezentace	M
	Datový typ hodnot datového prvku	M
	Maximální délka hodnot datového prvku	M

Minimální délka hodnot datového prvku	M
Schéma prezentace	C
Přípustné hodnoty datového prvku	M
<i>Administrativní atributy</i>	
Pověřená organizace	O
Registrační status	C
Správce datového prvku	O
Komentář	O

Některé atributy jsou uváděné povinně u každého datového prvku ve sborníku (M - mandatory), jiné jsou volitelné (O - optional), jiné zase podmíněné (C - conditional) existencí jiného atributu, tj. povinné za jistých specifikovaných podmínek.

Standard se většinou týká **abstraktní úrovně** a nikoliv aplikační úrovně datových prvků. Podle ISO 11179 jsou datové prvky charakterizovány na nejvyšší úrovni, na níž se vyskytují: jako **objektová třída, vlastnost a prezentace** (representation). Každý datový prvek má hodnotovou doménu, která je též dokumentována. **Hodnotová doména jsou přípustné hodnoty datového prvku**. Poskytují prezentaci, ale nelze z nich vyvodit, s kterým datovým konceptem jsou hodnoty spojeny ani co znamenají.

Dobře navržené datové prvky a jejich hodnotové domény (množiny přípustných hodnot) se zaznamenávají v datovém registru jako vzory pro potencionální znovupoužití.

Avšak chceme-li vědět jak je datový prvek použit v aplikaci, jsou třeba ještě další atributy. Příklady některých potencionálních atributů konkrétních datových prvků: *Účel, pro který jsou data zaznamenávána, Statistická metodologie použitá v datovém souboru*, apod.

Další úroveň vývoje datového prvku je tedy **aplikační úroveň**. Obecný datový prvek bude např. přizpůsoben aplikaci zařazením jen určité části své hodnotové domény nebo zúžením příslušné definice (nebo obojím), aby se daly zahrnout pouze určité hodnoty, které bude daná aplikace používat. Obecný datový prvek **Název země** se týká všech zemí, ale datový prvek aplikace by se mohl jmenovat **Název země obchodního partnera**. Hodnotová doména by pak byla podmnožinou zemí uvedených v ISO 3166. Termín pro tzv. kvalifikátor - *obchodní partner* - představuje sám o sobě objektovou třídu. Tyto vztahy by měly být vyjádřeny hierarchicky v datovém modelu.

Poslední typ datového prvku je **fyzický datový prvek**. Jsou to prvky, které se skutečně objeví v záhlaví sloupce databázové tabulky, v popisu souboru, v uspořádání transakčního souboru EDI, atd.

Navrhovaná **revize normy** ISO/IEC 11179, část 3 modeluje datový prvek a s ním související koncept datového prvku. **Datový prvek sestává z konceptu datového prvku a jeho prezentace.** Datový koncept je definován jako vlastnost souboru objektů (třídy) popsaná nezávisle na jednotlivých prezentacích, např. *barva objektu*. Prezentace je definována minimálně jako kombinace hodnotové domény (všechny možné barvy objektu, které připadají v úvahu), datového typu (numerický, alfanumerický...) a , je-li třeba, měrné jednotky.

Návrh **datového prvku aplikace** často vyžaduje dodatečnou kvalifikaci objektové třídy a (nebo) vlastnosti (např. *původní barva objektu*). Termín, který takto upřesňuje definici a pomáhá prokazovat jednoznačnost konceptu, se nazývá kvalifikátor.

Rozlišení různých úrovní datového prvku, konceptu datového prvku a jejich vzájemných vztahů má pomáhat při tvorbě dobře navržených dat, která se dají sdílet. Je třeba specifikovat vztahy a dědičnost počínaje nejobecnějším datovým prvkem a konče konkrétním aplikačním datovým prvkem. Při analýze v této oblasti se nyní uplatňuje objektově orientovaný přístup, zejména při sestavování hierarchie objektových tříd. Další problém, který doposud standardy neřešily, je znovupoužití datové hodnotové domény.

### **Metamodel podle standardu ANSI X3.285**

Tento článek vychází též z ANSI X3.285 Standardu - Metamodel pro správu sdílených dat, 1999, který specifikuje strukturu datového registru ve formě konceptuálního modelu a je doplňkem k šestidílnému standardu ISO/IEC 11179, Informační technologie - Specifikace a standardizace datových prvků. (ANSI - American National Standard Institution).

Modelování dat je založeno na teorii, že všechna data popisují vlastnosti (atributy) objektů v realitě. Základní jednotky dat jsou **datové prvky**.

K vyjádření struktury datového sborníku se používají **konceptuální** (pojmové) **datové modely**. Popisují schéma registru, kde jsou uloženy zmíněné popisy sdílených dat, nazývané také **metadata**. Datový model metadat je potom **metamodel**. Metamodel je nezbytný pro koordinaci datové reprezentace mezi různými softwarovými nástroji, jež ukládají data, manipulují s nimi a vyměňují si je.

Datové sborníky vytvářené podle standardů mohou pak využívat informační manažeři, tvůrci archivů (repository) a CASE produktů, správci dat a ti, kteří jsou zodpovědní za sdílení dat.

Konceptuální datový model (KDM) zobrazuje, jak je relevantní informace strukturována v realitě, t.j. jak ji vnímáme, jak je člověk zvyklý na informaci pohlížet. KDM představuje

o jednu úroveň vyšší abstrakci než logický datový model a je vhodný pro první fázi modelování dat. Je to také vhodná úroveň, na které lze **integrovat různé datové modely**. Logický datový model popisuje v podstatě táz data, ale jak už jsou stukturována v informačním systému (model informačního systému ). Logický datový model může být pak přímo použit pro návrh databáze.

Jednotky složené z více prvků jsou **složené datové prvky** a mají odlišné vlastnosti než datové prvky. V registru je popis datového prvku dokumentován zároveň s každým složeným datovým prvkem, který jednotlivý prvek zahrnuje.

Přístup k problematice datových prvků a jejich atributů ve sborníku je podle ANSI X3.285 Standardu - Metamodel pro správu sdílených dat, 1999, obdobný, ale atributy jsou detailněji rozebrány zejména co se týče hodnotové domény. První věc, na kterou se klade důraz při registraci datového prvku, je úvaha, jakým způsobem bude prvek reprezentován, tedy jeho **hodnotová doména a třída prezentace** (forma vyjádření prvku). Seznam možných prezentačních tříd je zde oproti ISO 11179 rozšířen a jednotlivé termíny vyjadřující prezentaci jsou součástí názvu prvku.

Třídy prezentace:

částka, kód, grafika, ikona, míra, název, číslo, množství, text

### **Atributy datových prvků podle standardu ANSI X3.285**

#### **Prezentace datového prvku**

třída prezentace  
hodnotová doména  
*identifikátor*  
*název domény*  
*definice*  
*datový typ (soubor znaků)*  
*typ domény*  
*extrémní hodnoty (délka)*  
*formát*  
*měrná jednotka*  
*přesnost*  
*zdroj*

#### **Definice datového prvku s kontextem**

#### **Název datového prvku s kontextem**

#### **Identifikátor datového prvku a číslo verze**

#### **Správní atributy**

registrační status  
předkládající organizace  
správce dat

## Koncept datového prvku, název konceptu, definice, identifikátor

### Název a definice domény konceptu

### Přípustné hodnoty

### Příklad

### Referenční dokument (původ)

K ilustraci použijí datový prvek ze Standardu ISO 3166 - Kódy pro prezentaci názvů

zemí - **Krátký anglický název země.**

<b>Atribut datového prvku:</b>	<b>hodnota</b>
třída prezentace	název
název hodnotové domény	Krátké názvy zemí v angličtině
definice domény	Všechny krátké anglické názvy zemí
datový typ	alfanumerický
typ domény	výčtová
extrémní hodnoty (délka)	minimálně 4 znaky (Peru) maximálně 44 znaků
formát	A(44), A=alfanumerický typ, 44 je přizpůsobeno nejdelšímu názvu ISO 3166
zdroj	Krátký název země prezentovaný v anglickém jazyce.
definice datového prvku	Krátký anglický název země (termíny „krátký“ a „anglický“ jsou kvalifikátory)
název datového prvku	Identifikátor země
koncept datového prvku - název	Identifikátor pro primární geopolitickou entitu.
koncept datového prvku - definice	Země světa
doména konceptu - název	Primární geopolitické entity
doména konceptu - definice	Názvy uvedené v ISO 3166 pod příslušným záhlavím.
přípustné hodnoty	Dánsko

K registračnímu statutu prvku: datový prvek může být *nekompletní, certifikovaný, standardní, vyřazený*, apod.

Tento standard nerozlišuje povinné, podmíněné nebo volitelné atributy.

Z příkladu je patrné, jak je důležité názvosloví datových prvků a formulace definic.

Přesná a jednoznačná definice datového prvku je nezbytná pro zajištění možnosti jeho sdílení.

Tvorba názvů datových prvků a formulace definic je samostatná kapitola, částečně regulovaná standardem ISO 11179, *část 4 - Pravidla pro formulaci datových definic a část 5 - Názvoslovné a identifikační principy datových prvků.*

### Literatura:

ČSN ISO/IEC 11179, Informační technologie - Specifikace a standardizace datových prvků, část 3 Základní atributy datových prvků, 1996, ČNI Praha

ISO/IEC NWI xxx, Information Technology, Procedures for Achieving Data Registry Content Consistency, Working paper Draft 2.1, 1999