

# OPREDELITEV KAKOVOSTI PROSTORSKIH PODATKOV

doc.dr. Radoš Šumrada

FGG-Oddelek za geodezijo, Ljubljana

Prispelo za objavo: 1996-06-20

Pripravljeno za objavo: 1996-06-20

## Izvleček

*V članku sta predstavljena pomen in vloga osnovnih pokazateljev za določanje kvalitete prostorskih podatkov. Opisani elementi so pojmovno opredeljeni v kakovostnem modelu in predstavljani v poročilu o kakovosti, ki je ponavadi obvezen sestavni del standardov za izmenjavo prostorskih podatkov. Primerjalno so določeni tudi identifikatorji kakovosti med poročilom komisije za kakovost prostorskih podatkov ICA, ameriškim standardom SDTS ter predlogom novega evropskega standarda tehničnega odbora CEN/TC 287.*

*Ključne besede: elementi, opredelitev kakovosti, osnovni pojmi, prostorski podatki, standardi*

## Abstract

*This paper presents the role and importance of crucial elements or indicators that determine the quality of spatial data. Described quality indicators are conceptually defined in the quality model and presented in the quality report, which is generally a mandatory constituent part of spatial data transfer standards. The paper also gives a short comparative discussion and the definitions of spatial data quality elements as they are implemented in the american national standard SDTS, described in the special report of the ICA Commission on Spatial data quality and defined in the proposals for the new European spatial data standards developed under the auspices of CEN/TC 287.*

*Keywords: basic terms, definition of quality, elements, spatial data, standards*

## 1 UVOD

Kakovost prostorskih podatkov je ključen problem v obdobju skokovitega razvoja elektronske tehnologije in komunikacijskih omrežij, ki zagotavljajo hiter in enostaven dostop do podatkov. Veliko zbirk prostorskih podatkov lahko uporabniki dosežejo z vključitvijo v lokalno ali svetovno medomrežje (Internet). Brez poglobitvenih elementov<sup>1</sup>, ki opredeljujejo kakovost podatkov, se zdijo vsi pridobljeni podatki natančni in popolni. Med obdelavo pa bo uporabnik lahko opazil, da podatki iz različnih virov, ki jih primerja, niso enakovredni in se ne ujemajo v celoti. Pri grafični predstavitvi in povezovanju takšnih raznovrstnih podatkov se točke, linije in območja

ne prekrivajo natančno, čeprav predstavljajo iste prostorske pojave. Brez ustreznih pokazateljev kvalitete prostorskih podatkov, pridobljenih iz različnih virov, uporabnik nima nobenega oprijemljivega merila, da bi lahko razrešil nastala odstopanja.

V obdobju analogne kartografije je bila glavna pozornost za kvaliteto prostorskih podatkov usmerjena v natančnost planimetričnega kartiranja objektov (položajna pogreška  $m_y$  in  $m_x$  v ravnini upodobitve) v odnosu do njihove dejanske lege v prostoru. Na računalniških prikazih so videti digitalni prostorski podatki dovršeni. Zelo hitro se zato lahko ustvari varljiv občutek o njihovi navidezni natančnosti vkljub vsem vsebovanim pogreškom v samih podatkih. Izkušen uporabnik mora zato imeti na voljo ustrezne metode za opredelitev poznane kakovosti izbranih podatkovnih nizov pri njihovi analizi ali prikazovanju. Čeprav obstaja soglasje, da kvaliteta podaja neko vrsto zaželenih vrednosti, pa je dejanski pomen takšne vrednosti relativen oziroma intuitiven. ISO-va (International Standards Organisation) opredelitev pojma kvalitete v okvirih kontrole kakovosti je naslednja (ISO WWW):

- kvaliteta je skupek značilnosti in lastnosti proizvoda ali storitve, ki podpirajo njeno sposobnost, da zadovolji izražene ali vsebovane potrebe.

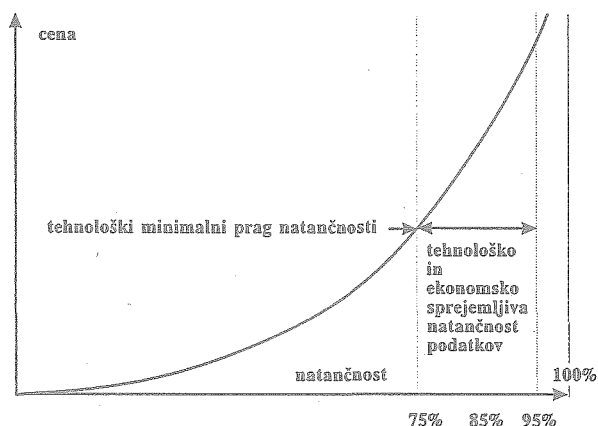
Taka opredelitev kvalitete poudarja njen relativen značaj. Kakovost se vedno meri v odnosu do potreb, ki so lahko izražene izrecno ali pa posredno. Zaradi take definicije se lahko kvaliteta določenega objekta spreminja, če obstajajo različne potrebe. Da bi zagotovili podatke o kvaliteti, ki so neodvisni od načina uporabe, je treba opredeliti kvalitetne norme. V primeru prostorskih podatkov temeljijo takšne kakovostne norme na osnovnih elementih, ki opredeljujejo kvaliteto podatkov in morajo biti sestavni del poročila o kvaliteti podatkov. Če se delno priredi že navedeno opredelitev ISA, potem je lahko kvaliteta prostorskih podatkov definirana (Guptill et al., 1995):

- kvaliteta prostorskih podatkov je skupek značilnosti in lastnosti podatkovne zbirke, ki podpirajo njeno sposobnost, da zadovolji izraženi niz zahtev.

Kakovost prostorskih podatkov opredeljujejo standardni pokazatelji, ki so povezani in se ne nanašajo samo na položajno natančnost ali točnost. Opredelitev ključnih elementov, ki so pomembni za določitev kakovosti podatkov, je samo prvi korak. Naslednja pomembna stopnja je opredelitev metodologije za specifikacijo kvalitete prostorskih podatkov glede na vse pomembne dejavnike, ki jo določajo. Pri izboru in izdelavi metodologije za določitev kakovosti so torej pomembne predvsem naslednje tri faze:

- razvoj in opredelitev ustreznega standardnega niza kriterijev oziroma elementov za opredelitev kvalitete prostorskih podatkov
- razvoj in opredelitev enostavnih metod za testiranje ter opredelitev kakovosti prostorskih podatkov in povezav med njimi
- opredelitev primernih metod za predstavitev in upodobitev kvalitete prostorskih podatkov (standardno poročilo) ter ustrezna grafična ponazoritev elementov kakovosti.

Tržna oziroma ekonomska vrednost prostorskih podatkov je prav tako tesno povezana z njihovo kakovostjo. Kvalitetnejši podatki so dražji zaradi številnih dejavnikov in postopkov, ki so potrebni za pridobivanje, zajemanje, urejanje, obdelavo, noveliranje (ažuriranje), hranjenje, predstavitev ter posredovanje prostorskih podatkov.



Slika 1

Stoodstotna natančnost<sup>2</sup> ali točnost podatkov je zgolj hipotetična predpostavka tako s tehničnega in z metodološkega kot tudi z ekonomskega vidika. Cena takšnih podatkov bi bila za proizvajalca in zlasti za uporabnika ekonomsko nesprejemljiva. Glede na praktične izkušnje oziroma hevristične predpostavke se kot minimalni tehnološki prag pričakuje nad 75-odstotna natančnost. Tehnološko izvedljiva in hkrati za nekatere uporabnike ekonomsko še sprejemljiva se ocenjuje do 95-odstotna natančnost podatkov. Na sliki 1 so prikazani opisani odnosi.

## 2 DOLOČANJE IN PODOJANJE KVALITETE PROSTORSKIH PODATKOV

**P**red izdelavo poročila o kakovosti je treba določiti oziroma izbrati ustrezno metodologijo za testiranje in formalno določitev kvalitete prostorskih podatkov v podatkovnih zbirkah. Kvaliteta podatkov se lahko preizkusi glede na vse elemente, ki jo določajo. Obstajajo uveljavljene metode za določitev kakovosti podatkov, ki temeljijo na ustreznem testiranju in statistični obdelavi izbranega podatkovnega vzorca. Takšni testi se običajno izvajajo ločeno za vsak posamični pokazatelj kvalitete podatkov. Prav tako je treba določiti tudi ustrezne načine za numerično in grafično predstavitev kvalitete prostorskih podatkov. Poročilo o značilnostih in kakovosti prostorskih podatkov mora vsebovati vsaj naslednje pomembne postavke:

- podatki o primernosti za uporabo<sup>3</sup> morajo biti na voljo neodvisno od podatkov, tako da lahko možni uporabniki ocenijo kakovost podatkovne zbirke pred njihovo uporabo,
- zanesljivost kakovostnih postavk predstavlja odgovornost za proizvajalca in uporabnika podatkov. Proizvajalec mora sporočiti vse standardne elemente o poznani kvaliteti podatkov. Uporabnik se mora zanesljivo odločiti o primernosti podatkov za nameravano uporabo,

- prožnost omogoča širjenje in dopolnjevanje elementov kakovosti podatkov. Najboljša je uporaba ustreznega metapodatkovnega standarda<sup>4</sup> o kvaliteti prostorskih podatkov.

Glede na raven, na katero se nanašajo razni elementi kvalitete prostorskih podatkov, se kakovost veže na posamezne podatkovne elemente, ki so lahko: podatkovna zbirka ali karta, podatkovna tematika, podatkovni niz ali sestavljeni objekt, objekt, entiteta, pojav ali podatkovni zapis, podatkovni zapis ali polje. Standardizirane metode in načini za podajanje poročila o kvaliteti prostorskih podatkov so prav tako zelo pomembni ter so lahko zlasti naslednji: poročilo in tekstualni opisi, grafični prikazi kakovosti, statistična opredelitev in obdelave izbranih testnih vzorcev.

### 3 OPREDELITEV ELEMENTOV KVALITETE PROSTORSKIH PODATKOV

**P**rostorski podatki so običajno organizirani v podatkovne nize, ki so razpoznavne zbirke podatkov. Prostorski podatki se lahko pridobijo iz neposrednih opazovanj, izpeljejo se lahko iz drugih podatkovnih virov, lahko so inačica ali posredna izpeljava iz opazovanj, ali pa so pridobljeni iz različnih interpretacij podatkov (CEN/TC 287, 1995c). Uporabnik kljub sodobni tehnološki podpori ni sposoben zajeti in obdelati vseh možnih podatkov o opazovanih prostorskih pojavih. Rezultat je posploševanje oziroma izbor in abstrakcija samo tistih podatkov, ki so pomembni za določeno vrsto uporabe ali aplikacije. V postopkih zajemanja in urejanja prostorskih podatkov je navzoča pristranskost tistega, ki podatke zajema. Pomembni so tudi njegov osebni izbor, obravnava in interpretacija.

**Z**a jasno razpoznavanje posameznih podatkovnih nizov mora biti na voljo dovolj metapodatkov. Obvezni del metapodatkov je tudi poročilo o kakovosti podatkov, ki mora vedno spremljati dejanske podatke pri standardnem prenosu. Kvaliteto prostorskih podatkov v poročilu ICA opredeljuje naslednjih sedem pomembnih elementov:

- izvor podatkov (vir in tehnologija)
- položajna natančnost (georeferenčna ali lokacijska) grafičnih podatkov
- opisna natančnost (zanesljivost opisnih podatkov)
- popolnost podatkov (podatkovni model in opisi)
- logična usklajenost podatkov (doslednost različnih povezav med geometričnimi ter opisnimi podatki)
- semantična natančnost podatkov (odnos med podatkovnim modelom ter stvarnostjo)
- ažurnost podatkov (časovne značilnosti in sistem noveliranja).

Določitev ali ocena vrednosti posameznih pokazateljev kvalitete se mora izvesti v skladu z izbrano metodologijo za testiranje kakovosti. Poročilo o kakovosti podatkov mora biti na voljo ne samo skupaj z dejanskimi prostorskimi podatki, temveč tudi neodvisno od samih podatkov. Tako lahko uporabniki predhodno ocenijo primernost podatkov pred njihovo uporabo oziroma nakupom.

### 4 OBRAVNAVA ELEMENTOV KVALITETE PROSTORSKIH PODATKOV

**V** naslednji preglednici so prikazani obravnavi pokazatelji kvalitete prostorskih podatkov v ameriškem nacionalnem standardu za prostorske podatke SDTS (Spatial Data Transfer Standard), v že omenjenem poročilu posebne komisije ICE in

predlogu standarda CEN tehničnega odbora TC 287 za Geografske informacije (CEN/TC 287, 1995b). Prikazan vrstni red izhaja iz izvornega vrstnega reda predloga ICE, ki je strokovno najbolj celovito opredeljen in predpostavlja obstoj sedmih ključnih elementov kakovosti podatkov.

<i>SDTS (1992)</i>	<i>ICA (poročilo, 1995)</i>	<i>CEN/TC 287 (1996)</i>
<i>izvor podatkov</i>	<i>izvor podatkov</i>	<i>izvor podatkov in uporaba</i>
<i>opisna natančnost</i>	<i>opisna natančnost</i>	<i>tematska natančnost</i>
<i>natančnost opisnih podatkov</i>	<i>natančnost opisnih podatkov</i>	<i>natančnost opisnih podatkov</i>
<i>popolnost podatkov</i>	<i>popolnost podatkov</i>	<i>popolnost podatkov</i>
<i>logična usklajenost podatkov</i>	<i>logična usklajenost podatkov</i>	<i>logična usklajenost podatkov</i>
	<i>semantična natančnost podatkov</i>	
	<i>ažurnost podatkov</i>	<i>časovna natančnost</i>

## 5 CEN/TC 287 MODEL KAKOVOSTI PROSTORSKIH PODATKOV

Predlog CEN/TC 287 temelji na tradicionalni triravenski arhitekturi kot podlagi za opisovanje različnih ravni podatkovne abstrakcije. Za opredelitev potrebnih shem se na konceptualni ravni uporablja formalni opisni jezik Express (ISO 10303-11). Shema vsebuje opis pomena, vsebine, sestave, odnosov in pravil v podatkovnem modelu. Predlog TC 287 za kakovostni model prostorskih podatkov je prikazan na sliki 2. Kvalitetni model je formalno opredeljen v kvalitetni podshemi. Ta je podlaga za oceno in določitev pokazateljev kvalitete, ki se predstavijo standardizirano v izkazu o kakovosti prostorskih podatkov.

### 5.1 Opredelitve osnovnih pojmov in razlaga

Poreklo<sup>5</sup> je edini obvezni pokazatelj kvalitete prostorskih podatkov, ki opisujejo zgodovino geografskega podatkovnega niza ali podniza v smislu odgovornega proizvajalca, tehničnega opisa izvornih materialov in izvedenih obdelav.

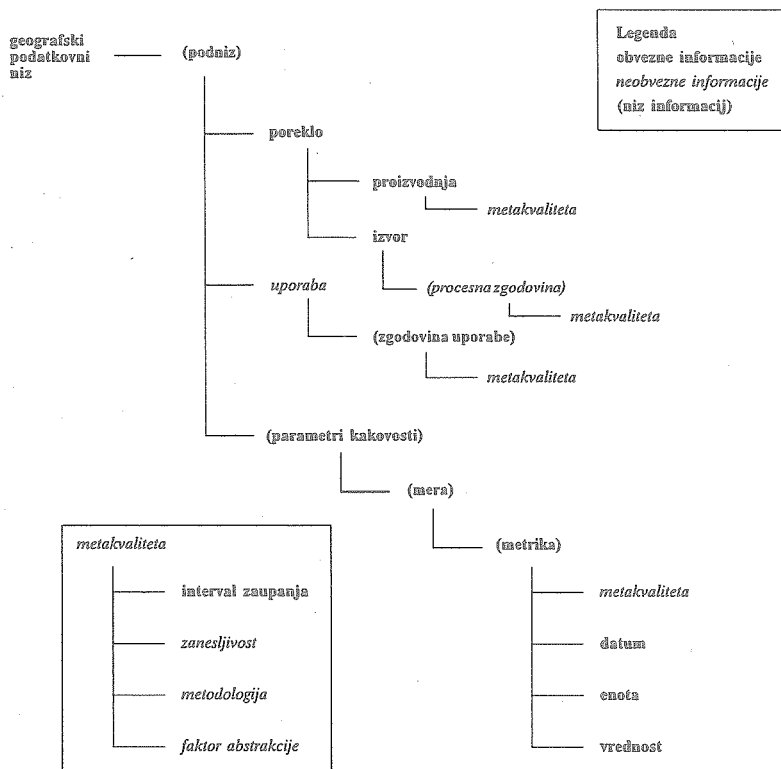
Uporaba<sup>6</sup> se lahko uporabi za opis predhodne uporabe prostorskega podatkovnega niza ali podniza.

Kakovostni parametri<sup>7</sup> so kvantitativni pokazatelji kvalitete geografskega podatkovnega niza ali podniza. Predlog kakovostnega modela CEN/TC 287 vsebuje pet primarnih in en sekundarni parameter kvalitete, katerih opredelitve so naslednje:

- pozicijska natančnost<sup>8</sup> je primarni kakovostni parameter, ki prikazuje točnost geografskega položaja v geografskem podatkovnem nizu ali podnizu
- tematska natančnost<sup>9</sup> je primarni kakovostni parameter, ki prikazuje točnost opisov v geografskem podatkovnem nizu ali podnizu
- časovna natančnost<sup>10</sup> je primarni kakovostni parameter, ki prikazuje točnost časovnih zapisov v geografskem podatkovnem nizu ali podnizu
- logična usklajenost<sup>11</sup> je primarni kakovostni parameter, ki prikazuje stopnjo strukturne celovitosti določenega podatkovnega niza

- popolnost<sup>12</sup> je primarni kakovostni parameter, ki prikazuje razliko med dejanskim podatkovnim nizom in njegovo opredelitvijo
- zanesljivost besedila<sup>13</sup> je sekundarni kakovostni parameter, ki prikazuje pravilnost besedila v geografskem podatkovnem nizu ali podnizu.

Metrika<sup>14</sup> je rezultat opisanega in datiranega preizkusa, ki ga tvori vrednost ali niz vrednosti opredeljenega podatkovnega tipa.



Slika 2

Mera<sup>15</sup> je niz ene ali več metrik, ki skupaj prikazujejo kvalitetno predstavo o določenem vidiku geografskega podatkovnega niza ali podniza.

Metakvaliteta<sup>16</sup> je vrednostna informacija o pokazatelju kakovosti. Vsak pojav informacije o kakovosti ima lahko podatek o lastni kvaliteti. Mero o metakakovosti sestavljajo:

- obvezna mera intervala zaupanja informacij o kakovosti
- neobvezna mera zanesljivosti informacij o kakovosti
- neobvezna mera o metodologiji, uporabljeni za pridobivanje informacij o kakovosti
- neobvezna mera o faktorju abstrakcije podaja razliko med nominalno osnovo in stvarnostjo.

Interval zaupanja<sup>17</sup> je ocena pravilnosti kakovostnega pokazatelja. Statistične indikatorje se predstavi kot interval z določeno verjetnostjo. Nestatistične pokazatelje se lahko prikaže s pomočjo ustreznih opisnih pojasnil.

Zanesljivost<sup>18</sup> podaja oceno predstaviteljnosti določenega kakovostnega pokazatelja.

Metodologija<sup>19</sup> je sestavina metakvalitete, ki prikazuje, kako je bil določen kakovostni pokazatelj določen.

Nominalna osnova<sup>20</sup> je idealizirana abstrakcija stvarnosti (pojmovni model), ki tvori podlago za opredelitev odnosa, na kateri se stvarnost odraža v geografskem podatkovnem nizu ali podnizu.

Faktor abstrakcije<sup>21</sup> podaja odmik pokazatelja kakovosti glede na zahteve postopka abstrakcije stvarnosti v nominalno osnovo (konceptualni model).

#### Literatura:

*CEN/TC 287/WG2 Geographic Information - Data Description - Metadata (N 370), September 1995a*

*CEN/TC 287/WG2 Geographic Information - Data Description - Quality (N 369), August 1995b*

*CEN/TC 287/WG1 Geographic Information - Reference Model (prEN 12009), July 1995c*

*CEN/TC 287 home page (URL): <http://ilm425.nlh.nolgis/cen.tc287>*

*FGDC meta standard home page (URL): <http://fgdc.er.usgs.gov/metaover2.html>*

*Fegas, G.R. et al., An Overview of FIPS 173. The Spatial Data Transfer Standard. ACSM, 1992, Vol. 19, No. 5*

*GIS Standards (OII) home page (URL): <http://www2.echo.lu/impact/oii/gis.html>*

*Guptill, C. S., Morrison L.J. (editors), Elements of Spatial Data Quality. ICA Commission on Spatial Data Quality. Pergamon Press, 1995*

*ISO/TC 211 home page (URL): <http://www.statkart.no/isotc211/>*

*Skogan D., Basic Principles of Geographic Information Sharing and Standardisation. ScanGIS'95 Proceedings, Trondheim, 1995*

- 1 Sedem elementov: izvor podatkov, položajna natančnost, natančnost opisnih podatkov, popolnost podatkov, logična usklajenost podatkov, semantična natančnost podatkov in ažurnost podatkov.
- 2 Natančnost ali točnost izraža, kako zanesljivo lahko določene meritve predstavljajo merjeno količino. Natančnost se običajno določa v smislu intervala, v katerem se predpostavlja, da leži prava vrednost merjene količine. V praktičnem pomenu se natančnost opredeljuje kot razlika med trenutno vrednostjo izbranega podatka in neko primerljivo bolj natančno vrednostjo istega podatka.
- 3 Primernost za uporabo je skupek značilnosti in lastnosti podatkovne zbirke, ki podpirajo njeno sposobnost, da zadovolji izražene zahteve določene aplikacije.
- 4 Metapodatki so podatki, ki opisujejo prostorske podatke in njihove poslovne vidike. Metastandard za prostorske podatke opredeljuje vsebinska in formalna pravila za opis prostorskih podatkov.
- 5 Lineage is the history of a geographic dataset or subset in terms of its producer (responsible organisation), technical description of source material and processing applied.
- 6 Usage may be utilised to describe the previous applications of spatial dataset or subset.
- 7 Quality parameters are quantitative indicators of quality of a geographic dataset or subset.
- 8 Positional accuracy is a primary parameter indicating accuracy of geographic position within a geographic dataset or subset.

- 9 Thematic accuracy a primary quality parameter indicating the accuracy of thematic attributes of a geographic dataset or subset.
- 10 Temporal accuracy is a primary quality parameter indicating the accuracy of time recorded within a geographic dataset or subset.
- 11 Logical consistency is a primary quality parameter indicating the degree of structural integrity of a given dataset.
- 12 Completeness is a primary quality parameter indicating the difference between an actual dataset and its specifications.
- 13 Textual fidelity is a secondary quality parameter indicating the correctness of text within a geographic dataset or subset.
- 14 Metrics results from a described and dated test consisting of a value or set of values of a defined type.
- 15 Measure is a set of one or more metrics that together indicate the quality performance of a particular aspect of a geographic dataset or subset.
- 16 Metaquality is a qualitative information about a quality indicator.
- 17 Confidence interval is an estimate of the correctness of a quality indicator. For statistical indicators this will be represented as an interval with a given probability. For non-statistical indicators it might be represented by explanatory text.
- 18 Reliability is an estimate of the representativity of a quality indicator.
- 19 Methodology is the component of metaquality that indicates how a quality indicator has been achieved.
- 20 Nominal ground is an idealised abstraction of reality (conceptual model) forming the basis for specifications defining how a geographic dataset or subset relates to reality.
- 21 Abstraction modifier indicates the distortion of a quality indicator by the requirements of the process of abstraction from reality to the nominal ground.

*Recenzija: Matjaž Ivačič (v delu)  
mag. Božena Lipej*