

# KVALITETA PODATAKA I NJEN ZNAČAJ DANAS

Datum prijave: 20.04.2015.  
Datum prihvaćanja: 29.05.2015.

UDK: 004.6+007  
Pregledni rad

mr. sc. Damir Vuk,  
Enes Ciriković, dipl. ing.  
Dominik Suk, student  
Visoka škola Virovitica  
Matije Gupca 78, 33000 Virovitica  
Telefon: 033 721 099 E-mail: damir.vuk@vsmti.hr

**SAŽETAK** - Iako nije sasvim nova, paradigma „kvaliteta podataka“ (data quality), danas još uvijek ne gubi na značaju. Unapređenje kvalitete podataka može predstavljati značajan izvor konkurentne prednosti. S druge strane, posljedice loše kvalitete podataka mogu biti kritične za uspjeh pojedinih organizacija. U radu se analizira pojam kvalitete podataka kao multidimenzionalna osobina. U fokusu je pojam kvalitete podataka sa stajališta korisnika odnosno konzumenta kvalitete podataka. Razmatra se koncept „informativnog proizvoda“ i informativnog proizvodnog sustava. Na kraju se daje sažeti pregled vrsta troškova kao posljedica niske kvalitete podataka, a isto tako i troškova koji nastaju kao rezultat investiranja u kvalitetu podataka.

**Ključne riječi:** kvaliteta podataka, informativni proizvod, troškovi kvalitete podataka

**SUMMARY**- Although not entirely new, the paradigm of "data quality", still today remains important. Improving data quality can be a significant source of competitive advantage. On the other hand, the consequences of poor data quality can be critical to the success of individual organizations. The paper analyzes the concept of data quality as multidimensional feature. It focuses on the concept of data quality from the standpoint of the user or consumer of data quality. We consider the concept of "information products" and information production system. Finally, a short review of possible costs types as a result of low-quality data, and costs incurred as a result of analyzing and investing in data quality are discussed.

**Keywords:** data quality, information product, data quality cost

## 1. UVOD

U današnjoj, informativskoj ekonomiji, gdje se poduzeća i organizacije susreću sa sveprisutnom globalnom konkurencijom, podaci i informacije predstavljaju ključni faktor njihove uspješnosti i dugoročne održivosti. Nije stoga neobično da pojam kvalitete podataka već cijeli niz godina zaokuplja pažnju kako informatičkih profesionalaca, tako i poslovnih menadžera. Očigledno je da loša kvaliteta podataka može imati značajne posljedice na poslovanje organizacije odnosno poduzeća. Opće je prihvaćen stav da se utjecaj loše kvalitete podataka odražava kako na kvalitetu procesa donošenja odluka, odnosno na kvalitetu donesenih odluka, tako i na organizacijsko povjerenje u raspoložive podatke te na zadovoljstvo kupaca i poslovnih partnera u cjelini.

Iako se kvaliteta podataka više od dvadeset godina sustavno proučava na akademskoj razini, a metode procjene i upravljanja kvalitetom podataka prakticiraju u nizu naprednih organizacija, danas je problem kvalitete podataka aktualniji više nego ikada. Dva su razloga za to.

Prvi razlog proizlazi iz iznimne eksplozije gomilanja velikih količina podataka obuhvaćenih u digitalnoj formi na memorijama s direktnim pristupom. Ta eksplozija se događa osobito proteklih nekoliko godina i to s trendom još bržeg rasta, uzrokovanog dostupnošću brzih memorija s direktnim pristupom ogromnih kapaciteta i razumnih cijena, ali također uzrokovana najrazličitijim izvorima podataka od kojih su najznačajniji online internet korisnici, te automatski izvori kao posljedica različitih IoT uređaja (Internet of

Things). Prema nekim procjenama, u 2013. godini je u svijetu već bilo prisutno preko dvjesto milijuna IoT uređaja (Jarr, 2015.).

Drugi razlog je dostupnost zrelih kako komercijalnih alata, tako i alata otvorenog koda, koji omogućuju efikasno analiziranje velikih količina podataka (Big Data), bilo sa svrhom dobivanja fleksibilnih uvida u podatke, bilo sa svrhom ekstrakcije znanja iz tih podataka. Radi se o tzv. OLAP alatima i alatima za rudarenje podataka (Data Mining). Ti alati se oslanjaju na podatke organizirane u tzv. skladišta podataka (Data Warehouse). U najnovije vrijeme, kada te količine podataka postaju nehomogene i iznimno velike (Big Data), koristi se naziv jezero podataka (data lake) (Jarr, 2015.).

Sve to dovodi do paradigme o podacima kao trajnom resursu organizacije (Levitin i Redman, 1998). U klasičnom odlučivanju podaci imaju svoj životni vijek, nakon toga oni prestaju biti aktualni i postaju povijesni te se brišu ili eventualno arhiviraju. Novost je u tome što danas podaci kako aktualni, ali isto tako i povijesni, postaju trajni resurs organizacije koji se može i mora opet iznova koristiti kako bi se dobili novi uvidi odnosno znanje koje proizlazi iz velikih količina povijesnih podataka. Dakle, danas podaci u principu ne zastarijevaju čak i kad postaju povijesni. To je moguće zahvaljujući mogućnosti pohranjivanja velikih količina podataka, kao i mogućnosti efikasnog manipuliranja tim podacima.

Prema tome, i povijesni podaci imaju potencijal da se iz njih dobiju aktualne informacije pa čak i nove spoznaje odnosno znanje. Ali, nisu svi podaci pogodni za to. Uvjet je da budu kvalitetni, tj., da posjeduju

odgovarajući stupanj kvalitete. Podaci loše kvalitete nisu dobar resurs. Pritom, ovdje treba pojam kvalitete podataka shvatiti relativno.

Netočni, nepravovremeni ili nekonzistentni podaci su uvijek loše kvalitete. Međutim, podaci mogu biti točni, potpuni, čitljivi, dostupni i pogodni za jednu (predviđenu) upotrebu, ali mogu istovremeno biti neupotrebljivi za neku drugu upotrebu. Pogodnost upotrebe (fitness for use) je važan kriterij kvalitete podataka. Idealno bi bilo kad bi podaci bili pogodni za bilo koju upotrebu, ali to nije uvijek moguće.

Podaci sami po sebi nisu dovoljni za upotrebu. Da bi smo od „sirovih“ podataka dobili informacije, potrebni su i dodatni podaci o podacima - tzv. metapodaci. Podaci su uvijek predstavljeni u formi nekog modela podataka. Taj model podataka može biti sasvim jednostavan ili kompleksan, ali u svakom slučaju on određuje kako će biti interpretirano značenje podataka (tj. semantika podataka) koji su njime predstavljeni. Metapodaci su podaci koji semantički opisuju podatke u nekoj reprezentaciji. Jedino podaci uz odgovarajuću semantiku mogu biti interpretirani, odnosno iz takvih podataka je moguće ekstrahirati informacije ili čak novo znanje odnosno spoznaje. Podaci s neodgovarajućom ili nedostatnom semantikom ne mogu se dobro interpretirati - njihova kvaliteta je loša s obzirom na dimenziju „pogodnost upotrebe“. Unatoč tome, i takvi podaci mogu biti upotrebljivi, ali njihova upotreba je ograničena. Dakle, semantička izražajnost odnosno potpunost modela podataka (reprezentacije podataka) direktno utječe na kvalitetu podataka.

## 2. DIMENZIJE KVALITETE PODATAKA

Danas među istraživačima kvalitete podataka prevladava stav da se kvaliteta podataka mora definirati kroz više dimenzija. Postoje međutim, izvjesne razlike u tome koje dimenzije su prisutne, te kako ih kategorizirati. U literaturi često citirani, Wang i Strong, razlikuju tri pristupa problemu kvalitete podataka: intuitivni, teorijski i empirijski.

Intuitivni pristup uzima u obzir jedan ili svega nekoliko svojstava kvalitete podataka koji proizlaze uglavnom iz prethodnog iskustva ili intuicije istraživača. Najčešće se uzimaju atributi kvalitete u užem smislu kao što je točnost podataka, a u nekim slučajevima i pouzdanost. Prednost ovoga pristupa je u tome što u praktičnom smislu omogućuje usredotočenje na one attribute kvalitete podataka koji su u konkretnom slučaju najznačajniji.

U teorijskom pristupu problemu kvalitete podataka fokus je na istraživanju načina na koji podaci tijekom procesa proizvodnje podataka postaju loši. Tako Lukyanenko i Parsons svoj „teorijski pristup kvaliteti podataka“ zasnivaju na konceptu klasa (Lukyanenko i Parsons, 2011.).

Empirijski pristup istražuje karakteristike kvalitete podataka sa stajališta korisnika, odnosno u kojoj mjeri podaci zadovoljavaju njihove potrebe. Prema tome, empirijski je pristup usmjeren na po-

datke kao proizvode koji nastaju u procesu „informativne proizvodnje“.

Strong, Lee i Wang koji zastupaju empirijski pristup, smatraju da se problem kvalitete podataka mora promatrati u širem kontekstu gdje se podaci nalaze odnosno upotrebljavaju, a to je kontekst informacijskog sustava. Sukladno tome, autori definiraju konceptualni okvir (framework) za koji iznose da je provjeren kako u industriji tako i u državnoj upravi. Okvir sadrži petnaest dimenzija, grupiranih u četiri kategorije kvalitete podataka:

- unutrašnja kvaliteta (intrinsic),
- kvaliteta dostupnosti (accessibility),
- kontekstualna kvaliteta (contextual) i
- reprezentacijska kvaliteta podataka (representational).

Unutrašnja kvaliteta podataka koja se često u pojednostavljenoj formi shvaća kao točnost i objektivnost, predstavlja kvalitetu podataka u užem smislu. Ona obuhvaća četiri dimenzije:

- točnost,
- objektivnost,
- vjerodostojnost i
- povjerljivost izvora.

Kategorija dostupnosti podataka može biti odlučujuća za upotrebu. Budući da današnji online sustavi omogućuju globalnu dostupnost u svako vrijeme i na svakom mjestu, korisnici to i očekuju, ali pritom očekuju i sigurnost pristupa podacima:

- Dostupnost podataka
- Sigurnost pristupa

Kategorija kontekstualnih dimenzija definira dimenzije kvalitete podataka sa stajališta konteksta aktivnosti, odnosno zadataka u kojima konzumenti koriste podatke. Ova kategorija uključuje pet pojedinačnih dimenzija:

- Relevantnost,
- Dodanu vrijednost,
- Pravovremenost,
- Kompletnost i
- Količinu podataka.

Korisnici podataka moraju biti u stanju shvatiti značenje podataka i to s lakoćom, bez mnogo napora. Pred toga potrebno je da podaci budu koncizno prikazani, te da budu konzistentni. Reprezentacijske dimenzije su:

- Interpretabilnost,
- Lakoća razumijevanja,
- Konciznost reprezentacije i
- Konzistentnost reprezentacije.

Ovaj konceptualni okvir kvalitete podataka ne treba shvatiti sasvim strogo, tim više što je on rezultat empirijskog istraživanja pa kao takav može varirati u zavisnosti od slučaja do slučaja, stoga on predstavlja samo dobru ideju vodilju u pristupu analizi kvalitete podataka.

Uzmemo li u obzir upravo prikazanu multidimenzionalnost kvalitete podataka, vidimo da je u praktičnom slučaju teško postići potpunu kvalitetu podataka tj. po svim dimenzijama kvalitete. U stvarnosti, to nije niti nužno, a niti poželjno. Kvaliteta podataka

je rezultat odgovarajućih napora u organizaciji, a to razumljivo ima svoju cijenu. Ako trošak postizanja odgovarajućeg stupnja kvalitete podataka ne opravdava korist, onda je takav trošak nepoželjan. Sintagma „prikladnost za upotrebu“ („Fitness for use“), označava upravo to - onu razinu i vrstu kvalitete podataka koja je prikladna odnosno potrebna za odgovarajuću upotrebu, onako kako ju definiraju konzumenti podataka.

### 3. PODACI KAO INFORMACIJSKI PROIZVODI

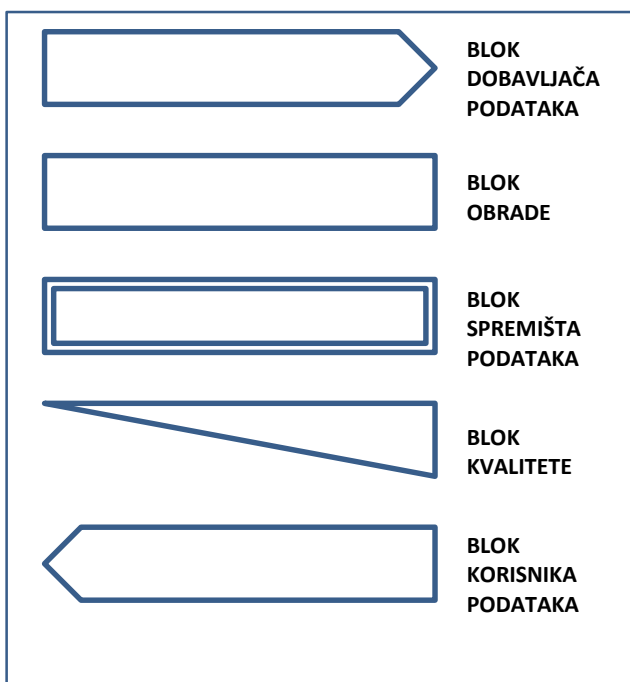
Kao što smo vidjeli, empirijski model kvalitete podataka je zasnovan na koceptu korisnika-konzumenta podataka, te pogledu na dimenzije kvalitete sa stajališta potreba korisnika.

Na istoj ideji su Ballou et al., postavili teorijski model tzv. „proizvodni pristup“, koji obuhvaća slijedeće temeljne koncepte:

- informacijski proizvodni sustav,
- informacijski proizvod i
- korisnik kao konzument kvalitete podataka.

Analogno klasičnom industrijskom proizvodnom sustavu, oni definiraju informacijski proizvodni sustav (information manufacturing system), kojemu je output informacijski proizvod (information product), čijom kvalitetom treba upravljati. „Informacijski proizvodni sustav se odnosi na informacijski sustav koji proizvodi predefinirane informacijske proizvode“ (Ballou et al., 1998.).

SLIKA 1: GRAFIČKI SIMBOLI INFORMACIJSKOG PROIZVODNOG MODELA



Izvor: Prilagođeno prema (Ballou et al. 1998.)

Proizvodni pristup se zasniva na četiri atributa informacijskih proizvoda:

- pravovremenost (timeliness),
- kvaliteta podataka (data quality),

- trošak (cost) i
- vrijednost (value).

Informacijski proizvodni sustav se promatra odnosno analizira, kroz različite elemente nazvane blokovi. Za tu svrhu razvijena je dijagramska notacija koja omogućuje zorniji prikaz informacijskog proizvodnog sustava (Baillou et al. 1998.). Na slici 1. su prikazani elemente te grafičke notacije.

U tom modelu je definirano pet blokova :

- blok dobavljača podataka (data vendor block),
- blok obrade (processing block),
- blok spremišta podataka (data storage block),
- blok kvalitete (quality block) i
- blok korisnika (customer block).

Blok dobavljača je namijenjen za prikaz različitih izvora „sirovih“ podataka. Ovaj blok nema ulaz iz nijednog drugog bloka. On prikazuje ulaz jednog ili više, bilo eksternih, bilo internih podataka.

Blok obrade predstavlja blok u kome se podaci obrađuju odnosno transformiraju, te im se dodaje vrijednost za korisnika.

Blok spremišta podataka omogućuje pohranjivanje podataka u datoteke ili bazu podataka.

Blok kvalitete obrađuje podatke tako da im povećava njihovu kvalitetu na izlazu iz bloka u odnosu na kvalitetu koju su imali na ulazu u blok.

Autori ove notacije ističu da je moguće povući paralelu s notacijom dijagrama toka podataka (DFD-data flow diagrams). Blokovi dobavljača i korisnika odgovaraju eksternim entitetima u DFD. Blok obrade je analogan konceptu funkcije, a blok spremišta je analogan spremištu podataka. Jedino blok kvalitete nema adekvatan koncept u DFD. Unatoč toga, autori ne isključuju mogućnost upotrebe DFD za potrebe modeliranja informacijskog proizvodnog sustava (Ballou et al., 1998.).

Pored grafičkog instrumentarija za modeliranje i analizu, informacijski proizvodni model uključuje numeričku analizu kojoj je cilj ne samo povećati kvalitetu podataka, nego i upravljati optimizacijom vremenske i troškovne komponente modela.

Ideja informacijskog proizvodnog modela koji proizvodi informacijske proizvode, usmjerena je ne samo na analizu takvog sustava, nego prije svega na trajno upravljanje s tim sustavom te tretiranjem informacije kao proizvoda, umjesto informacije kao nuz-proizvoda. S tim u vezi, Wang, Lee, Pipino i Strong naglašavaju da ako se želi tretirati informacija kao proizvod, treba poštivati sljedeća četiri principa (Wang et al., 1998.b):

1. Razumjeti informacijske potrebe konzumenata informacija.
2. Upravljati informacijom kao proizvodom dobro definiranog proizvodnog procesa.
3. Upravljati informacijom kao proizvodom koji ima svoj životni ciklus.
4. Zadužiti osobu kao menadžera informacijskih proizvoda da upravlja informacijskim procesima kao i rezultirajućim informacijskim proizvodima.

Razumijevanje informacijskih potreba odnosno zahtjeva korisnika, je u neposrednoj vezi s kvalitetom podataka odnosno informacija koje korisnici dobivaju. Proizvodnja informacija mora biti u dobro definiranom sustavu, u suprotnom informacijski proizvodi će biti samo nusproizvodi upitne kvalitete. Svaki informacijski proizvod ima svoj životni ciklus koji mora biti poznat i upravljan. Zadaća menadžera za informacijske proizvode je da upravlja promjenjivim informacijskim potrebama korisnika te da na bazi objektivnih metrika doprinosi unapređenju informacijskih proizvoda.

#### 4. ZNAČAJ SKLADIŠTENJA PODATAKA

Značaj upravljanja kvalitetom podataka u današnje vrijeme raste iz dva razloga. Prvo, količine podataka koje organizacije posjeduju su sve veće, a trend povećanja količine spremljenih podataka kao i onih koje treba obraditi, stalno raste. Tome doprinosi spoznaja da povijesni podaci mogu biti izvor bitnih strateških, ali taktičkih odluka ako se nad njima primjene suvremene metode OLAP (OnLine Analytical Processing) analize, a posebno metode rudarenja podataka.

Drugo, tradicionalni izvor strukturiranih podataka u tipičnom poduzeću, je bila SQL baza podataka. Danas sve više pored strukturiranih, pojavljuju i razni drugi polu strukturirani ili nestrukturirani izvori upitne ili barem nedorečene semantike. Nadalje, sve veći je broj različitih izvora podataka. Mnogi od tih izvora su lokalni, često pod nedovoljnom kontrolom.

Problem kvalitete podataka i analitičke obrade podataka (OLAP) iz različitih, često nehomogenih izvora podataka, trebao je riješiti koncept skladišta podataka kao konsolidiranog jedinstvenog izvora podataka kontrolirane kvalitete. Unatoč toga, skladište podataka automatski ne rješava sve probleme. Prije svega, skladište podataka u svom konceptu nije usmjereno na poboljšanje kvalitete podataka po svim dimenzijama. Samo su neke od dimenzija u fokusu. U ETL procesu punjenja skladišta podacima, fokus je uglavnom na objedinjavnju različitih izvora, djelomičnom poboljšanju točnosti te stvaranju dostupnosti podataka. Problem je u tome što se semantika ulaznih podataka u pravilu ne može nadoknaditi ako prije toga nije postojala na izvornim podacima. Nadalje, problem null-vrijednosti je veći nego što se misli, budući da postojanje null-vrijednosti u podacima odražava u pravilu, pogrešno modeliranu shemu baze podataka, a to ukazuje na pogrešnu ili barem nedostajuću semantiku podataka.

Kontrola dozvoljenih vrijednosti podataka koje ulaze u bazu podataka putem aplikacijskih programa, bez obzira radi li se o desktop ili web aplikacijama, nije dovoljna. Nužno je kontrolirati podatke koji ulaze u bazu podataka putem integritetskih pravila unutar baze podataka kad god je to moguće, a samo u slučajevima kad DBMS to ne omogućuje, onda jedino što preostaje je ugradnja semantičkog odnosno integritetskog ograničenja na razini aplikacijskih programa. Prirodno mjesto za ugradnju semantike je

baza podataka, jer je na taj način moguće cjelovito i centralno upravljati semantikom podataka te izbjeći da u bazu podataka uđu nečisti podaci (dirty data). To ne samo da smanjuje potrebu pročišćavanja nečistih podataka u ETL procesu, nego i direktno doprinosi boljoj kvaliteti podataka. Praktičari često zanemaruju te činjenice, te daju prednost rješenjima koja daju bolje performanse na račun semantike. Također, često se zanemaruje činjenica da podaci u bazi podataka mogu biti nečisti, odnosno da im je kvaliteta barem sa stajališta točnosti i konzistentnosti nezadovoljavajuća.

Ako takvi podaci dođu bez popravka u skladište podataka, onda je njihova upotrebljivost sa stajališta konzumenata, upitna. Na tu činjenicu upozorava Kim u svom radu (Kim, 2002.).

Čak i ako zanemarimo navedene probleme pa hipotetski pretpostavimo da su podaci koji ulaze u skladište točni i konzistentni, problem neće biti isključen. Naime, postojeći podaci u svakoj bazi podataka pa tako i skladištu imaju svoj životni ciklus te s vremenom degeneriraju i postaju netočni odnosno nekonzistentni. Uzmimo za primjer podatke o kupcima u skladištu podataka. Neke procjene govore da mjesečno dva posto tih podataka degenerira zbog zastarijevanja (Eckerson, 2002.). U ovom slučaju riječ je o matičnim podacima (master data). Slična je situacija i s drugim podacima. Transakcijski podaci su uvijek referencijski povezani s matičnim podacima pa svaka degeneracija matičnih podataka izaziva povezanu degeneraciju transakcijskih podataka.

Prema tome, skladištenje podataka nije odgovor na sve probleme kvalitete podataka, kako se često pogrešno smatra, jer samo po sebi ne poboljšava sve dimenzije kvalitete podataka.

#### 5. IZVORI I POSLJEDICE LOŠE KVALITETE PODATAKA

Loša kvaliteta podataka može negativno utjecati na organizaciju odnosno poduzeće. Osobito ona loša kvaliteta koja nije identificirana i upravljana, može izazvati negativne posljedice (Wang i Strong, 1996.c), (Redman, 1998.), (Haug et al., 2011.).

Tipični problemi mogu se manifestirati kao: nezadovoljstvo kupaca, povećani ili nepotrebni troškovi poslovanja, nemogućnost efikasnog donošenja odluka kako na taktičkoj, tako i na strateškoj razini, niska organizacijska efikasnost, a često i nezadovoljstvo zaposlenika te općenito pogoršana organizacijska kultura. Po nekim rezultatima detaljno rađenih istraživanja, loša kvaliteta podataka može utjecati na povećanje operativnih troškova u iznosima od 8-12% prihoda, a kod uslužnog sektora čak 40-60% izdataka (Redman, 1998.). Ova istraživanja ukazuju da je trošak uzrokovan lošom kvalitetom evidentno prisutan, međutim egzaktni iznos troška ostaje samo na razini procjene (Redman, 1998.).

Kao što je navedeno, posljedice loše kvalitete podataka su evidentne i potvrđene kroz empirijska istraživanja. S druge strane, poboljšanje kvalitete podataka nije besplatno. Zbog toga kada govorimo o

troškovima kvalitete podataka, moramo uzeti u obzir dvije kategorije troškova: jedno su troškovi koji proizlaze kao posljedica niske kvalitete podataka; dok su drugo troškovi koji proizlaze iz nastojanja da se održi potrebna kvaliteta podataka ili da se loša kvaliteta podataka popravi odnosno poboljša. Eppler i Helfert, 2004., klasificiraju troškove kvalitete podataka na te dvije kategorije, a zatim svaku od njih na još dvije hijerarhijske razine.

Prema toj podjeli troškovi koji su posljedica niske kvalitete podataka dijele se na: a) direktne troškove (verifikacijske troškove, troškove ponovnog unosa, kompezacijske troškove), b) indirektno troškove (troškovi proizašli iz loše reputacije, troškovi proizašli iz loših odluka ili akcija, troškovi nepovratnih investicija).

Troškovi koji proizlaze iz unapređenja odnosno održavanja kvalitete podataka, oni dijele na tri podkategorije: c) prevencijski troškovi (troškovi treninga, nadzorni troškovi, troškovi razvoja i primjene standarda), d) troškovi detekcije (troškovi analize, troškovi izvještavanja), e) troškovi popravka (troškovi planiranja popravka, troškovi provedbe popravka) (Eppler i Helfert, 2004.).

Ako uzmemo u obzir svu kompleksnost problema održavanja odnosno postizanja potrebne razine kvalitete podataka kao multidimenzijske varijable, nužno se nameće zaključak da je potrebno odrediti optimalnu strategiju koja je definirana optimalnim miksom troškova zbog loše kvalitete i troškova zbog održavanja određene ciljane kvalitete podataka.

## 6. ZAKLJUČAK

O kvaliteti podataka se u literaturi i praksi raspravlja cijeli niz godina. Unatoč tome, kvaliteta podataka ne gubi na značaju, a niti na interesu istraživača i praktičara. U praksi ne bi trebalo problem kvalitete podataka svesti na jednostavne tehničke karakteristike kao što je točnost i ispravnost. Kvaliteta je više od toga - ona je višedimenzionalna. Dimenzije kvalitete proizlaze iz potreba korisnika - konzumenta podataka. Kvaliteti podataka odnosno kvaliteti informacija, treba pristupiti kroz koncept informacijskog proizvoda, jer tako dolazi u fokus njena multidimenzionalnost.

Važnost procjene, unapređenja i održavanja kvalitete podataka za poduzeća i organizacije općenito, može biti osobito bitna. Ali, poboljšanje kvalitete podataka, kao i njeno održavanje, predstavlja dvodimenzionalni trošak. U praksi je potrebno pronaći optimalni splet tih dviju dimenzija troškova kvalitete podataka.

## LITERATURA

1. Ballou, D., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modeling information manufacturing systems to determine information product quality. *Management Science*, 44(4), 462-484.
2. Eckerson, W. (2002). *Data Warehousing Special Report: Data quality and the bottom line*. The Data Warehousing Institute, Preuzeto 10.01.2015.: [http://www.es-tgv.ipv.pt/PaginasPessoais/jloure-iro/ESI\\_AID2007\\_2008/fichas/TP06\\_anexo1.pdf](http://www.es-tgv.ipv.pt/PaginasPessoais/jloure-iro/ESI_AID2007_2008/fichas/TP06_anexo1.pdf)
3. Eppler, M., & Helfert, M. (2004, November). A classification and analysis of data quality costs. In *International Conference on Information Quality*.
4. Haug, A., Zachariassen, F., & Van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, 4(2), 168-193.
5. Jarr, S. (2015.) *Fast Data and the New Enterprise Data Architecture*, Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472., ISBN: 978-1-491-91388-8
6. Kim, W. (2002). On Three Major Holes in Data Warehousing Today. *Journal of Object Technology*, vol.1(4), 39-47.
7. Levitin, A. V., & Redman, T. C. (1998). Data as a resource: Properties, implications, and prescriptions. *Sloan Management Review*, 40 (1), pp.89-101.
8. Lukyanenko, R., & Parsons, J. (2011). Data Quality as an Outcome of Conceptual Modeling Choices. In 16th the International Conference of Information Quality (ICIQ2011).
9. Madnick, S. E., Wang, R. Y., Lee, Y. W., & Zhu, H. (2009). Overview and framework for data and information quality research. *Journal of Data and Information Quality (JDIQ)*, 1(1), 2.
10. Parker, G. G., Van Alstyne, M. W. (2005). Two-Sided Network Effects: A Theory of Information Product Design. *Management Science*, Vol.51, No. 10., pp. 1494-1504.
11. Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211-218.
12. Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), 79-82.
13. Sadiq, S., Yeganeh, N. K., & Indulska, M. (2011, January). 20 years of data quality research: themes, trends and synergies. In *Proceedings of the Twenty-Second Australasian Database Conference-Volume 115* (pp. 153-162). Australian Computer Society, Inc..
14. Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103-110.
15. Wang, R. Y. (1998.a). A product perspective on total data quality management. *Communications of the ACM*, 41(2), 58-65.
16. Wang, R. Lee., Pipino, Y. W. and Strong, D. (1998.b) *Manage your information as a product*. *Sloan Management Review*, 4: 95-105.
17. Wang, R. Y., & Strong, D. M. (1996.c). Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 5-33.