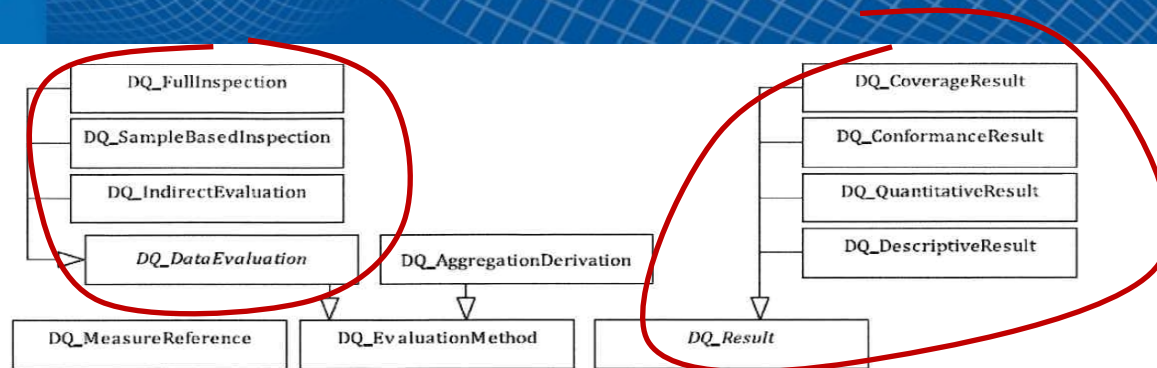


KVALITA (PROSTOROVÝCH) DAT

ISO 19157 Geografická informace – kvalita dat

- Soubor základních metodik hodnocení kvality (prostorových) dat v kartografii, geoinformatice i dálkovém průzkumu Země
- Deklarována také jako ČSN ISO 19157
 - verze ISO 19157:2013/Amd.1:2018
- Nahrazuje původní normy (ČSN) ISO 19113:2005 a ISO 19114:2005
 - oproti nim je výrazně přehlednější, s vyšším důrazem na příklady

Komponenty a elementy kvality dat



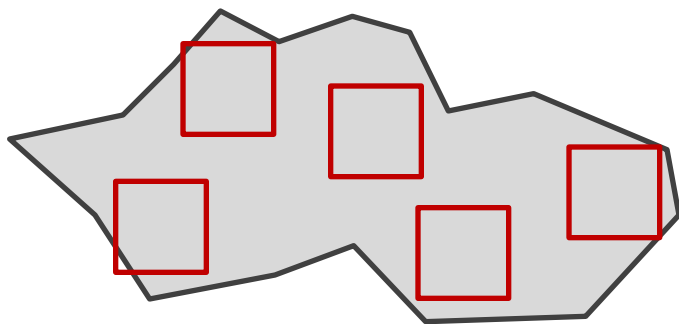
Elementy
kvality dat

Strategie vzorkování pro hodnocení kvality dat

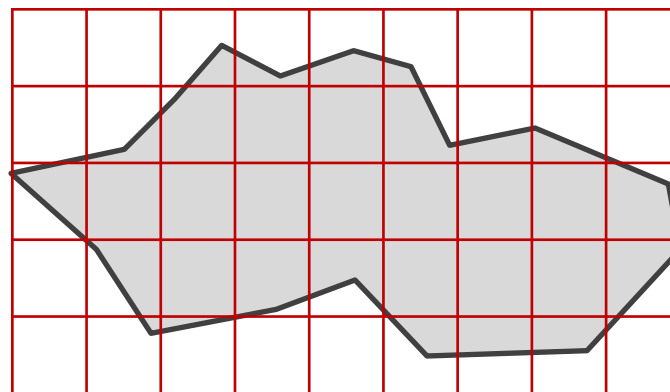
Pokud neanalyzují celý statistický soubor.

Počet prvků z celkového počtu vybraných pro hodnocení (například každý pětistý prvek podle ID).

jednoduché/stratifikované náhodné vzorkování ploch



nenáhodné vzorkování pravidelnými plochami



vzorkování založené na pravděpodobnosti

- musím znát pravděpodobnost výběru
- detaily ISO 2859

...

Deklarace výsledků

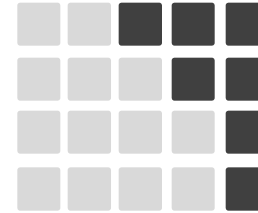
DQ_Result

DQ_CoverageResult

DQ_ConformanceResult

DQ_QuantitativeResult

DQ_DescriptiveResult



Grid s buňkami zachycujícími střední kvadratickou chybu datové sady.

Typ prvku Heliporty, 5. třída měřické přesnosti, podle ČSN 01 3410 (2014)

Instance prvku s ID7841 2D $e_i = 1,73 \text{ m}$

Typ prvku výšk. bod $\sigma_z (95\%) = 12 \text{ cm}$

První odraz laserového skenování obecně dosáhl nižší absolutní polohové přesnosti ve 3D (střední hodnota nejistot polohy) oproti ...

Míry kvality dat

- Konkrétní metodické pokyny, tzv. procedury, jak zjistit (vypočítat) daný element kvality dat
- Vždy se řeší, na jaké úrovni mohou (mají) být aplikovány, např.
 - série datových sad
 - datová sada
 - podmnožina datové sady
 - typ prostorového prvku
 - instance prostorového prvku
 - typ atributu
 - instance atributu

Míry kvality dat

Tabulka D.30 – Vychýlení poloh

Řádek	Komponenta	Popis
1	Jméno	vychýlení poloh (1D, 2D, 3D)
2	Náhradní jméno	
3	Jméno prvku	absolutní nebo vnější přesnost
4	Základní míra	nelze použít
5	Definice	vychýlení poloh pro množinu poloh, v nichž nejistoty poloh jsou definovány jako odchylka mezi měřenou polohou a tím, co je považováno za odpovídající skutečnou polohu
6	Popis	<p>Pro určitý počet bodů (N) jsou měřené polohy dány jako x_{mi}, y_{mi} a z_{mi} souřadnice v závislosti na rozměru, ve kterém je poloha bodu měřena. Odpovídající množina souřadnic x_{yi}, y_{yi} a z_{yi} se považují za skutečnou polohu. Odchylky a vychýlení jsou počítány jako:</p> <p>Jednoduché odchylky:</p> $e_{xi} = x_{mi} - x_{yi}$ $e_{yi} = y_{mi} - y_{yi}$ $e_{zi} = z_{mi} - z_{yi}$ <p>Vychýlení (zkreslení):</p> $a_x = \frac{\sum e_{xi}}{N_x}$ $a_y = \frac{\sum e_{yi}}{N_y}$ $a_z = \frac{\sum e_{zi}}{N_z}$ $a_p = \sqrt{a_x^2 + a_y^2}$ $a_{3D} = \sqrt{a_x^2 + a_y^2 + a_z^2}$ <p>Mělo by být stanoveno kritérium zavádějící soulad (např. povolující soulad k nejbližší pozici, soulad vrcholů nebo soulad podél linií). Kritérium/kritéria pro nalezení vzájemně korespondujících bodů musí být rovněž vykázána společně s výsledkem hodnocení kvality dat.</p>
7	Parametr	
8	Typ hodnoty	míra
9	Struktura hodnoty	-
10	Odkaz na zdroj	-
11	Příklad	-
12	Identifikátor	128

Tabulka D.10 – Počet položek, které nejsou ve shodě s pravidly konceptuálního schématu

Řádek	Komponenta	Popis
1	Jméno	počet položek ve shodě s pravidly konceptuálního schématu
2	Náhradní jméno	-
3	Jméno prvku	konceptuální bezspornost
4	Základní míra	celkový počet chyb
5	Definice	celkový počet všech položek v množině dat, které nejsou ve shodě s pravidly konceptuálního schématu
6	Popis	Pravidla, pokud jsou přímo nebo nepřímo popsána konceptuálním schématem, musí být sledována. Příkladem nedodržení těchto pravidel může být chybné umístění vzhledu v rámci definované tolerance, duplikace vzhledů a chybné překrývání vzhledů.
7	Parametr	-
8	Typ hodnoty	Integer
9	Struktura hodnoty	-
10	Odkaz na zdroj	-
11	Příklad	<p>Příklad 1: Věže s identickým atributem a ve vyhledávací toleranci (tolerance vyhledání = 10 m.)</p>  <p>Příklad 2: Most není povoleno využívat pro transport. Využije se kategorie silnice.</p> 

Přehled základních výpočtů měr kvality dat

Jméno základní míry kvality dat	Definice základní míry kvality dat	Příklad	Hodnotový typ kvality dat
Indikátor chyby	Indikátor položky je chybný	false	Boolean (hodnota 'true' značí, že jednotka není správná)
Indikátor správnosti	Indikátor položky je správný	true	Boolean (hodnota 'true' značí, že jednotka je správná)
Počet chyb	Celkový počet položek, které jsou předmětem nějaké chyby specifického typu	11	Integer
Počet správných položek	Celkový počet položek, které nevykazují chyby specifického typu	571	Integer
Poměr chybných položek	Počet chybných položek s ohledem na celkový počet položek, které by měly být uvedeny	0,0189	Real [procenta nebo poměr]
Poměr správných položek	Počet správných položek s ohledem na celkový počet položek, které by měly být uvedeny	0,9811	Real [procenta nebo poměr]

Úplnost (DQ_Completeness)

- Boolean, počet či podíl nekorektně zastoupených dat

- tj. **nad plánovaný rámec**
(DQ_CompletenessCommission)

Datová sada obsahuje 5 národních parků, ale v dané oblasti existují pouze 4.

nebo

- **chybějící data oproti záměru**
(DQ_CompletenessOmission)

V datové sadě chybí 40% všech biotopů zjištěných při terénním průzkumu.

- Způsoby zjišťování, např.

- zpětné vazby od uživatelů

14 adresních bodů v datové sadě je duplicitních.

- vnitřní revize průběžných kontrol (tj. kolik prvků vrátil supervisor operátorům)

Logická konzistence (DQ_LogicalConsistency)

Konceptuální konzistence (DQ_ConceptualConsistency)

- Dodržení pravidel konceptuálního schématu
- Implicitně se nastavuje s hodnotou „true“
- Hodnocení pomocí křížové tabulky prvků databáze
 - popis přípustnosti všech stavů vzájemného umístění prvků databáze (bodů, čar, ploch a jejich atributů) v rámci jejich dotyku, duplikace, tolerance přesnosti, překrytu a logické návaznosti těchto prvků

neplatné překrytí
železnice a jezera



Logická konzistence (DQ_LogicalConsistency)

Doménová konzistence (DQ_DomainConsistency)

- Obvykle „jen“ zda instance prostorového prvku je/není ve shodě s doménou hodnot
- Úzce souvisí se
 - způsobem sběru dat a
 - definicí aplikačního schématu

false, nebyla detekována žádná chyba – v množině dat jsou uvedeny pouze (povinné) typy prostorových prvků z aplikačního schématu

Logická konzistence (DQ_LogicalConsistency)

Formátová konzistence (DQ_FormatConsistency)

- Obvykle hodnoceno pouze na úrovni fyzické struktury
 - Jsou data v požadových formátech?
 - Často proto implicitní hodnota „true“

false: data jsou publikována jako shapefile oproti definovanému GeoPackage

5 souborů používá kódování Windows-1250 oproti definovanému UTF-8

Logická konzistence (DQ_LogicalConsistency)

Topologická konzistence (DQ_TopologicalConsistency)

- Obecně je doporučováno relativní (poměrové) vyjádření chyb v těch případech, kdy je to možné (i ve 3D)

- Poměr vadných spojení bod-křivka
- Počet chybějících spojení z důvodu nedotažení
- Počet chybějících spojení z důvodu přetažení
- Počet neplatných ploch štěrbin
- Počet chyb neplatných sebeprůseků
- Počet neplatných sebepřekryvů



Polohová přesnost (DQ_PositionalAccuracy)

Absolutní nebo vnější přesnost (DQ_AbsoluteExternalPositionalAccuracy)

- Definice tzv. ground truth (např. geodetická měření)
 - Soulad s hodnotami, které jsou považovány za správné
- Značné množství měř kvality dat (viz ISO 19157)
- Pro 1D, 2D i 3D, vektorová i gridová data
- Klíčové pro mapové/datové produkty NATO

Absolutní přesnost: dvojice odpovídajících bodů

● lokalizace prvku podle geodetického měření

● lokalizace prvku v datové sadě

Laboratoř geoinformatiky a kartografie

Relativní přesnost: jak konzistentní je např. polohová přesnost na mapovém listu



Tematická přesnost (DQ_ThematicAccuracy)

Správnost klasifikace (DQ_ThematicClassificationCorrectness)

- Porovnání tříd přiřazených k prvkům nebo atributům
- Komplexní, ale náročná matice chybných klasifikací
 - zpětná vazba od uživatelů, případně obohacené o hodnotu vnitřní revize (supervisor → operátor)

	Klasifikační třída v datové sadě				
Klasifikační třída v realitě		železnice	silnice	dálnice	SUMA
	železnice	7	2	1	10
	silnice	1	2	2	5
	dálnice	1	1	3	5
	SUMA	9	5	6	20

Tematická přesnost (DQ_ThematicAccuracy)

Správnost nekvantitativních atributů (DQ_NonQuantitativeAttributeAccuracy)

- Oproti správnosti klasifikace se vyjadřuje jen jako sumární absolutní či relativní informace
 - zpětná vazba od uživatelů, případně obohacené o hodnotu vnitřní revize (supervisor → operátor)

5 geografických jmen v datové sadě nepoužívá oficiální geografické názvosloví

5 z 10 000

0,05%

Tematická přesnost (DQ_ThematicAccuracy)

Přesnost kvantitativních atributů (DQ_QuantitativeAttributeAccuracy)

- Většinou není řešena pro celou datovou sadu
 - Stanovení reprezentativních vzorků
- Typicky řešení nejistoty hodnot atributů na určité hladině významnosti

Uvedení šířky vodního toku 2,15 metru odpovídá nejistotě hodnot atributů na 68,3% hladině významnosti.

Na hladině významnosti 99% je výška stromu definována s odchylkou maximálně 0,3 metru.

Časová kvalita (DQ_TemporalQuality)

- Přesnost měření času

- Standardně na 50%, 68,3%, 90%, 95%, 99% a 99,8% hladině významnosti

Atribut HLUK_MER vykazuje přesnost měření času ± 5 sekund na hladině významnosti 95%.

- Časová bezespornost (konzistence)

true, 150 údajů z měření NO_x je uvedeno v datové sadě, ale nejsou správně chronologicky uspořádány.

- Časová platnost

Pro 25% predikovaných prostorových prvků vypršela časová platnost.

Použitelnost (DQ_UsabilityElement)

- Samostatná zpráva o kvalitě dat, která je založena na **požadavcích uživatele**
- Může se opírat o kterýkoli prvek kvality i být jakkoli rozšířeno
 - například na základě specifických uživatelských požadavků
 - nicméně se doporučuje používat deskriptory prvků kvality jako u ISO 19157
- Výsledkem je deklarace shody množiny dat se speciálním určením

Routovatelná datová sada pro navigační účely nákladních automobilů s parametrickým nastavením výšky vozidla...
polohová přesnost silničních segmentů ve 2D $e_i = 0,25$ m
podjezdová výška mostu ID 7832 s přesností σ_z (95%) = 0,10 cm

Závěrem

Doporučený postup (tzv. top-down approach):

1. Definuje se účel, co a proč je třeba ověřit z hlediska kvality dat.
2. V souladu s účelem se vybírají konkrétní komponenty kvality dat – od obecných prvků kvality ke konkrétním mírám kvality.
3. Vybírá se mezi absolutním, relativním, maticovým či jiným komplexním vyjádřením výsledků podle účelu.
4. Aplikuje se metodika (výpočet) z dané míry dat a procedury na reprezentativním vzorku dat nebo celku.
5. Výsledky se publikují jako technická zpráva, v registru a v metadatech, poté vše sémanticky propojím.