

# Lekce 7 – Kvalita dat

1	Cíle lekce.....	1
2	Kvalita vstupních dat .....	1
2.1	Hledisko mikroúrovně.....	1
2.2	Hledisko makroúrovně .....	1
2.3	Hledisko uživatelské úrovně .....	1
3	Řízení kvality dat (procesy QA/QC) .....	2
4	Příklady funkcí kontrolujících kvalitu dat .....	4

## 1 Cíle lekce

- Podat informace o kritériích kvality prostorových dat
- Popsat typických proces aktualizace s kontrolou kvality
- Uvést příklady kontrolních funkcí

## 2 Kvalita vstupních dat

Kvalitu vstupních dat lze posuzovat z několika hledisek.

### 2.1 Hledisko mikroúrovně

Kvalitu dat lze posoudit lokálně – analýzou jednoho objektu a jeho okolí. Příklady kritérií kvality této úrovně:

- Přesnost
  - Polohová a výšková – přesnost určení souřadnic (očekávaná odchylka souřadnic od skutečné polohy)
  - Časová přesnost – aktuálnost
  - Tematická přesnost - přesnost určení atributů (hodnota atributu odpovídá číselníku nebo se nachází v zadaném intervalu)
- Konzistence a integrita (jednoznačné relace mezi daty)
- Měřítko - rozlišení dat (nejmenší reprezentovatelná jednotka, nejmenší objekt zobrazený na mapě)

### 2.2 Hledisko makroúrovně

Kvalitu dat lze posoudit pouze analýzou celé datové sady. Příklady kritérií kvality této úrovně:

- Úplnost
  - úplnost pokrytí území
  - úplnost klasifikace (jak kvalitně zvolená klasifikace reprezentuje data)
  - úplnost verifikace dat
- Definice dat (existence a kvalita metadat), rodokmen dat (historie dat - zdroje a postup zpracování dat)
- Jazyková kvalita
- Projekce - kartografické zobrazení

### 2.3 Hledisko uživatelské úrovně

Kvalita dat z pohledu uživatele. Příklady kritérií:

- přístupnost dat
  - privátní data, veřejné databáze, ochrana údajů
  - interoperabilita dat
- přímá cena dat
- nepřímá cena dat (nepřímá cena dat způsobený vynuceným přepracováním nekvalitních dat může být vyšší než cena kvalitních dat)

### 3 Řízení kvality dat (procesy QA/QC)

QA - Quality assurance (zabezpečení kvality) je proces, kterým je měřena kvalita dat a ve kterém jsou implementovány prostředky pro zajištění kvality dat

QC - Quality control (řízení kvality) je proces, který zabezpečuje kvalitu dat v uživatelem očekávané úrovni

Pro zajištění procesu QA/QC disponují systémy kontrolními funkcemi pro kontrolu prostorových a popisných dat. Systémy umožňují provádění těchto kontrol

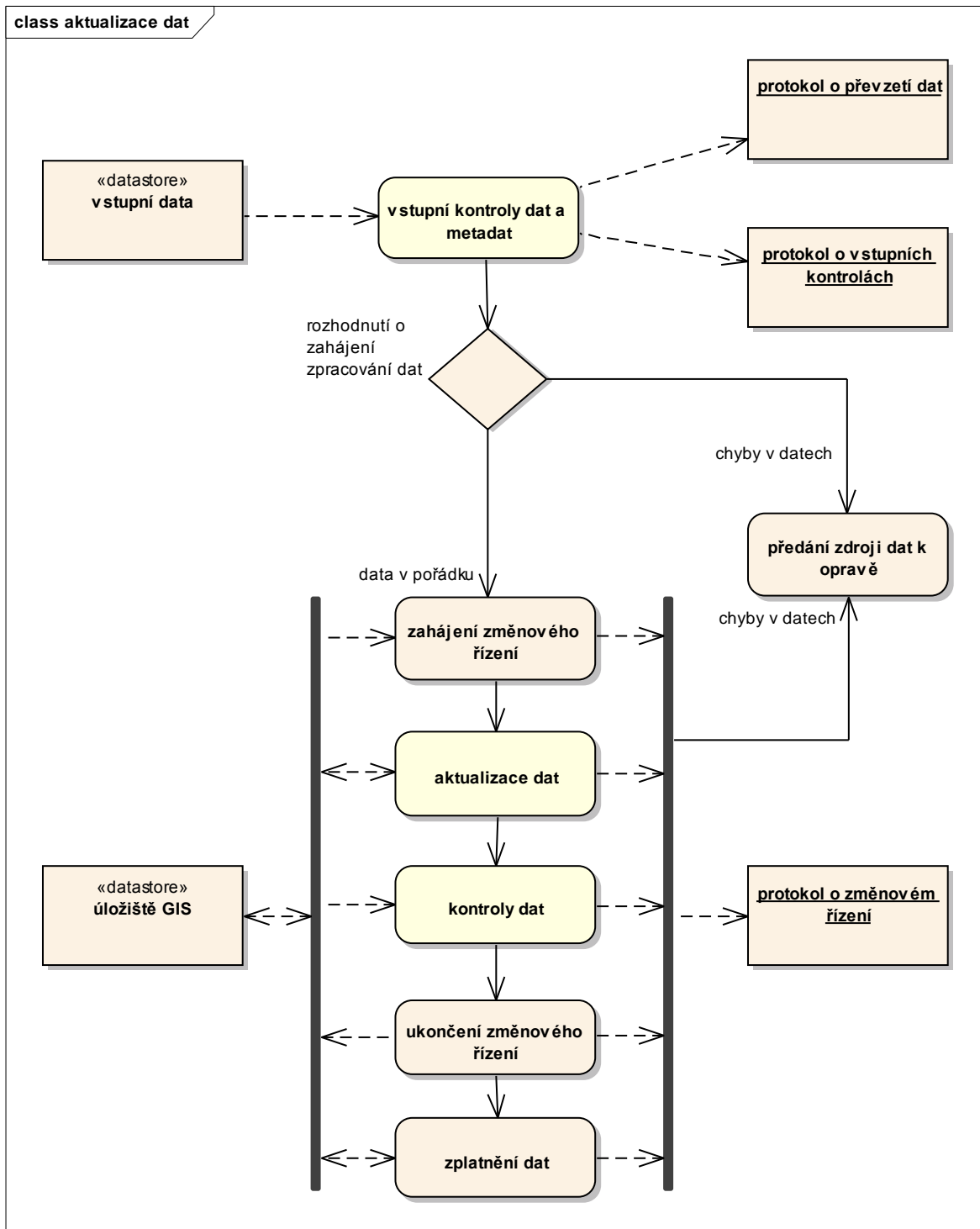
- při pořízení dat (speciálním případem jsou dynamické kontroly dat ala kontrola pravopisu a podtrhávání chyb v textových editorech
- při ukončení transakce
- dávkově v pravidelných intervalech nebo na vyžádání uživatele

*Průběh* kontrolních funkcí je řízen parametrickými předpisy, výsledky kontrolních funkcí jsou vizualizovány (vytváří se datové množiny obsahující chybové objekty nebo značky definující polohu chyby). Kontrolní funkce zásadně nemodifikují žádná data. Funkce, které upravují chybná data, se spouštějí v samostatných transakcích. Před zplatněním transakcí opět proběhnou kontroly dat.

Základní pracovní postup provádění kontrol obsahuje kroky:

- výběr dat, na která se mají aplikovat kontroly
- výběr typů kontrol, které se mají provést a případné nastavení parametrů
- výběr kontextových dat pro některé typy kontrol
- vyhodnocení výsledků kontrol
- případná oprava dat určených kontrolním nástrojem jako chybová data

## Typický proces aktualizace dat, jehož součástí jsou kontroly dat



## 4 Příklady funkcí kontrolujících kvalitu dat

Příklady kontrolních funkcí jsou převzaty z dokumentace ZABAGED.

### Kontrola nadbytečného rozdělení liniových objektů

Význam kontroly:

Vyhledává liniové objekty stejného feature typu, které jsou rozděleny na více napojujících se linií, aniž je důvod k jejich rozdělení podle definovaných pravidel (v místech, kde dochází ke změně hodnoty atributu nebo v místě, kde se v lomovém bodě stýkají více jak 2 liniové objekty - "T napojení prvků").

Kontrolované objekty:

Kontrolují se specifikované liniové objekty. Předpokládá se, že se budou kontrolovat všechny feature typy, které mají geometrický typ definovaný jako "LINE".

Bude možné pustit kontroly pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění 2 bodů.
- Výčet kontrolovaných feature typů.

Ke každému kontrolovanému feature typu pak lze dále definovat:

- Množinu atributů jejich sémantické tabulky, jejichž hodnoty musí být shodné proto, aby se rozdělení linií vyhodnotilo jako nadbytečné.
- Množinu feature typů liniových objektů, které se mohou napojovat ve styčném bodě kontrolovaných linií, aniž by to znamenalo důvod k rozdělení kontrolovaných linií. Bude možno uvést i negaci seznamu.
- Množinu feature typů bodových objektů, které mohou ležet ve styčném bodě kontrolovaných linií, aniž by to znamenalo důvod k rozdělení kontrolovaných linií. Bude možno uvést i negaci seznamu.

Výsledek kontroly:

Kontrola umístí do chybové množiny značku chyby umístěnou v místě napojení nadbytečně rozdělených linií.

### Kontrola nutného rozdělení liniových objektů

Význam kontroly:

Vyhledává liniové objekty, které v jejich průběhu kdekoli kříží (i bez společného vrcholu) jiný liniový objekt daný množinou feature typů, případně ve vnitřním vrcholu leží nějaký bodový objekt, což je důvodem k rozdělení kontrolované linie v daném vnitřním vrcholu. Jsou definovány vztahy mezi jednotlivými feature typy, které určují pravidla pro vyhodnocení nutnosti "rozdělení".

Kontrolované objekty:

Kontrolují se specifikované liniové objekty. Bude možné pustit kontroly pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění 2 bodů.
- Výčet kontrolovaných feature typů.

Ke každému kontrolovanému feature typu pak lze dále definovat:

- Množinu atributů jejich sémantické tabulky, jejichž hodnoty musí být shodné proto, aby se rozdělení linií vyhodnotilo jako nadbytečné.
- Množinu feature typů liniových objektů, které se mohou napojovat ve styčném bodě kontrolovaných linií, aniž by to znamenalo důvod k rozdělení kontrolovaných linií (bude možné uvést i "negaci" seznamu).
- Parametr určující, zda důvodem k rozdělení kontrolované linie je napojení jiné linie vnitřními vrcholy nebo jen napojení jiné linie počátečním resp. koncovým vrcholem.

- Množinu feature typů bodových objektů, které mohou ležet ve styčném bodě kontrolovaných linií, aniž by to znamenalo důvod k rozdělení kontrolovaných linií (bude možné uvést i "negaci" seznamu).

Výsledek kontroly:

Kontrola umístí do chybové množiny značku chyby umístěnou v místě, kde má být provedeno rozdělení linie.

### **Kontrola volných konců liniových objektů**

Význam kontroly:

Vyhledává liniové objekty, na jejichž počáteční nebo koncový vrchol není napojen jiný povolený liniový objekt, případně povolený bodový objekt.

Za napojení lze ve specifikovaných případech považovat i napojení do vnitřního vrcholu jiné linie.

Kontrolované objekty:

Kontrolují se specifikované liniové objekty. Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění dvou bodů.
- Výčet kontrolovaných feature typů.
- Výčet feature typů, které definují objekt "použitelné" pro napojení kontrolovaného prvku.
- Příznak určující, zda kontrolovaná linie může/nemůže končit ve vnitřním vrcholu jiné linie.
- V případě možnosti ukončení ve vnitřním vrcholu, pak výčet liniových feature typů, ve kterých se může kontrolovaný objekt napojovat do vnitřního vrcholu.

Výsledek kontroly:

Kontrola umístí do chybové množiny liniový objekt a značku chyby v místě volného konce.

### **Kontrola prostorových vztahů mezi objekty**

Význam kontroly:

Kontroluje prostorový vztah (prostorovou souvztažnost) objektů oproti předepsaným pravidlům. Především se jedná o kontrolu prostorových vztahů bodových nebo liniových objektů vůči liniovým objektům. Pokud daný typ bodového nebo liniového objektu neleží na specifikovaném výčtu typů liniových objektů, pak je vyhodnocen jako chybový. Typ požadovaného prostorového vztahu je určen parametrickým předpisem. V rámci vyhodnocování chyb se zohledňuje definované okolí grafických prvků. Součástí této kontroly může být také kontrola shodnosti hodnot určených atributů pro objekty v definovaném prostorovém vztahu.

Kontrolované objekty:

Kontrolují se specifikované bodové nebo liniové objekty. Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění 2 bodů.
- Výčet kontrolovaných feature typů.

Pro každý kontrolovaný feature typ lze dále určit:

- Výčet feature typů liniových objektů, na kterém může být umístěn (v případě, že se kontroluje liniový objekt pak musí být umístěn všemi jeho vrcholy a počet vrcholů v souběhu musí souhlasit).
- Výčet dalších feature typů liniových objektů, na kterých musí mít umístěný alespoň 1 vrchol (v případě, že se kontroluje liniový objekt).
- Příznak určující, zda kontrolovaný objekt může/nemůže ležet na vnitřním vrcholech nebo jen koncových vrcholech liniových prvků.
- Pokud se má kontrolovat hodnota popisného atributu, pak se určuje název atributu kontrolovaného feature typu a název atributu feature typu, který vytváří daný prostorový vztah.

Výsledek kontroly:

Kontrola umístí do chybové množiny bodové nebo liniové objekty, které nesplňují požadovaný prostorový vztah k liniovému objektu.

### **Kontrola chybného překrývání malých plošných objektů**

Význam kontroly:

Vyhledává prvky mající charakter menších plošných objektů a které se svojí obvodovou hranicí vzájemně překrývají. Charakter kontrolovaných prvků je dán také jejich geometrickým vyjádřením prvků, a to uzavřeným polygonem (který je pořízen jako 1 grafický prvek).

Kontrola také vyhledává umístění jednoho plošného útvaru v druhém plošném útvaru ("budova v budově").

Kontrolované objekty:

Kontrolují se specifikované plošné prvky, které mají charakter menších plošných objektů.

Parametrizace:

- Výčet kontrolovaných feature typů.
- Pro každý kontrolovaný feature typ lze určit výčet jiných feature typů plošných prvků, které buď překrývají / jsou uvnitř kontrolovaného prvku a přesto se tato situace nebude považovat za chybovou (např. uvnitř budovy může být obvod vnitrobloku).

Výsledek kontroly:

Kontrola umístí do chybové množiny takové prvky, které vykazují chybu.

### **Kontrola uzavřenosti plochy a jednoznačného přiřazení centroidu**

Význam kontroly:

Vyhledává liniové objekty reprezentující hranice definovaných ploch, které netvoří uzavřené polygony. Vyhledává také liniové objekty, které se kříží s jinými liniovými objekty za předpokladu, že tyto liniové objekty reprezentující hranice definovaných ploch. Následně kontroluje počet centroidů v uzavřených polygonech seskládaných z jednotlivých liniových objektů. V případě, že se kontrola provádí v rámci změnového řízení a během aktualizace dat byly modifikovány objekty reprezentující hranice ploch nebo centroidy, pak dojde nejdříve k aktualizaci (přegenerování) ploch a aktualizaci příslušných topologických struktur. Výsledné přegenerované plochy bude možné zobrazit pomocí standardních funkcí klientského prostředí, což umožní i případnou vizuální kontrolu ploch.

Kontrolované objekty:

Kontrolované feature typy budou ty, které reprezentují všechny objekty reprezentující hranice ploch nebo centroidy ploch.

Parametrizace:

Příznak určující jaké podtypy kontrol se budou provádět:

- Vyhledání liniových objektů, které netvoří uzavřené polygony.
- Vyhledání ploch bez centroidu.
- Vyhledání ploch s více centroidy.

Výsledek kontroly:

Kontrola umístí do chybové množiny liniové objekty, které netvoří uzavřené polygony. Do chybové množiny se budou také ukládat plochy, které nemají jednoznačně přiřazený centroid.

### **Kontrola vyhledání sousedních ploch stejného typu (významu)**

Význam kontroly:

Vyhledává sousedící plochy stejného feature typu resp. plochy stejného významu daného určeným atributem. Kontrola se de facto provádí nad liniovými objekty, které tvoří hranice ploch.

Kontrolované objekty:

Kontrolují se specifikované plošné objekty, jejichž hranice jsou "seskládané" z jednotlivých liniových objektů.

Předpokládá se, že se budou kontrolovat všechny feature typy, které budou tvořit plošnou topologii. Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Výčet kontrolovaných feature typů.
- Pro každý kontrolovaný feature typ lze určit atributy popisných dat, jejichž hodnota představuje význam plochy.
- Výčet feature typů objektů reprezentujících hranice ploch, které mohou dělit plochy stejného typu a přesto tato situace nebude považována za chybovou.

Výsledek kontroly:

Kontrola umístí do chybové množiny značku liniové objekty, které tvoří hranice sousedních ploch stejného významu.

### **Kontrola nepovolených prostorových kombinací objektů**

Význam kontroly:

Vyhledává nepovolené kombinace prostorového umístění bodových, liniových nebo "malých" plošných objektů uvnitř daných ploch. Plochy zde budou reprezentovány plochami generovanými z jednotlivých liniových hranic.

Kontrolované objekty:

Kontrolují se specifikované plochy. Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Výčet kontrolovaných feature typů reprezentující plochu.
- Pro každý kontrolovaný feature typ bude možné určit výčet feature typů bodových, liniových nebo "malých" plošných objektů, které nesmí ležet uvnitř kontrolované plochy a pro každý takový feature typ bude také určeno, zda je za chybu považováno to, že leží uvnitř plochy celý nebo je za chybu považováno to, když leží uvnitř plochy jen jeho část.

Výsledek kontroly:

Kontrola umístí do chybové množiny bodové, liniové nebo "malé" plošné objekty vyhodnocené jako chybové.

### **Kontrola návaznosti liniových objektů včetně jejich atributů**

Význam kontroly:

Vyhledává liniové objekty, které jsou stejného typu a jsou vzájemně napojeny, ale u kterých se liší hodnota specifikovaného popisného atributu. Současně se může kontrolovat i fakt, zda navazující linie se stejnými hodnotami atributů tvoří uzavřený útvar (např. pro hranici chráněného území).

Kontrolované objekty:

Kontrolují se specifikované liniové objekty. Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění 2 bodů.
- Výčet kontrolovaných feature typů.
- Pro každý kontrolovaný feature typ bude možné určit množinu atributů popisných dat, jejichž hodnoty mají být v rámci návaznosti linií stejné.
- Příznak určující, zda kontrolované linie mají tvořit uzavřený útvar.

Výsledek kontroly:

Kontrola umístí do chybové množiny liniové objekty, které vykazují chybu. Pravidla pro změnu hodnoty popisného atributu nelze vždy přesně algoritmicky predikovat, proto mnohdy půjde jen o "potenciální" chybu.

### **Kontrola prostorově vzájemně souvisejících objektů**

Význam kontroly:

Kontrola vyhledává liniové objekty, které mohou prostorově souviset s jinými typy liniových objektů. Prostorovou souvislostí je zde myšlen přesný souběh linií nebo částí linií.

Kontrolované objekty:

Kontrolují se specifikované liniové objekty.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění 2 bodů.
- Výčet kontrolovaných feature typů.
- Výčet feature typů, které reprezentují potenciálně prostorově souvisejícím liniové objekty.
- Vzdálenost, ve které se vyhledávají prostorově související objekty.

Výsledek kontroly:

Kontrola umístí do chybové množiny liniové objekty, které jsou vyhodnoceny jako chybové.

### **Kontrola duplicity liniových objektů**

Význam kontroly:

Vyhledává liniové objekty, které jsou prostorově totožné (leží na sobě), případně jejichž některá část je prostorově totožná s částí jiného liniového objektu.

Kontrolované objekty:

Kontrolují se specifikované liniové objekty.

Předpokládá se, že se budou kontrolovat všechny feature typy, které mají geometrický typ definovaný jako "LINE". Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění 2 bodů.
- Výčet kontrolovaných feature typů.
- Příznak určující, zda "duplicitní pár" musí tvořit objekty se stejným feature typem nebo zda jej mohou tvořit i objekty s různým feature typem (ve druhém případě se jako další parametr přidává výčet takových feature typů).
- Příznak určující, zda prostorově duplicitní linie musí/nemusí mít shodnou orientaci.
- Parametr určující u každého kontrolovaného liniového feature typu, jakým způsobem se vyhodnocuje prostorová duplicita:
- Celek - duplicita je vyhodnocena jen v případě, že celý průběh liniových objektů je totožný.
- Více segmentů - duplicita je vyhodnocena v případě, že 2 a více segmentů (úseček) liniových objektů jsou totožné.
- 1 Segment - duplicita je vyhodnocena již v případě, že 1 segment (úsečka) liniových objektů je totožný.

Výsledek kontroly:

Kontrola umístí do chybové množiny duplicitní liniové objekty a značku chyby umístěnou v místě těch vrcholů linií, které jsou prostorově totožné u obou linií.

### **Kontrola duplicity bodových objektů**

Význam kontroly:

Vyhledává bodové objekty, které jsou prostorově totožné (leží na sobě).

Kontrolované objekty:



Kontrolují se specifikované bodové objekty. Předpokládá se, že se budou kontrolovat všechny feature typy, které mají geometrický typ definovaný jako "POINT". Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Vzdálenost pro prostorové ztotožnění 2 bodů.
- Výčet kontrolovaných feature typů.
- Příznak určující, zda "duplicitní pár" musí tvořit objekty se stejným feature typem nebo zda jej mohou tvořit i objekty s různým feature typem.

Výsledek kontroly:

Kontrola umístí do chybové množiny značku chyby v místě vztažného bodu bodového objektu.

### **Kontrola minimální povolené velikosti objektů**

Význam kontroly:

Vyhledává liniové a plošné objekty, jejichž velikost (délkový rozměr nebo plocha) je menší než povolená hodnota. U liniových objektů a plošných objektů, jejichž hranice jsou tvořeny liniovými objekty lze kontrolovat i minimální povolenou velikost jednotlivých segmentů (úseček).

Kontrolované objekty:

Kontrolují se specifikované liniové a plošné objekty. Bude možné pustit kontroly i pro jednotlivě definované skupiny feature typů.

Parametrizace:

- Výčet kontrolovaných feature typů.
- Parametr určující, zda se kontroluje plocha, celková délka linie nebo délka jednotlivých úseček, případně povolený poměrový vztah plocha vs. minimální velikost hranice plochy.
- Parametr obsahující minimální povolenou hodnotu velikosti (poměru).

Výsledek kontroly:

Kontrola umístí do chybové množiny takové objekty, které vykazují chybu.

### **Kontrola hodnot popisných atributů**

Význam kontroly:

Kontrola je zaměřená na vyhledání objektů, jejichž hodnoty jsou mimo definovaný rozsah hodnot nebo jejichž hodnoty nejsou vyplněné.

U některých popisných atributů bude zadání hodnoty v rámci povoleného rozsahu případně zadání povinného atribut řízeno přímo definicí metamodelu projektu nebo na úrovni definice struktury databázových objektů.

Kontrolované objekty:

Kontrolují se specifikované atributy s popisnými daty u daných feature typů.

Parametrizace:

Programově budou kontroly zřejmě rozděleny na dva základní typy, kterým bude poskytnut jiný výčet parametrů.

Pro kontrolu na vyhledání atributů, jejichž hodnoty nejsou vyplněné:

- Výčet kontrolovaných feature typů a jejich atributů, které nesmí být prázdné (musí být vyplněné).

Pro kontrolu na vyhledání atributů, jejichž hodnoty jsou mimo povolený rozsah:

- Výčet kontrolovaných feature typů a jejich atributů, které se mají kontrolovat.
- Příznak určující, zda prázdná hodnota je pro hodnotu atributu akceptovatelná.
- Parametr určující typ operátoru pro vyhodnocení hodnot (=, >, <, !=, IN, BETWEEN, LIKE).
- Hodnoty (hodnoty) určující výčet nebo rozsah povolených hodnot.

Výsledek kontroly:

Datový výběr obsahující prostorová a popisná data pro objekty vyhodnocené jako chybové.

Připomínky a dotazy k obsahu lekce posílejte, prosím, na adresu:  
Rudolf Richter, [richter@fi.muni.cz](mailto:richter@fi.muni.cz)