

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**ZAGOTAVLJANJE KAKOVOSTI PODATKOV PRI IZVAJANJU
NOVEGA KAPITALSKEGA SPORAZUMA BASEL II**

Ljubljana, april 2011

NIKOLA MURK

IZJAVA

Študent Nikola Murk izjavljam, da sem avtor tega magistrskega dela, ki sem ga napisal v soglasju s svetovalcem prof. dr. Jurijem Jakličem, in da v skladu s 1. odstavkom 21. člena Zakona o avtorskih in sorodnih pravicah dovolim njegovo objavo na fakultetnih spletnih straneh.

V Ljubljani dne _____

Podpis: _____

KAZALO

UVOD	1
1 SPLOŠNO O BASELSKIH SMERNICAH	3
1.1 Zgodovina in prihodnost.....	3
1.2 Zahteve Basla II.....	4
1.2.1 Minimalne kapitalske zahteve – prvi steber	6
1.2.2 Regulativni nadzor – drugi steber.....	7
1.2.3 Tržna disciplina – tretji steber	8
1.3 Basel III	9
2 OBLIKOVANJE OKOLJA ZA KAKOVOST PODATKOV	9
2.1 Podatki.....	10
2.2 Informacije	10
2.3 Znanje	11
2.4 Modrost.....	11
2.5 Kakovost.....	12
2.6 Opredelitev kakovosti informacij	12
2.6.1 Naravna kakovost informacij	12
2.6.2 Pragmatična ali stvarna kakovost informacij.....	13
2.6.3 Opredelitev kakovosti informacije	14
2.6.3.1 Doslednost	14
2.6.3.2 Izpolnjevanje	15
2.6.3.3 Intelektualni delavec in končni uporabnik.....	15
2.6.3.4 Pričakovanja	16
2.6.4 Komponente kakovosti informacij	16
2.7 Opredelitev podatkov in kakovost arhitekture informacij.....	17
2.8 Kakovost predstavitve podatkov	18
3 PRIMERJAVA METODOLOGIJ.....	19
3.1 Primerjalna perspektiva za analiziranje metodologij.....	19
3.2 Dimenzije in mere	19
3.2.1 Točnost	27
3.2.2 Popolnost	27
3.2.3 Doslednost	28
3.2.4 Časovne dimenzije: Aktualnost, Volatilitnost ali Nihajnost in Pravočasnost	28
3.3 Strategije in tehnike	29
3.3.1 Podatkovno usmerjene tehnike.....	32
3.3.2 Procesno usmerjene tehnike	33
3.4 Faze in koraki	33
3.4.1 Faza ocenjevanja	35
3.4.2 Faza izboljšanja	36
3.5 Stroški.....	38
3.6 Vrste podatkov.....	41

3.7 Vrste informacijskih sistemov	43
4 OPREDELITEV KAKOVOSTI PODATKOV	46
4.1 Predstavitev poslovne banke X	46
4.2 Razlaga pojmov	46
4.3 Osnovni model Basel II sistema	47
4.4 Predstavitev okvira kakovosti podatkov	50
4.4.1 Temeljni procesi za zagotovitev kakovosti podatkov	50
4.4.1.1 Proces dostave	50
4.4.1.2 Proces spremljanja in eskalacije	51
4.4.1.2.1 Proces spremljanja	51
4.4.1.2.2 Proces eskalacije	52
4.4.1.3 Proces upravljanja s specifikacijo	53
4.4.1.3.1 Upravljanje z izdajami SPEC	53
4.4.1.3.2 Zaključevanje nove različice SPEC	53
4.4.2 Opredelitev delovnih mest	54
4.4.2.1 Pravila komuniciranja	55
4.4.2.2.2 Centralni poslovni strokovnjaki	57
4.4.2.2.3 Decentralizirani upravljalci kakovosti podatkov	58
4.4.2.2.4 Decentralizirani poslovni strokovnjaki	59
4.4.2.2.5 Lokalni upravljalci kakovosti podatkov	60
4.4.2.2.6 Lokalni poslovni strokovnjaki	61
4.4.2.2.7 Lokalni IT upravljalci kakovosti podatkov	62
4.5 Poročanje o statusu kakovosti podatkov	63
SKLEP	66
LITERATURA IN VIRI	69
PRILOGE	

KAZALO TABEL

Tabela 1: Metodologije kakovosti podatkov	20
Tabela 2: Metodologije in dimenzije kakovosti podatkov	20
Tabela 3: Dimenzije in mere	22
Tabela 4: Metodologije in mere kakovosti	25
Tabela 5: Opredelitev dimenzije popolnosti	27
Tabela 6: Opredelitev časovnih dimenzij	28
Tabela 7: Metodologije in vrste strategij	31
Tabela 8: Metodologije in koraki ocenjevanja	35
Tabela 9: Metodologije in koraki izboljšanja – korak 1	36
Tabela 10: Metodologije in koraki izboljšanja – korak 2	37
Tabela 11: Opredelitev stroškov nekakovostnih podatkov	39
Tabela 12: Različne predstavitve istega predmeta v realnem svetu	42
Tabela 13: Metodologije in vrste podatkov	43

Tabela 14: Metodologije in informacijski sistemi.....	44
Tabela 15: Poročilo o statusu kakovosti podatkov	65

KAZALO SLIK

Slika 1: Tristebni sistem določanja kapitalske ustreznosti.....	5
Slika 2: Osnovni model sistema	49
Slika 3: Hierarhična struktura odgovornih za kakovost podatkov	55
Slika 4: Komunikacijski tokovi	56
Slika 5: Poročanje o statusu kakovosti podatkov	63

UVOD

Problematika in namen. Podatki so osnova informacijske tehnologije. Računalniška informacijska revolucija in globalizacija sta povzročili, da so združbe začele zbirati podatke iz različnih virov. Pozornost združb je bila na začetku usmerjena predvsem v zbiranje podatkov, nihče, oziroma zelo malo združb, pa se je dejansko ukvarjalo s kakovostjo podatkov. Bilo je samo še vprašanje časa, kdaj se bodo združbe v hudem tekmovalnem boju osredotočile na kakovost podatkov v smislu vzpostavitve mehanizma obvladovanja kakovosti podatkov (Bowen, Fuhrer & Guess, 1998).

Tudi na raziskovalnem področju kakovost podatkov dolgo časa ni bila popularna, saj so se raziskovalci posvečali predvsem umestitvi informacijskih sistemov v združbe. Danes je kakovost podatkov izredno aktualno, pomembno in zelo hitro razvijajoče se področje – eni jo sicer dojemajo samo kot trenutno »modno« muho, in če se s tem ukvarja konkurenca, zakaj se ne bi tudi oni. Drugi pa z njo resnično želijo izboljšati svoje poslovanje, zmanjšati stroške in postati bolj konkurenčni – postati boljši. Če leta 2006 avtorja Batini in Scannapieca (2006, str. 1) navajata, da iskana beseda »Data Quality« v enem od popularnih internetnih iskalnikov vrne približno 3 milijone zadetkov, vrne marca 2011 enaka iskana beseda več kot 7 milijonov zadetkov. Pomembnost tega področja nakazujejo tudi ukrepi Evropske unije, na primer EUROSTAT na svojih spletnih straneh navaja zagotavljanje kakovostnih podatkov kot enega glavnih ciljev delovanja te združbe.

Da bi zaščitile vlagatelje in preprečile bančne neuspehe, urejajo vse države finančne dejavnosti v okviru svojih pristojnosti. V zadnjem desetletju se je količina prostovoljnih ali obveznih regulatornih standardov, s katerimi se srečajo ponudniki finančnih storitev, zvišala kar nekajkrat. Torej lahko ponudnikom finančnih storitev oprostimo, če imajo občutek, da so preobremenjeni z nenehnim usklajevanjem svojega poslovanja z regulatornimi zahtevami. Vendar s to »skladnostno utrujenostjo« tvegamo – pripelje nas lahko do tega, da bodo ponudniki finančnih storitev sprejeli samo minimalne ukrepe in s tem zadostili le regulatorni skladnosti, ali pa sprejeli nove standarde samo za javne namene. Standardi, ki so trenutno na mizi, predstavljajo priložnost za preoblikovanje razumevanja poslovanja, radikalno izboljšanje zmogljivosti modela podatkov in analiz podatkov, v zameno za bolj donosni raspored virov v celotni združbi ter večjo možnost za dobiček.

V magistrskem delu obravnavam izzive, ki jih za ponudnike finančnih storitev predstavljajo kapitalski sporazum Basel II in druge pobude po kapitalski ureditvi, predvsem izzive, vezane na podatke, kakovost podatkov in kakovost informacij, dobljenih iz teh podatkov. Basel II predstavlja korenito spremembo kapitalske ureditve finančnih storitev, je pa tudi najpomembnejši poslovni izziv, katerega daljnosežne posledice vplivajo na strukturo trga, načrtovanje izdelkov oziroma storitev, izbiro strank in oblikovanje cen. Z osredotočenjem na spremembe, s katerimi bi dosegli najugodnejši vpliv Basla II na

poslovanje, lahko Basel II namesto regulatorne obremenitve postane konkurenčna prednost, vendar je to na voljo samo ponudnikom finančnih storitev, katerih finančni podatki so dovolj kakovostni za uporabno analizo. Ankete kažejo (AIM Software, 2007), da visok odstotek ponudnikov finančnih storitev uporablja neustrezne podatke v modelu njihove finančne transakcije, in posledično zamujajo priložnost za ustvarjanje večjega dobička na različnih poslovnih področjih.

Basel II ponuja priložnost za razvoj natančnih podatkovnih modelov in tako razvoj bolj dobičkonosne poslovne strategije. Od leta 2007 so nekatere oblike skladnosti z Baslom II obvezne, vendar je v interesu komercialnega ponudnika finančnih storitev izboljšanje podatkov. Iz rezultatov katerekoli analize stroškov in koristi lahko sklepamo, da pridobijo pomembne prednosti tisti ponudniki finančnih storitev, ki prej dosežejo večjo skladnost z Baslom II.

Kot sem že poudaril, je kakovost podatkov pri Baslu II ključnega pomena. Konkretno obravnavam manjšo banko v lastništvu velike tuje bančne skupine, ki ima hčerinske banke v skoraj vseh državah vzhodne in centralne Evrope. Glede na to, da je banka del tuje bančne skupine, mora upoštevati pravila poslovanja, ki jih določa matična banka. Zelo pomembno je, da so podatki, ki jih pošljemo v skupen Basel II sistem, kakovostni: natančni, objektivni, vredni zaupanja in da ohranjajo ugled banke. Največ problemov izhaja iz naslova nacionalnih diskrecij. Na nivoju bančne skupine je izdelana skupna rešitev za vse članice skupine (pošiljanje in obdelava podatkov, izračun kapitalskih zahtev, itd.), vendar imajo posamezne banke lastno in raznoliko lokalno zakonodajo, ki jo morajo upoštevati pri izračunu svojih kapitalskih zahtev. Ukvarjam se tudi z vprašanjem, v kakšni višini je sploh smotrno zagotavljati kakovost podatkov in na kakšen način. Za doseganje kakovosti podatkov je namreč potrebno upoštevati več vidikov, in ne le kaj nam ponuja tehnologija sama po sebi. Predvsem je narobe pričakovati, da bo kakšna programska rešitev uredila problem nekovostnih podatkov. **Namen magistrskega dela** je pokazati pomembnost kakovosti podatkov, ki neposredno ali posredno vplivajo na učinkovitost poslovanja, oziroma na skladnost z Baslom II, seznaniti poslovodstvo s stroški pri eventuelni uporabi nekovostnih podatkov ter vsakemu sodelujočemu v življenjskem ciklu podatkov približati pojem kakovosti podatkov ter jim določiti pomembnost.

Cilj. Cilj magistrskega dela je ustvariti okvir za ocenjevanje kakovosti podatkov in zagotoviti postopke, ki omogočajo dosledno in natančno oceno kakovosti podatkov vseh zbirk podatkov, potrebnih pri novem kapitalskem sporazumu, ki bodo omogočile zagotoviti, da bo imela banka na voljo kapital, ki ustreza njeni tveganosti in poslovni strategiji.

Metode dela. Metodološki prijem pri izdelavi magistrskega dela vsebuje teoretično-analitični ter izkustveni del. Teoretično-analitični del vsebuje pregled pomembnejše svetovne literature, ki se nanaša na izbrano problematiko. Pri tem uporabim metodo

opisovanja. Ta je primerna, kadar gre za opis danih dejstev, ki jih kaže upoštevati pri razvijanju spoznanj in modela za razrešitev izbranega problema raziskovanja. V teoretičnem delu se osredotočam na ključne pojme, ki so vezani na kapitalski sporazum in kakovost podatkov. Pri analitičnem in izkustvenem delu, pri samem kreiranju okvira za ocenjevanje kakovosti podatkov, se bom naslanjal na že obstoječe metodologije (Total Information Quality Management, v nadaljevanju TIQM, Total Data Quality Management, v nadaljevanju TDQM, Methodology for the Quality Assessment of Financial Data, v nadaljevanju QAFD, The Datawarehouse Quality Methodology, v nadaljevanju DWQ,...).

Magistrsko delo je sestavljeno iz štirih poglavij, v katerih se skozi poglavja prepletajo obravnavane teme posameznega področja. V uvodu je predstavljena proučevana problematika in opredeljen načrt raziskave. V prvem poglavju bom teoretično razložil preteklost, namen, učinke ter prihodnost baselskih smernic. Da bi bolje razumeli vlogo podatkov, je v drugem poglavju nekaj pozornosti namenjeno osnovnim pojmom v zvezi s podatki, in sicer o zaznavanju stvarnega sveta ter prenosu na abstraktno raven, o odločanju, kje potrebujemo podatke oziroma informacije, kaj pomenijo neakovosti podatki za združbo in s kakšnimi stroški so povezane neakovostne informacije. V tretjem poglavju so opredeljena teoretična izhodišča kakovosti podatkov, pogled na kakovost z vidika združbe in posameznikov ter dejavniki, ki posredno ali neposredno vplivajo na kakovost. Izhodiščna metodologija pri pridobivanju in izboljšavi kakovosti informacij je metodologija TIQM, ki je hkrati glavno vodilo pri definiranju okvira za ocenjevanje kakovosti podatkov v četrtem poglavju.

1 SPLOŠNO O BASELSKIH SMERNICAH

Kapital banke lahko zelo hitro uničijo nenadne spremembe v kakovosti in vrednosti premoženja. V zgodovini so bile bančne krize povezane z velikim gospodarskim razkrojem in recesijo, predvsem iz razloga, ker višino zahtevanega kapitala za banke regulirajo oblikovalci politike ter zahtevajo visoke standarde pri upravljanju z naložbami, vključno z upravljanjem likvidnosti, računovodstva, revizije in posojilne prakse.

1.1 Zgodovina in prihodnost

Baselski sporazumi so pripravljene s strani Baselskega odbora za nadzor bank (angl. *Basel Committee on Banking Supervision*, v nadaljevanju BCBS), ki je bil ustanovljen leta 1974 s strani guvernerjev centralnih bank desetih najrazvitejših držav, takoimenovana skupina G-10.

Kapitalska regulativa Basel I je začela veljati decembra 1992 (razvoj in posvetovanja od leta 1988). Cilji so bili zagotoviti, da razpolagajo banke z dovolj kapitala za absorbirajnje eventuelnih izgub, in s tem ne povzročajo sistemskih težav ter enake pogoje za vse udeležence na mednarodni ravni (v izogib konkurenčnim konfliktom).

Basel I je dal bankam možnost nadzora nad višino zahtevanega kapitala s premikanjem bilančnih sredstev po segmentih z različnimi utežmi ter možnostjo listninjenja sredstev in njihovim prestavljanjem na zunajbilanco – možnost ukinitve posredništva. Banke so hitro akumulirale kapital na podlagi predpisanih minimalnih kapitalskih zahtev, ampak kapital v bistvu ni imel omejujoč vpliv na tveganja banke. Glavna kritika sporazuma Basel I je bila, da premalo upošteva nekreditna tveganja, ki so pridobila na pomenu v zadnjih letih ter se poleg tega osredotoča zgolj na kvantitativno merjenje kapitalске ustreznosti, zanemarja pa kvalitativni vidik. Kapitalska regulativa naj bi sledila spremembam na trgu. Revidirani sporazum, znan kot Basel II, je stopil v veljavo junija 2004 (BCBS, 2004). Glavne novosti so poudarek na učinkovitejšem obravnavanju kreditnega tveganja, dodana je kapitalska zahteva za operativno tveganje.

Osrednja naloga kapitalskih regulativ je preprečevanje kriz. Pristopa po Baslu I in II očitno nista delovala zadostno – nista bila ustrezna, da bi preprečila vpliv gospodarske krize na poslovanje finančnih združb. Svetovno gospodarstvo in nekatere finančne združbe namreč še vedno okrevaajo od največje finančne krize po veliki depresiji.

Guvernerji centralnih bank in finančni regulatorji iz 27 držav so 13. septembra 2010 sprejeli strožja pravila glede kapitalске ustreznosti bank, imenovana Basel III. Po novem bo morala višina temeljnega kapitala Tier 1 za banke znašati vsaj 6 %, medtem ko je bila z dosedanjimi standardi predpisana pri 4 %. Strožja pravila se bodo začela postopno uvajati po letu 2013, višina temeljnega kapitala Tier 1 naj bi se na zahtevanih minimalnih 6 % povzpela do leta 2015. Z uporabo novih pravil naj bi preprečili izbruh morebitnih novih finančnih kriz.

Pravila v finančnem sektorju je bilo nujno potrebno obnoviti, rezultat teh obnovitev in posodobitev je Basel III. Poleg samega sprejema pravil je mogoče še pomembnejša njihova uveljavitev. Lahko imamo najboljša pravila za banke na svetu, vendar bodo nekoristna, če ne bomo nadzorovali njihovega upoštevanja.

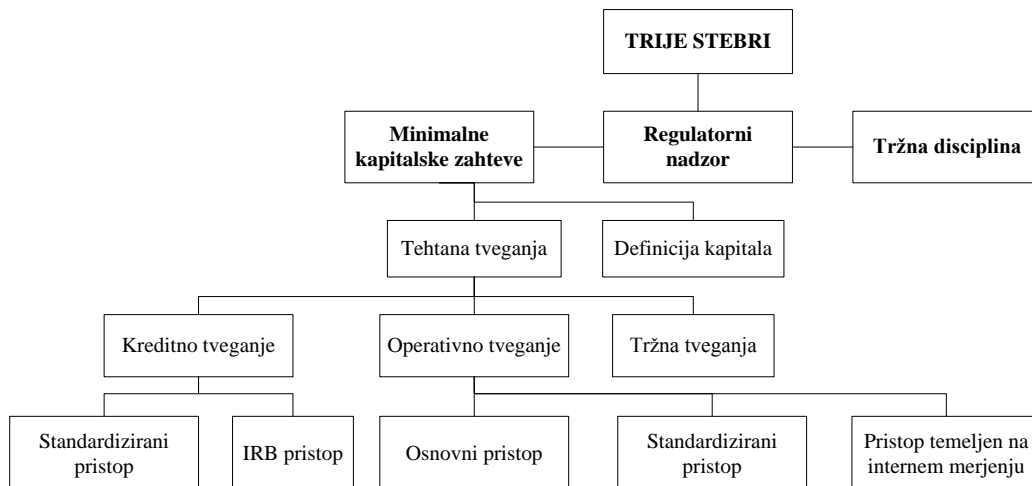
1.2 Zahteve Basla II

Kapitalski sporazum Basel II je priporočilo za oblikovanje politik v zvezi z obvladovanjem tveganj in določa, koliko kapitala mora imeti banka oziroma druga finančna združba glede na nivo finančnih in operativnih tveganj, s katerimi se sooča pri svojem poslovanju. S tem so postavljene zahteve za upravljanje s prevzetimi tveganji in izračun potrebnega kapitala, ki naj bi zagotavljale stabilnost združbe in s tem finančnega sistema.

Sporazum ni zavezujoč, torej ni zakonska obveza za banke oziroma druge finančne združbe, vendar je mednarodno uveljavljen in se ga uporablja po vsem svetu. Uporabljen je kot podlaga za oblikovanje evropskih bančnih smernic oziroma direktiv, ki jih morajo

upoštevati članice Evropske unije (v nadaljevanju EU) v svoji zakonodaji. Basel II je bil oblikovan leta 2004 (sprejet 2006). V Sloveniji ga banke uporabljajo od leta 2007.

Slika 1: Tristeborni sistem določanja kapitalske ustreznosti



Vir: Banka Slovenije, *Posledice uveljavitve novega baselskega sporazuma za slovenske banke*, 2006.

Basel II je prinesel nov sistem ocenjevanja tveganj, ki temelji tako na kvantitativnih kot kvalitativnih merilih, in bolj upošteva sisteme obvladovanja tveganj v posameznih bankah, širi vlogo preostalih tržnih udeležencev in spodbuja boljši nadzor nad tveganji v bankah. Sporazum združuje vse tri vrste tveganj (tržno, kreditno in operativno) in temelji na treh stebrih (Tajnikar, 2009, str. 11):

- steber 1: **minimalne kapitalske zahteve** (angl. *Minimal Capital Requirements*),
- steber 2: **regulativni nadzor** (angl. *Supervisory Review*) in
- steber 3: **tržna disciplina** (angl. *Market Discipline*).

V primerjavi s predhodnikom (Basel I) se bistvene novosti navezujejo na zadnja dva stebra. Osrednji del še vedno predstavlja določanje minimalnih kapitalskih zahtev, kjer pa je, zaradi spremenjene metodologije merjenja tveganj, ugotavljanje različnih stopenj kreditnega tveganja bolj občutljivo in zato bolj prilagojeno realnosti. Temeljnemu kvantitavnemu stebru sta v novi shemi dodana dva kvalitativna stebra, regulativni nadzor in tržna disciplina.

Pri regulativnem nadzoru gre za preverjanje nadzornikov, ali banka korektno ocenjuje tveganja, ki jim je izpostavljena pri opravljanju svojih storitev. Nadzorniki imajo pooblastilo zahtevati višjo kapitalsko ustreznost od zakonsko določene, če je banka pomembna iz sistemskega vidika.

Vse pomembnejšo vlogo pri vzpodbujanju stabilnega in varnega finančnega sistema igra tudi tržna disciplina kot tretji steber Basla II. Eden izmed vzvodov za to je večja transparentnost poslovanja bank, kar pomeni večji obseg javnega razkritja ter objave podatkov in informacij, povezanih z bančnim poslovanjem. Tržna disciplina zahteva aktivnejšo vlogo tržnih udeležencev, ki preko svojih poslovnih odločitev »nagrajujejo« banke, katerih poslovanje je transparentno, in »kaznujejo nedisciplinirane« banke (Sušnik, 2002, str. 2).

1.2.1 Minimalne kapitalske zahteve – prvi steber

Prvi steber Basla II določa, da mora koeficient kapitalske ustreznosti, ki ga izračunamo kot razmerje med celotnim kapitalom in tehtano aktivo, predstavljati najmanj 8 % celotnega kapitala banke. Tehtana aktiva je seštevek knjigovodskih stanj vseh aktivnih bilančnih in zunajbilančnih postavk, zmanjšanih za oblikovane oslabitve oziroma rezervacije za te postavke in tehtanih po stopnji tveganosti.

Koeficient kapitalske ustreznosti je prikazan v naslednji enačbi (Secretariat of the Basel Committee on Banking Supervision, 2001, str. 3)

$$\text{Min.kap. koef. (8 \%)} = \frac{\text{celotni kapital}}{\text{kreditno tveganje} + \text{tržna tveganja} + \text{operativno tveganje}} \quad (1)$$

Za merjenje kapitalske zahteve za kreditno tveganje predvideva Basel II naslednje možne pristope (Secretariat of the Basel Committee on Banking Supervision, 2001, str. 3):

- **Standardiziran pristop** in
- **Pristop temelječ na internih ratingih** (v nadaljevanju IRB), ki se deli še na:
 - Osnovni pristop in
 - Napredne pristope.

Za merjenje kapitalske zahteve za tržna tveganja obstajata dva pristopa (Standardiziran pristop in Pristop temelječ na internih modelih), za merjenje kapitalske zahteve za operativno tveganje pa so možni sledeči pristopi (Basel Committee on Banking Supervision, 2001, str. 28):

- **Enostavni pristop** (angl. *Basic Indicator Approach*),
- **Standardiziran pristop** (angl. *Standardised Approach*),
- **Napredni pristopi** (angl. *Advanced Measurement Approaches*), ki se deli še na:
 - Pristop z internim merjenjem (angl. *Internal Measurement Approach*),
 - Točkovaalni pristop (angl. *Scorecard Approach*),
 - Pristop s statistično porazdelitvijo izgube (angl. *Loss Distribution Approach*).

Basel II daje velik poudarek spremljanju in obvladovanju operativnega tveganja. Obstajata dve težavi, vezani na obvladovanje operativnega tveganja: definicija tega tveganja, ki je zelo široka in zato tudi ohlapna, in (posledično izhaja iz prve) merjenje tega tveganja. Za merjenje in upravljanje operativnega tveganja ni neke splošno uveljavljene metodologije, ki bi veljala za dobro bančno prakso. Razen tega na tem področju še vedno zeva ogromna podatkovna vrzel, saj niti največje svetovne banke tega tveganja še ne spremljajo dalj časa ali vsaj ne sistematično. Ključnega pomena pri obvladovanju operativnega tveganja so vodstveni organi banke in sistem notranjih kontrol. Značilnost tega tveganja je, da v večini primerov nastopa skupaj z drugimi tveganji, na primer tržnim in kreditnim tveganjem. Dejavniki tveganja so interne narave (kakovost notranje revizije, obseg poslovanja, število transakcij, stopnja storjenih napak, ipd.), pri čemer korelacija med dejavniki in višino tveganja večinoma ni znana, kar onemogoča merjenje tveganja (Majič, 2002, str. 59).

1.2.2 Regulativni nadzor – drugi steber

Drugi steber predstavlja kvalitativni dodatek k prvemu. Regulativni nadzor zahteva od nadzornikov da preverjajo, ali ima vsaka banka ustrezen notranji proces ocenjevanja tveganja in računanja kapitalskih zahtev (Secretariat of the Basel Committee on Banking Supervision, 2001, str. 5).

8-odstotni koeficient kapitalske ustreznosti je le potrebni, ne pa tudi zadostni pogoj za zagotovitev varnega poslovanja banke. Nadzorniki imajo pooblastilo zahtevati višjo kapitalsko ustreznost, kot jo določi banka sama ali kot jo izračuna v skladu z izbrano metodologijo. To je posebej pomembno takrat, ko je banka pomembna iz systemskega vidika in bi njene težave lahko ogrozile stabilnost finančnega sistema (Sušnik, 2001, str. 41). Cilj nadzornikov pri preverjanju kapitalske moči bank na tem področju je zagotoviti, da je obseg kapitala skladen s tveganostjo njihovega celotnega poslovanja.

Vsebina regulativnega nadzora je strnjena v štirih temeljnih načelih (angl. »*Four key principles of supervisory review*«), ki se medsebojno dopolnjujejo (Basel Committee on Banking Supervision, 2001, str. 30–31):

1. Banke absolutni znesek svojega kapitala prilagajajo tveganosti svojega poslovanja tako, da ves čas poslujejo nad predpisanim minimumom; v nasprotnem primeru imajo nadzorniki moč zahtevati povečanje kapitala nad omenjeno minimalno raven.
2. V bankah morajo biti vzpostavljeni interni sistemi za celostno oceno njihove kapitalske ustreznosti, izdelane morajo biti strategije za doseganje in ohranjanje njihove želene oziroma ustrezne kapitalske moči.
3. Nadzorniki morajo preverjati in vrednotiti interne ocene kapitalske ustreznosti banke ter njene strategije, pa tudi izpolnjevanje predpisanega minimuma.
4. Nadzorniki morajo ukrepati takoj, ko obstaja nevarnost, da kapital banke pade pod raven, ki jo označujejo za varno.

Drugi steber naj bi v bistvu pomenil formalizacijo obstoječih nadzornih praks na področju nadzora kapitala in kapitalne ustreznosti bank. Minimalni standardi, ki naj bi jih uporabljali nadzorniki pri ocenjevanju kapitala, naj bi izhajali iz revizije obstoječih tehnik nadzora, njihovih nadgraditev oziroma dopolnitev ter razvoja novih pristopov, temelječih na bančni praksi in izkušnjah. Nadzorniki se pri spremljanju poslovanja bank opirajo na svojo presojo, temelječo na poznavanju in poglobljenem razumevanju razmer v banki, zato je nadzor fleksibilen in hkrati individualen.

Nadzor je tudi delovno precej bolj intenziven od regulative, saj zahteva številne dobro usposobljene in izkušene nadzornike. Medtem, ko daje shema za merjenje tveganj iz Basla I relativno objektivne rezultate, ki jim je težko oporekati, pa regulativni nadzor iz Basla II od nadzornikov zahteva odločitve na podlagi presoje, ki jih je težavno zagovarjati zaradi subjektivne narave procesa njihovega oblikovanja. To še posebej velja za primere, ko bi nadzorniki zahtevali od bank višji koeficient kapitalne ustreznosti od minimalno zahtevanega ali takšnega, ki bi ga banke s stroškovnega vidika še lahko prenesle (Sušnik, 2001, str. 32).

1.2.3 Tržna disciplina – tretji steber

Tretji steber povečuje tržno disciplino s povečanjem obsega podatkov, ki jih morajo banke razkriti. Večji obseg javnega razkritja ter objave podatkov in informacij, povezanih z bančnim poslovanjem in profilom tveganja banke, naj bi tržnim udeležencem omogočili boljši pregled nad poslovanjem bank. Banke morajo razkriti področje delovanja, strukturo kapitala in kapitalno ustreznost, način izračunavanja kapitalne ustreznosti ter način upravljanja s tveganji (Basel Committee on Banking Supervision, 2001, str. 1).

Banka, ki velja za varnejšo in uspešno voden, je močnejši pogajalec v razmerju do investitorjev, kreditojemalcev, depozitarjev in drugih strank, razen tega tudi tržni pritiski silijo banke v učinkovito alokacijo sredstev in uporabo kapitala. Tržna disciplina naj bi tako z vzpostavitvijo tretjega stebra prvič dobila eksplicitno opredeljeno vlogo v kapitalnih standardih (Sušnik, 2001, str. 32).

Vključitev tržnih udeležencev kot opazovalcev upravljanja s tveganji in procesa interne alokacije kapitala v bankah, naj bi bilo pomembno dopolnilo k regulativnemu nadzoru (drugi steber). Žal pa tržna disciplina navadno ne deluje prav v tistih državah, kjer bi bilo to najbolj potrebno – to je v državah s šibkim nadzornim sistemom. Odprtje bančnega sektorja tuji konkurenci prispeva k okrepitvi učinkovite tržne discipline, saj večja konkurenca na trgu prisili tudi domače banke k izboljšanju njihovega poslovanja in k bolj preglednemu komuniciranju z vsemi zainteresiranimi javnostmi. K večji tržni disciplini naj bi največ prispevali predvsem tržni udeleženci sami. Razkritje in preglednost pa sta le potrebna, vendar ne tudi zadostna pogoja za učinkovito delovanje tržne discipline. Tržni udeleženci morajo biti motivirani, da analizirajo razpoložljive informacije. S tem, ko se

odzivajo nanje, dosegajo večjo disciplino. Tržna disciplina namreč deluje v praksi le v primeru, če se udeleženci trga odzovejo na tržne signale s spremembo svojega vedenja. Tretji steber torej prinaša povečanje javnih razkritij, kar pomeni javne objave podatkov in informacij s strani bank. V zvezi s tem pa banke skrbijo stroški zbiranja velikega obsega podatkov in vprašanje zaupnosti podatkov.

1.3 Basel III

Predstavniki najvplivnejših svetovnih centralnih bank in regulatorji so 13.09.2010 v Baslu dosegli soglasje o novem kapitalskem sporazumu (Basel III). Da bi preprečili možnost ponovnih zlomov bank, kot smo jim bili priča v letu 2008, ter banke naredili odporne na udarce in krize, bodo morale banke v prihodnosti potrojiti svoje kapitalske rezerve kot previdnostni ukrep morebitnih izgub.

Paket ukrepov, pod imenom Basel III, bo zvišal minimalni kapital na 4,5 % (tako imenovani Core Tier I), kar je več kot podvojeno v primerjavi z 2 % iz Basla II (plus 2,5 % dodatnega kapitala za zaščito). Če pade količnik kapitalske ustreznosti pod 7 %, bo banki prepovedano izplačevanje dividend delničarjem in nagrad poslovodstvu, vendar pa banka ne bo imela nobene obveze zbiranja svežega kapitala.

Minimalni koeficient kapitalske ustreznosti ostaja 8 %. Večina držav, ki jih je prizadela bančna kriza, vključno z ZDA in Veliko Britanijo, se je zavzemala za hitro uvedbo strožjih standardov, vendar je temu močno nasprotovala Nemčija, tako da se bodo nova pravila uvajala postopoma, z začetkom v letu 2013, pa vse do leta 2019. Basel III naj bi temeljito okrepil mednarodne kapitalske standarde, ki bodo prispevali k dolgoročni finančni stabilnosti in rasti.

Na drugi strani bi lahko strožja merila in pospešitev izvajanja Basla III upočasila dajanje posojil in ustavila že tako ohlapno okrevanje svetovnega gospodarstva. To je razlog za relativno dolgo časovno obdobje pred pričetkom izvajanja novih pravil – analitiki ocenjujejo to kot uslugo bančnemu lobiju.

2 OBLIKOVANJE OKOLJA ZA KAKOVOST PODATKOV

Bivstvenega pomena za uresničevanje Basla II so kakovostni podatki, iz katerih lahko dobimo kakovostne in uporabne informacije. Bistvo Basla II je dobra analiza tveganja, ki pa je odvisna od kakovostnih podatkov, to pomeni razpoložljivosti pomembnih in potrebnih podatkov za učinkovito analizo in modeliranje. Vsaka zamuda pri procesu čiščenja in konsolidacije podatkov, da bodo ustrezali baselskim merilom tveganja, bo vodila k zamudi pri končni skladnosti s kapitalskimi zahtevami, višjim stroškom in zamujeni priložnosti prihodkov.

Da bi razumeli kakovost podatkov in kakovost informacij, je potrebno opredeliti njuna ključna pojma – podatke in informacije. Opredeljeni so tudi pojmi znanje, modrost in kakovost. Informacije bistveno vplivajo na uspešnost poslovanja; nasprotno lahko nekakovostne informacije škodijo uspešnosti poslovanja.

2.1 Podatki

Podatek (lat. *Datum* – kaj pomeni »nekaj dano«) je trditev, sprejeta kot nominalna vrednost. V kontekstu klasične računalniške znanosti pojem podatek ponavadi pomeni številko ali druge informacije, ki so zastopane na način, da jih lahko računalnik procesira. Podatke lahko opredelimo tudi iz poslovnega vidika, neodvisno od informacijske tehnologije. Preprosto povedano, podatek je predstavitev dejstva o stvareh.

Podatki so surovine, iz katerih pridobimo informacije, ki so osnova za inteligentna dejanja in odločitve. Na primer, podatek 112 predstavlja dejstvo in je v realnem svetu brez opisne opredelitve ali konteksta nesmiseln.

Podatki predstavljajo stvari ali entitete v resničnem svetu. Wikipedia (2009) definira entiteto kot filozofski pojem, ki je vezan na obstojnost česarkoli v danem prostoru in času, pa kljub temu ni nujno vezano na materialno pojavnost; nekaj kar je, kar obstaja. Beseda entiteta se uporablja v mnogih področjih, kjer običajno z ustreznimi pridevniki ali kot pridevnik sam tvori smiselno območje ali obseg. Pojem entiteta navadno obravnavamo kot neko smiselno zaključeno enoto, ki je lahko sestavljena iz več podenot. Na primer v času, ko to pišem, sem jaz, Nikola, izredni študent Ekonomske fakultete v Ljubljani. Nikola in fakulteta sta entiteti, pomeni da obstajava.

2.2 Informacije

Če so podatki surovina, je informacija končni izdelek. Informacija je podatek v kontekstu, uporaben podatek. Podatek dobi pomen, dejstva postanejo razumljiva. Prejšnji primer podatkov postane razumljiva informacija, kadar nekdo ve, da je 112 telefonska številka za klic v sili.

Kakovost informacij zahteva kakovost treh komponent (English, 1999, str. 18):

- **jasno opredelitev ali pomen podatkov,**
- **pravilno vrednost,**
- **razumljivo predstavitev** (oblika predstavljanja intelektualnim delavcem – knowledge worker).

Nekakovost katerekoli od teh treh komponent lahko povzroči propad poslovnega procesa ali napačno odločitev. Informacija je uporaben podatek in jo lahko predstavimo kot enačbo (English, 1999, str. 19):

$$\text{Informacija} = f(\text{Podatki} + \text{Opredelitev} + \text{Predstavitev}) \quad (2)$$

S poslovnega vidika je informacija lahko opredeljena, vrednosti so natančne, predstavitev smiselna, vendar še vedno ne more biti dragocen vir za združbo. Kakovostna informacija je sama po sebi nekoristna, vendar razumljiva ljudem in lahko vodi do dodane vrednosti.

2.3 Znanje

Kakovostna informacija postane močan vir, ko jo lahko ljudje asimilirajo. Intelektualni delavci (angl. *knowledge workers*) plus kakovostne informacije dajo potencial, da dobijo informacije vrednost. Baze podatkov brez intelektualnih delavcev, ki jih uporabljajo, pomenijo enako vrednost kot banka brez denarja.

Znanje ni samo znana informacija, temveč informacija v kontekstu, razumevanje pomena informacije in lahko se predstavi kot enačba (English, 1999, str. 20):

$$\text{Znanje} = f(\text{Ljudje} + \text{Informacija} + \text{Pomen}) \quad (3)$$

Znanje je dodana vrednost informaciji s strani ljudi, ki imajo izkušnje in sposobnosti razumeti njihov pravi potencial. Z neprestanim razvojem informacijske tehnologije so združbe zmožne zajemati znanje, ga shraniti in deliti. Napredek interneta, intraneta, podatkovnega rudarjenja, itd. širijo obzorja »deljenih« podatkov v podatkovnih skladiščih in operativno – transakcijskih (v nadaljevanju OLTP) podatkovnih bazah.

Združba ima lahko veliko znanja, vendar to še ne pomeni, da bo uspešno. Znanje ima vrednost le v primeru, da imajo ljudje pooblastila delovati v obsegu svojega znanja. Rečemo lahko, da ima znanje vrednost samo takrat, ko deluje.

2.4 Modrost

Cilj vsake združbe je maksimiranje vrednosti njenih sredstev za izpolnjevanje poslanstva. Vir informacij je maksimiran, ko je upravljan na način, da je kakovosten, in ki omogoča enostaven dostop tistim, ki jih potrebujejo. Človeški viri v združbi so maksimirani, ko so zaposleni ustrezno strokovno usposobljeni, imajo dostopne vire (zagotovljena sredstva) vključno z informacijami ter so pooblaščen za delovanje, izvajanje svojega dela in zadovoljevanje končnega uporabnika.

Modrost je uporabljeno znanje in se lahko izrazi s enačbo (English, 1999, str. 20):

$$\text{Modrost} = f(\text{Ljudje} + \text{Znanje} + \text{Delovanje}) \quad (4)$$

Cilj kakovostnih informacij je opremiti intelektualne delavce s strateškimi viri in tako omogoči inteligentno učečo se združbo. Peter Senge opredeljuje učečo se združbo kot združbo, katera »nenehno širi svojo sposobnost za oblikovanje svoje prihodnosti« s pomočjo učenja in skupnega učenja. Inteligentna učeča se združba je tista, ki najbolj maksimira tako svoje izkušnje kot vire informacij v procesu učenja.

2.5 Kakovost

Ljudje zavestno ali podzavestno sodimo kakovost stvari na podlagi svojih preteklih izkušenj. Zavestna uporaba merjenja kakovosti je, ko na primer oseba primerja proizvode in izbere enega od njih kot »boljši« izdelek. »Boljši« pomeni izdelek, ki najbolj ustreza njenim splošnim potrebam, vendar to ne pomeni, da ima nujno najboljše značilnosti v vsaki kategoriji. Po nakupu določamo utemeljeno kakovost s tem, ali izdelek za svojo ceno izpolnjuje naša pričakovanja za nameravano uporabo. Nezavestno merjenje kakovosti je frustracija, povezana z nakupom nekakovostnega izdelka ali storitve (English, 1999). Primer nekakovostne storitve: čakanje v dolgi vrsti pred blagajniškim okencem banke, medtem ko se drugi zaposleni privatno pogovarja po telefonu, medtem ko bi lahko pomagal. Takšen primer nam sporoča, da banka ne skrbi za čas svojih strank.

Kakovost ni razkošje ali superiornost. Kakovost obstaja le v očeh uporabnikov in temelji na vrednosti, kakšno mnenje imajo uporabniki o tem, kako nekaj ustreza njihovim potrebam. Kar je kakovost za enega uporabnika, je lahko za drugega ravno nasprotno.

Kakovost ni primernost za namen. Na primer pri vnosu posebne položnice na spletni strani banke X, ima polje »Vrsta posla« vpisano privzeto vrednost; večina uporabnikov vrednosti ne spreminja, ker se položnica s tako vrednostjo omenjenega polja brez težav obdela. Rečemo lahko, da je vrednost v polju »primerna namenu« plačilo, vendar ni primerna za kakršnokoli analizo. Kakovost pomeni primernost za vse namene iz podatkov, vključno z verjetno prihodnjo uporabo. Kakovostno informiranje bodo uporabili v mnogih novih načinih za inteligentne učeče se združbe. Kakovost pomeni tudi srečanje potreb uporabnikov, vendar ne nujno tudi njihovo preseganje.

2.6 Opredelitev kakovosti informacij

2.6.1 Naravna kakovost informacij

Naravna kakovost informacij (angl. *inherent*) je, preprosto povedano, točnost podatkov oziroma stopnja, do katere podatkovna točnost v realnem svetu odraža predmet, katerega podatki ga predstavljajo (English, 1999). Vsi podatki so abstrakcije ali predstavitve nečesa realnega. Jean Baudrillard, francoski semiologist (semiologija je veda o znakih), ugotavlja,

da »natančna opredelitev resničnega postane ta, iz katere je mogoče dati enako reprodukcijo«.

Če so vsa dejstva, ki jih mora o entiteti poznati združba, točna, so ti podatki naravno kakovostni – to je elektronska reprodukcija resničnosti. Naravna kakovost informacij pomeni, da so podatki pravilni.

2.6.2 Pragmatična ali stvarna kakovost informacij

Pragmatična kakovost informacije je stopnja uporabnosti in vrednosti podatkov za podporo procesov združbe, ki omogočajo uresničevanje njegovih ciljev. V bistvu je pragmatična kakovost informacij stopnja uporabnikovega zadovoljstva, ki so jo pridobili intelektualni delavci pri opravljanju svojega dela.

Podatki v bazah ali skladiščih podatkov nimajo nobene dejanske vrednosti; imajo samo potencial vrednosti. Podatek ima spoznano (realizirano) vrednost samo, ko ga nekdo uporablja za kaj uporabnega; na primer za pošiljanje nove kreditne kartice komitentu banke, ali za določitev pravilnega ratinga komitenta. Pragmatična kakovost informacije je stopnja, do katere omogočajo podatki intelektualnim delavcem doseganje ciljev združbe na zadovoljiv in učinkovit način.

Kakovost informacije leži v njeni sposobnosti, da zadovoljuje svoje uporabnike; tiste, ki uporabljajo podatke pri delu. Na primer če je komitent banke odplačal kredit in je banka potrdila, da je bil komitent reden plačnik (ni imel težav pri odplačevanju, ni zamujal več kot 90 dni s plačilom obroka), je ta podatek naravno kakovosten – je pravilen. Če uporablja banka ta podatek kot sredstvo za samodejno pošiljanje ponudbe komitentom za nov kredit, ima ta podatek pragmatično kakovost informacije. Ob uporabi pravilnih podatkov o vrednostih, je banki omogočeno, da izpolni cilj – poveča število kreditov z manjšim tveganjem.

Možno je, da imamo naravno kakovostno informacijo brez tega, da imamo pragmatično kakovost informacije. Podatek, ki ni potreben za podporo poslovnim procesom, kakršnikoli odločitvi, ali nam ne koristi v nobeni analizi, je nepomemben. Tudi če so vrednosti pravilne in je zato naravno kakovosten, je takšen podatek nekoristen in za združbo nima vrednosti. Dejansko je nekakovostna informacija za združbo ta, ki za pridobivanje in vzdrževanje potrebuje denar in sredstva, vendar ne daje nobene dodane vrednosti. Ima celo negativno neto vrednost. Če na primer banka zbira podatke o barvi oči komitentov (ta podatek pridobi pri sklepanju posla), vendar to dejstvo (podatek) ni uporabljeno v njenih poslovnih procesih, manjka kakovost. Zbiranje takega podatka pravzaprav povečuje stroške poslovanja banke in se odraža v višjih stroških za komitenta.

2.6.3 Opredelitev kakovosti informacije

Enako izhodišče kot za kakovost potrošniških proizvodov velja tudi za kakovost informacije. Če želimo opredeliti kakovost informacije, moramo opredeliti uporabnike podatkov, tako imenovane intelektualne delavce, ki potrebujejo podatke za opravljanje svojega dela. Kakovost informacij je »dosledno izpolnjevanje pričakovanja intelektualnih delavcev in končnega uporabnika« skozi informacije in informacijske storitve, ki jim bodo omogočale opravljanje njihovega dela uspešno in učinkovito. Kakovost informacij opisuje attribute informacij, ki izhajajo iz uporabnikovega zadovoljstva.

Kakovost informacij obstaja, kadar omogoča informacija intelektualnim delavcem da izpolnijo svoje »poslovne« cilje. Kakovost informacij se ne meri samo po neposrednih upravičencih, temveč tudi po navzdolnjih intelektualnih delavcih.

V nadaljevanju so proučeni elementi kakovosti informacij: dosledno izpolnjevanje pričakovanih intelektualnih delavcev in končnega uporabnika.

2.6.3.1 Doslednost

Ko intelektualni delavci dobijo informacije o entiteti ali dogodku, pričakujejo dosledno kakovost, da že vnaprej poznajo raven kakovosti podatkov, s katerimi delajo. Za nekatere procese pri podpori odločanja lahko intelektualni delavci prenašajo določeno stopnjo napake in opustitve, če se zavedajo stopnje in vrste napake. Če v zanesljivosti podatkov obstajajo velika nihanja, lahko intelektualni delavci uporabijo svojo intuicijo kot odločilni sistem podpor, namesto da zaupajo nezanesljivim podatkom iz nezanesljivega elektronskega sistema (English, 1999).

Doslednost pomeni, da kakovost informacij izpolnjuje potrebe vseh intelektualnih delavcev in ne samo nekaterih. Kaj pomeni, če eni intelektualni delavci zahtevajo 95-odstotno natančnost, ostali pa 99-odstotno natančnost? To pomeni, da je za dosledno izpolnjevanje pričakovanih potrebna 99-odstotna natančnost. Dosledno tudi pomeni, da intelektualni delavci uporabljajo podatke o isti stvari iz dveh različnih baz podatkov, bodisi iz dveh OLTP ali operativne baze podatkov in podatkovnega skladišča in pričakujejo, da se podatki »ujemajo«. Če vzamemo na primer informacije o uporabniku iz centralnega informacijskega sistema, iz baze podatkov marketing oddelka, iz računovodske baze podatkov in iz podatkovnega skladišča, pričakujemo skladnost atributov, ki naj bi bili enaki v vseh štirih bazah podatkov.

Neuspeh pri vzdrževanju redundantnih baz podatkov še vedno ostaja ena od najbolj razširjenih težav kakovosti informacij. Če obstaja poslovni vidik izgradnje (ali nakupa) redundantne baze podatkov, obstaja tudi poslovni vidik vzdrževanja njene doslednosti.

2.6.3.2 Izpolnjevanje

Nekateri podatki zahtevajo 100-odstotno natančnost (angl. *zero-defect data*). Domena referenčni podatki, kot so na primer računi komitentov ali tarife, morajo imeti 100-odstotno natančnost, če želimo komitentom pravilno obračunati stroške. 100-odstotno natančni podatki so potrebni, če njihova nekakovost posledično povzroči odpoved glavnega procesa ali za banko katastrofalne posledice, kot na primer pomešani računi komitentov in posledično napačne transakcije.

Vendar ni za vse podatke zahtevano, da so popolni oziroma niti ni potrebno, da so ravno natančni. Veliko odločitev se lahko sprejme na podlagi podatkov iz podatkovnega skladišča, ki so nepopolni. Pravilno se lahko odločimo celo na podlagi podatkov, ki vsebujejo določeno stopnjo napake, v primeru, da je to upoštevano pri odločitvi – se tega zavedamo (ko vemo, da je odločitev sprejeta z določeno stopnjo napačnih podatkov).

Nekateri podatki, zlasti podatki o poslovnih dogodkih, morda ne bodo mogli biti zajeti po začetnem dogodku poslovne priložnosti brez obsežne preiskave in ponavljanja dogodka. Na primer, spremenljivke v znanstvenih eksperimentih, ki niso zajete v točki stika z dogodkom, se morda ne morejo ponoviti pri nobenem strošku. Tudi nameni, ki vodijo stranko, ko povprašuje po izdelku ali storitvi, so lahko za vedno izgubljeni, če niso zajeti tekom povpraševanja (Eppler, 1999).

2.6.3.3 Intelektualni delavec in končni uporabnik

Kdo lahko zazna kakovost informacij? Intelektualni delavci (angl. *knowledge workers*), ki potrebujejo podatke za opravljanje svojega dela. Izraz intelektualni delavec pomeni vlogo, v kateri nekdo zahteva ali uporablja podatke v kakršnikoli obliki, kot del funkcije njegovega delovnega mesta, ali v času izvajanja procesa. Torej je intelektualni delavec uporabnik informacije. Intelektualni delavci – uporabniki informacij ugotavljajo, ali so podatki kakovostni ali ne, temelječi na tem, kako dobro ti podatki podpirajo njihovo sposobnost za opravljanje dela.

Skoraj vsi zaposleni delavci so intelektualni delavci. Vodstvo, ki odloča, so očitno intelektualni delavci. Poslovni analitiki, ki potrebujejo natančne podatke o trendu, so glavni uporabniki podatkovnih skladišč, blagajniki, kreditni referenti,... Vsaka funkcija, ki spodbuja kakovost, mora imeti uporabnika za svoj edini cilj. Kakovostna funkcija, ki se ne osredotoča na potrebe in zahteve svojih uporabnikov, bo na koncu propadla.

Takojšne informacijske stranke so tisti intelektualni delavci, ki so v istem oddelku ali poslovnem področju kot proizvajalci podatkov. Na primer, delavci na šalterju banke so hkrati proizvajalci in intelektualni delavci podatkov o strankah. Šalterist ustvari zapis za stranko, ko stranka prvič posluje z banko. Za nadaljnje posle postane šalterist intelektualni

delavec, in samo pridobi dodatne podatke ter s stranko sklene nov posel. Ker proizvajalci podatkov o strankah tudi sami uporabljajo te informacije, je delež pravilno pridobljenih podatkov visok.

Oddelčni intelektualni delavci niso edini uporabniki podatkov. Podatke o strankah ne potrebujejo samo šalteristi, potrebujejo jih tudi oddelek za izvršbo, servisni oddelek, oddelek terjatev, oddelek trženja, oddelek kontrolinga in poročanja,... Ti navzdolnji intelektualni delavci prav tako pričakujejo kakovostne informacije za opravljanje svojega dela. Podatki v bazi podatkov o določeni entiteti so nekakovostni, če jih ne morejo uporabljati tudi drugi intelektualni delavci.

Kakovostne informacije so podatki, ki izpolnjujejo zahteve ne le neposrednega uporabnika, ampak tudi navzdolnjih uporabnikov informacij brez večjih transformacij. Stroški transformacijskih vmesnikov zmanjšujejo vrednosti podatkov, tako da zmanjšujejo dobiček, ki izhaja iz uporabe teh podatkov. Vmesniki predstavljajo tudi točke morebitne napake v procesu.

2.6.3.4 Pričakovanja

Vsi se strinjamo, da želijo vestni zaposleni narediti svoje delo dobro, in pričakujejo, da imajo na voljo potrebna sredstva za opravljanje svojega dela v zameno za pošteno plačilo. Intelektualni delavci, da bi svoje delo lahko opravljali uspešno in učinkovito, potrebujejo za opravljanje svojega dela visoko kakovostne informacije (vire).

Pravi cilj kakovosti informacij je povečati zadovoljstvo uporabnikov in delničarjev. Pravzaprav je mogoče videti in meriti kakovost informacije z zadovoljstvom strank. Recimo, stranka je preko spletne banke dala banki nalog za nakup funtov (menjava evro za funte) po tečaju, ki je bil objavljen na spletni banke. Banka čez kakšno uro ugotovi, da je tečaj napačen in o tej nerodni situaciji obvesti stranko. Seveda je stranka nezadovoljna zaradi nekakovostne informacije o tečaju za funt, podane na spletni banki, nezadovoljna pa je tudi banka, ker je uslužbenec svoje delo zelo slabo opravil in vpisal napačen tečaj ter zaradi škode provzročene ugledu banke in na koncu materialne škode zaradi razlike v tečaju. Kot vidimo, je nekakovostna informacija povzročila veliko nezadovoljstva.

2.6.4 Komponente kakovosti informacij

Prej smo pokazali, da je informacija lahko opredeljena s enačbo (English, 1999, str. 27):

$$\text{Informacija} = f(\text{Podatki} + \text{Opredelitev} + \text{Predstavitev}) \quad (5)$$

Tri komponente, ki sestavljajo končni proizvod, informacijo, so ločene in vsaka mora biti kakovostna, da dobimo kakovostno informacijo. Če ne poznamo pomena (opredelitve)

dejstva (podatka), bo vsaka vrednost brez pomena in smo nekakovostni. Če poznamo pomen (opredelitev) dejstva, vendar so vrednosti (podatki) nepravilni, smo nekakovostni. Če imamo pravilne vrednosti (podatki) za znana (definirana) dejstva, vendar je njihova predstavitev (bodisi kot pisno poročilo, na računalniškem zaslonu, ali računalniško ustvarjenih poročilih) nekakovostna, lahko intelektualni delavci napačno razložijo podatke, in smo spet nekakovostni.

2.7 Opredelitev podatkov in kakovost arhitekture informacij

Opredelitev podatkov se nanaša na specifikacijo podatkov - opredelitev, domene (področja) določenih vrednosti in poslovna pravila, ki urejajo podatke. Kakovost opredelitve podatkov je stopnja, do katere definicija podatkov natančno opisuje pomen tipa entitete v resničnem svetu. Med uporabnike informacij vključujemo poslovno in informacijsko sistemsko osebje (English, 1999, str. 28):

- **intelektualni delavci** morajo poznati pomen informacije, da lahko opravljajo svoje delo;
- **proizvajalec informacij** mora poznati pomen informacij, skupaj z veljavnimi vrednostmi in poslovnimi pravili, da jih lahko ustvari ali posodablja;
- **administratorji podatkov** morajo poznati pomen informacije, skupaj z veljavnimi vrednostmi in poslovnimi pravili, da bi razvili natančne podatkovne modele;
- **administratorji baz** morajo poznati pomen informacije skupaj z veljavnimi vrednostmi in poslovnimi pravili, da bi lahko oblikovali bazo;
- **sistemski analitiki** morajo poznati pomen informacije skupaj z veljavnimi vrednostmi in poslovnimi pravili, da bi oblikovali visoko celovit model aplikacije;
- **razvijalci aplikacij** morajo poznati pomen informacije, skupaj z veljavnimi vrednostmi in poslovnimi pravili, da bi razvili visoko celovito aplikacijsko logiko.

Kakovost arhitekture informacij je stopnja, do katere podatkovna struktura izvaja vsebinska in realna razmerja povezanih podatkov, ki predstavljajo realne predmete in dogodke. Njene lastnosti so (English, 1999, str. 28):

- je stabilna, omogoča novim aplikacijam ponovno uporabo izvirnih podatkov brez sprememb in zahteva samo kreiranje novih, neredundantnih entitet in novih lastnosti, ki se dodajo obstoječim podatkovnim modelom ali bazam podatkov. Stabilnost baze podatkov pomeni, da lahko nove aplikacije uporabljajo podatke iz obstoječih baz podatkov brez sprememb v strukturi podatkovnega modela ali baze podatkov, zadostuje že dodajanje novih podatkov;
- je prilagodljiva, podpira spremembe, ko izvaja združba svoje procese brez večjih sprememb podatkovnega modela ali baze podatkov. Prilagodljivost baze podatkov pomeni, da se lahko združita dve vrsti poslovanja in tako odpravi splošno podvajanje in

povečajo navzkrižne prodaje z minimalno spremembo strukture baze. Prilagodljivost baze podatkov pomeni, da lahko združbe spreminjajo procese z minimalno spremembo strukture podatkovne baze.

Za zagotovitev jasne komunikacije med vsemi upravitelji informacij je potrebna jasna in natančna opredelitev podatkov. Opredelitev podatkov je za podatke (vsebine) to, kar je za opredelitev besed slovar.

2.8 Kakovost predstavitve podatkov

Poslovni proces je lahko neuspešen, tudi če so podatki točni, popolni in ustrezajo jasnim natančnim opredelitvam. Procesi so neuspešni, če (English, 1999, str. 29):

- so podatki nedostopni;
- podatki niso na voljo pravočasno;
- so podatki predstavljeni na dvoumen način, so v neskladju z oznako podatka ali definicijo in povzročajo napačne razlage;
- so podatki predstavljeni na način, ki zahteva prekomerno delo z razlago, s čimer se uvaja možnost napak v dodatnih procesih, ki so potrebni, da bi bili podatki uporabni;
- so podatki napačno kombinirani z drugimi podatki, kar ima za posledico napačno izpeljane ali izračunane podatke.

Kakovost predstavitve podatkov se nanaša na dokumente in medije, ki nosijo informacije, kot so poročila ali forma, v kateri so predstavljeni rezultati poizvedbe iz baze podatkov. V praksi lahko velikokrat opazimo, kako različne aplikacije dajo različna obvestila po končanju določene akcije, na primer »Konec«, »Pripravljen«. Obvestilo lahko zmede uporabnika, lahko imamo različne odzive. Obvestilo »Konec« bi ponavadi pomenilo, »akcija je končana lahko delam naprej«. Na obvestilo »Pripravljen« bi bila lahko prva misel: »Pripravljen na kaj?«. Sporočilo »Pripravljen« je obvestilo, predstavljeno s strani aplikacije same – s strani sistema: »Pripravljen sem za delo naprej«. »Konec« je sporočilo, predstavljeno z vidika uporabnika aplikacije: »Obdelava je končala, lahko nadaljuješ«.

Predstavitev podatkov se mora osredotočiti na potrebe intelektualnih delavcev in njihovega vedenja o informaciji, da lahko hitro in enostavno razumejo smisel in pomembnost informacije, in jih pravilno uporabljajo pri svojem delu. Ker se informacija uporablja za različne namene, bo imela različne predstavitvene formate. Kakovost predstavitve pomeni, da je predstavljena oblika intuitivna za uporabo informacij.

3 PRIMERJAVA METODOLOGIJ

V literaturi najdemo široko paleto tehnik za ocenjevanje in izboljšanje kakovosti podatkov. Zaradi raznolikosti in kompleksnosti teh tehnik se bom v tem poglavju osredotočil le na primerjavo metodologij, prilagajanju in uporabi ocenjevanja kakovosti podatkov in tehnik za izboljšanje.

3.1 Primerjalna perspektiva za analiziranje metodologij

Obstaja več perspektiv, ki jih lahko uporabljamo za analiziranje in primerjavo metodologij kakovosti podatkov (Batini & Cappiello, 2009, str. 3):

1. **dimenzije in mere**, izbrane v skladu z metodologijo za ocenjevanje ravni kakovosti podatkov;
2. **strategije in tehnike**, ki so sprejete v metodologijo za ocenjevanje in izboljšanje ravni kakovosti podatkov;
3. **faze in koraki**, ki sestavljajo metodologijo;
4. **vrste stroškov**, ki so povezane s kakovostjo podatkov, vključno z vprašanji:
 - a) stroški povezani s slabo kakovosti podatkov - strošek procesa, ki je nastal iz napačnih podatkov in oportunitetni stroški zaradi izgubljenih in zgrešenih prihodkov; ti stroški so imenovani tudi **posredni stroški**;
 - b) stroški ocenjevanja in izboljšanja aktivnosti, navedeni kot **direktni stroški**;
5. **vrste podatkov**, obravnavane v metodologiji;
6. **vrste informacijskih sistemov**, ki uporabljajo, spreminjajo in upravljajo s podatki, obravnavani v metodologiji.

3.2 Dimenzije in mere

Tabela 1 prikazuje seznam metodologij s kraticami in razširjenim nazivom metodologije ter glavno referenco. V nadaljnjem tekstu bom uporabljal akronime za identifikacijo metodologij. Stroški, dimenzije in faze predstavljajo najbolj diskriminacijska merila, ki vodijo do identifikacije štirih vrst metodologij, ki jih obravnavam pozneje in poudarjam temeljne razlike med njimi.

V nalogi se osredotočam predvsem na kakovost dimenzij in mer prenesenih na vrednosti podatkov. Literatura za kakovost podatkov zagotavlja temeljito razvrstitev dimenzij kakovosti podatkov, vendar pa obstajajo številne razlike v opredelitvi večine dimenzij zaradi vsebinske narave kakovosti. Šest najpomembnejših klasifikacij dimenzije kakovosti so podali Wand in Wang (1996), Wang in Strong (1996), Redman (1996), Jarke, Lenzerini, Vassiliou in Vassiliadis (1995), Bovee, Srivastava in Mak (2001), in Naumann (2002). Z analizo teh klasifikacij lahko določimo osnovni sklop dimenzij kakovosti podatkov,

vključno z natančnostjo, popolnostjo, doslednostjo in pravočasnostjo, ki so v centru predstavljanja večine avtorjev (Catarci & Scannapieco, 2002).

Tabela 1: Metodologije kakovosti podatkov

Metodologija Akronim	Polni naziv metodologije	Glavna referenca
TDQM	Total Data Quality Management	Wang 1998
DWQ	The Datawarehouse Quality Methodology	Jeusfeld et al. 1998
TIQM	Total Information Quality Management	English 1999
AIMQ	A methodology for information quality assessment	Lee et al. 2002
CIHI	Canadian Institute for Health Information methodology	Long & Seko 2005
DQA	Data Quality Assessment	Pipino et al. 2002
IQM	Information Quality Measurement	Eppler & Munzenmaier 2002
ISTAT	ISTAT methodology	Falorsi et al 2003
AMEQ	Activity-based Measuring and Evaluating of product information Quality (AMEQ) methodology	Su & Jin 2004
COLDQ	Loshin methodology (Cost-effect Of Low Data Quality)	Loshin 2004
DaQuinCIS	Quality in Cooperative Information Systems	Scannapieco et al. 2004
QAFD	Methodology for the Quality Assessment of Financial Data	De Amicis & Batini 2004
CDQ	Comprehensive methodology for Data Qualitymanagement	Batini & Scannapieco 2006

Vir: C. Batini, C. Cappiello, C. Francalanci in A. Maurino, Methodologies for data quality assessment and improvement, 2009, str. 16:12.

V metodologijah lahko opazimo veliko različno definiranih dimenzij, kar prikazuje Tabela 2. To potrjuje zapletenost koncepta kakovosti podatkov, kar pa ni presenetljivo, saj lahko danes veliko število pojavov opišemo v smislu podatkov.

Tabela 2: Metodologije in dimenzije kakovosti podatkov

Akronim	Dimenzije kakovosti podatkov
TDQM	Dostopnost, Primernost, Verodostojnost, Celovitost, Jedrnat in dosleden prikaz, Enostavnost manipulacije, Dodana vrednost, Prosto napak, Interpretativnost, Objektivnost, Ustreznost, Ugled, Varnost, Pravočasnost, Razumljivost.
DWQ	Pravilnost, Celovitost, Minimalnost, Sledljivost, Interpretativnost, Razvoj metapodatkov, Dostopnost (sistema, transakciji, varnost), Koristnost (Interpretativnost), Pravočasnost (Aktualnost, Nihajnost), Odzivnost, Verodostojnost, Natančnost, Doslednost.

se nadaljuje

nadaljevanje

Akronim	Dimenzije kakovosti podatkov
TIQM	Naravne (inherentne) dimenzije: Opredelitev skladnosti (doslednosti), Celovitost, Skladnost poslovnih pravil, Točnost (do nadomestnega vira), Natančnost (za realnost), Preciznost, Nepodvajanje, Enakovrednost redundantnih podatkov, Vzpostojnost (hkratnost) redundantnih podatkov. Pragmatične dimenzije: Dostopnost, Pravočasnost, kontekstualna jasnost, Poreklo celovitosti, Uporabnost, Pravilnost (popolnost dejstev), Stroški.
AIMQ	Dostopnost, Primernost, Verodostojnost, Celovitost, Jedrnat in dosleden prikaz, Enostavnost manipulacije, Prosto napak, Interpretativnost, Objektivnost, Ustreznost, Ugled, Varnost, Pravočasnost, Razumljivost.
CIHI	Dimenzije: Natančnost, Pravočasnost, Primerljivosti, Uporabnost, Ustreznost. Značilnost: Preveč obravnavano, Premalo obravnavano, Enostavna/povezana varianca odgovora, Zanesljivost, Zbiranje in zajem, Enota/Artikel brez odgovora, Obdelava, Ocena, Pravočasnost, Razumljivosti, Integracija, Standardizacija, Enakovrednost, Sposobnost povezovanja, Primerljivost izdelkov/zgodovinska, Dostopnost, Dokumentacija, Interpretativnost, Prilagodljivost, Vrednost.
DQA	Dostopnost, Primerna količina podatkov, Verodostojnost, Celovitost, Brez napak, Doslednost, Jedrnat prikaz, Ustreznost, Enostavnost manipulacije, Interpretativnost, Objektivnost, Ugled, Varnost, Pravočasnost, Razumljivosti, Dodana vrednost.
IQM	Dostopnost, Doslednost, Pravočasnost, Jedrnatost, Vzdrževalnost, Aktualnost, Uporabnost, Udobnost, Hitrost, Razumljivost, Jasnost, Točnost, Sledljivost, Varnost, Pravilnost, Interaktivnost.
AMEQ	Dosleden prikaz, Interpretativnost, Primer dogovora, Jedrnat prikaz, Pravočasnost, Celovitost vrednosti, Ustreznost, Primernost, Smislenost, Pomanjkanje zmede, Sporazum, Berljiv, Razumljivost, Natančnost, Zanesljivost, Nepistranskost,
ISTAT	Točnost, Celovitost, Doslednost.
COLDQ	Shema: Jasnost opredelitve, Razumljivost, Prožnost (Fleksibilnost), Robustnost, Nujnost, Razdrobljenost atributov, Preciznost domene, Homogenost, Razpoznavnost, Dosegljivost, Ustreznost, Preprostost/Kompleksnost, Semantična doslednost, Sintaktična doslednost. Podatki: Točnost, Null Vrednosti, Celovitost, Doslednost, Aktualnost, Pravočasnost, Sporazum uporabe, Upraviteljstvo – odgovornost upravljanja (angl. <i>Stewardship</i>), Vseprisotnost. Predstavitev: Primernost, Pravilna interpretacija, Fleksibilnost, Natančnost oblike, Prenosljivost, Doslednost, Skladiščenje. Informacijska politika: Dostopnost, Metapodatki, Zasebnost, Varnost, Redundanca, Stroški.
DaQuinCIS	Točnost, Celovitost, Doslednost, Aktualnost, Zanesljivost (Verodostojnost).
QAFD	Sintaktična in semantična točnost, Notranja in zunanja doslednost, Celovitost, Aktualnost, Edinstvenost.
CDQ	Shema: Pravilnost z vzorcem, Pravilnost glede na zahteve, Celovitost, Primernost, Berljivost, Normalizacija. Podatki: Sintaktična točnost, Semantična točnost, Celovitost, Doslednost, Aktualnost, Pravočasnost, Nihajnost, Popolnost, Ugled, Dostopnost, Stroški.

Vir: C. Batini, C. Cappiello, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:18.

TIQM razvrsti dimenzije kot naravne in pragmatične ali stvarne, metodologija Cost-effect Of Low Data Quality, v nadaljevanju COLDQ, razlikuje med shemo, podatki, predstavitvijo in dimenzijo informacijske politike, metodologija Canadian Institute for Health Information methodology, v nadaljevanju CIHI, ponuja dve stopnji razvrstitve v smislu dimenzij in z njimi povezane značilnosti, metodologija Comprehensive methodology for Data Qualitymanagement, v nadaljevanju CDQ, predlaga sheme in dimenzije podatkov.

Tabela 3 prikazuje mere dimenzij kakovosti različnih metodologij. Ne vključuje mer za semantično natančnost, saj metodologije, ki jih obravnavajo QAFD in CDQ, ne določajo posebnih merilnih metod. Na splošno velja, da je opredeljeno več mer za vsako dimenzijo. Ustrezno temu je v Tabelo 3 vsaka dimenzija vpisana večkrat. Upoštevajmo, da so bile subjektivne mere, kot raziskave uporabnikov – vprašalniki, opredeljene za skoraj vse dimenzije kakovosti.

Tabela 3: Dimenzije in mere

Dimenzija	Naziv	Opredelitev metrike
Točnost (angl. <i>Accuracy</i>)	Acc1	Sintaktična točnost: meri se kot razdalja med vrednostjo, shranjeno v zbirki podatkov in pravilno vrednostjo Sintaktična točnost = število pravilnih vrednosti/število skupnih vrednosti
	Acc2	Število točno dostavljenih množic
	Acc3	Raziskava uporabnika – Vprašalnik
Popolnost (angl. <i>Completeness</i>)	Comp11	Popolnost = Število not null »vrednosti« / skupno število vrednosti
	Comp12	Popolnost = Število dostavljenih množic / Pričakovano število
	Comp13	Popolnost Web podatkov = $(T_{max} - T_{current}) * (PopolnostMax - PopolnostCurrent) / 2$
	Comp14	Raziskava uporabnika – Vprašalnik
Doslednost (angl. <i>Consistency</i>)	Cons1	Doslednost = Število doslednih vrednosti / skupno število vrednosti
	Cons2	Število množic, ki krše omejitve, število sprememb kodiranja
	Cons3	Število strani z odklonom vodnika sloga
	Cons4	Raziskava uporabnika – Vprašalnik
Pravočasnost (angl. <i>Timeliness</i>)	Time1	Pravočasnost = $(\max(0; 1 - Aktualnost / Nihajnost))$
	Time2	Odstotek izvršitve procesov, ki so se sposobni izvesti v zahtevanem časovnem obdobju
	Time3	Raziskava uporabnika – Vprašalnik

se nadaljuje

nadaljevanje

Dimenzija	Naziv	Opredelitev metrike
Aktualnost (angl. <i>Currency</i>)	Curr1	Aktualnost = Čas, v katerem so podatki shranjeni v sistem – čas, v katerem se podatki posodobijo v realnem svetu
	Curr2	Čas zadnje posodobitve
	Curr3	Aktualnost = Čas zahtevka – zadnja posodobitev
	Curr4	Aktualnost = Starost + (Čas dostave – Čas vnosa)
	Curr5	Raziskava uporabnika – Vprašalnik
Nihajnost (angl. <i>Volatility</i>)	Vol1	Dolžina časa, v katerem ostanejo podatki veljavni
Edinstvenost (angl. <i>Uniqueness</i>)	Uni1	Število dvojnikov
Ustrezna količina podatkov (angl. <i>Appropriate amount of Data</i>)	Appr1	Ustrezna količina podatkov = Min ((Število zagotovljenih podatkovnih enot / Število potrebnih podatkovnih enot); (Število potrebnih podatkovnih enot / Število zagotovljenih podatkovnih enot))
	Appr2	Raziskava uporabnika – Vprašalnik
Dostopnost (angl. <i>Accessibility</i>)	Access1	Dostopnost = max (0; 1– (Čas dostave – Čas zahteve) / (Skrajnji rok – Čas zahteve))
	Access2	Število neveljavnih linkov (pretrganih povezav) – Število lomljenih sider
	Access3	Raziskava uporabnika – Vprašalnik
Verodostojnost (angl. <i>Credibility</i>)	Cred1	Število množic z default vrednostjo
	Cred2	Raziskava uporabnika – Vprašalnik
Interpretativnost (angl. <i>Interpretability</i>)	Inter1	Število množic z interpretativnimi podatki, dokumentacija ključnih vrednosti (vrednosti ključa)
	Inter2	Raziskava uporabnika – Vprašalnik
Uporabnost (angl. <i>Usability</i>)	Usa1	Raziskava uporabnika – Vprašalnik
Izpeljana integriteta (angl. <i>Derivation Integrity</i>)	Integr1	Odstotek pravih izračunov pridobljenih podatkov, ki so v skladu z izpeljano formulo ali definicijo izračuna
Jedratost (angl. <i>Conciseness</i>)	Conc1	Število globin (zelo hierarhične) strani
	Conc1	Raziskava uporabnika – Vprašalnik
Vzdrževalnost (angl. <i>Maintainability</i>)	Main1	Število strani z manjkajočimi meta informacijami
Uporabnost (angl. <i>Applicability</i>)	App1	Število osirotelih strani
	App2	Raziskava uporabnika – Vprašalnik
Udobnost (angl. <i>Convenience</i>)	Conv1	Število navigacijskih poti: število izgubljenih / prekinjenih navigacijskih sledi
Hitrost (angl. <i>Speed</i>)	Speed1	Odzivni čas strežnika in omrežja
Izčrpnost (angl. <i>Comprehensiveness</i>)	Comp1	Raziskava uporabnika – Vprašalnik

se nadaljuje

nadaljevanje

Dimenzija	Naziv	Opredelitev metrike
Jasnost (angl. <i>Clarity</i>)	Clar1	Raziskava uporabnika – Vprašalnik
Sledljivost (angl. <i>Traceability</i>)	Trac1	Število strani brez avtorja ali vira
Varnost (angl. <i>Security</i>)	Sec1	Število slabih prijav
	Sec2	Raziskava uporabnika – Vprašalnik
Pravilnost (angl. <i>Correctness</i>)	Corr1	Raziskava uporabnika – Vprašalnik
Objektivnost (angl. <i>Objectivity</i>)	Obj1	Raziskava uporabnika – Vprašalnik
Ustreznost (angl. <i>Relevancy</i>)	Rel1	Raziskava uporabnika – Vprašalnik
Ugled (angl. <i>Reputation</i>)	Rep1	Raziskava uporabnika – Vprašalnik
Enostavnost delovanja (angl. <i>Ease of operation</i>)	Ease1	Raziskava uporabnika – Vprašalnik
Interaktivnost (angl. <i>Interactivity</i>)	Interact1	Število obrazcev – Število personaliziranih strani

*Vir: C. Batini, C. Cappiello, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:19.*

Zadnji stolpec Tabele 4 določa za vsako dimenzijo ter vsako mero, povezano z dimenzijo:

1. število metodologij, ki uporabljajo mero in
2. skupno število metodologij, ki omenjajo ustrezno dimenzijo.

Razmerje med temi vrednostmi meri stopnjo soglasja o merah dimenzij med metodologijami. Takšno soglasje je visoko za točnost, popolnost in doslednost, medtem ko je občutno nižje za dve časovno vezani dimenziji, pravočasnost in aktualnost in za skoraj vse druge dimenzije.

Večina mer, ki so v Tabeli 4 navedene samo enkrat, so del metodologije Information Quality Measurement, v nadaljevanju IQM, ki analizira kakovost spletnih informacij. Take mere so opredeljene z upoštevanjem orodij za merjenje, ki so na voljo za specifičen web kontekst. Na primer z uporabo analizatorja spletnega mesta je mogoče oceniti dimenzije kot so dostopnost, doslednost, pravočasnost, jedrnatost in vzdrževanje. Analizatorji prometa se lahko uporabijo za oceno uporabnosti in udobja, medtem ko so skenerji portov koristni za oceno varnosti.

Tabela 4: Metodologije in mere kakovosti

	TDQM	DWQ	TIQM	AIMQ	CIHI	DQA	IQM	ISTAT	AMEQ	COLDQ	DaQuinCIS	QAFD	CDQ	#metr/#dim
Acc1	X		X			X		X	X	X	X	X	X	9/13
Acc2		X												1/13
Acc3				X	X		X							2/13
Compl1	X			X		X			X	X	X	X	x	7/12
Compl2		X	X											2/12
Compl3													X	1/12
Compl4			X											1/12
Cons1	X					X								6/10
Cons2		X												1/10
Cons3							X							1/10
Cons4				X	X									2/10
Time1						X				X			X	3/7
Time2			X							X				2/7
Time3				X	X									2/7
Curr1								X			X			2/8
Curr2	X						X					X		2/8
Curr3				X	X									1/8
Curr4													X	1/8
Curr5				X	X									2/8
Voll	X												X	2/2
Uni1			X									X		1/2
Appr1						X								1/2
Appr2				X										1/2
Access1							X							1/4
Access2								X						1/4
Access3			X	X										2/4
Cred1		X												1/2
Cred2				X										1/2
Inter1		X												1/2
Inter2				X										1/2
Usa1			X											1/1
Integr1			X											1/1
Conc1														1/2
Conc2							X							1/2
Main1				X										1/1
App1														1/1

se nadaljuje

nadaljevanje

	TDQM	DWQ	TIQM	AIMQ	CIHI	DQA	IQM	ISTAT	AMEQ	COLDQ	DaQuinCIS	QAFD	CDQ	#metr/#dim
App2							X							1/1
Conv1							X							1/1
Speed1							X							1/1
Comp1				X			X			X				3/3
Clar1			X				X			X				3/3
Trac1							X							1/1
Sec1							X							1/1
Sec2				X										1/1
Corr1							X							1/1
Obj1				X										1/1
Rel1				X										1/1
Rep1				X										1/1
Ease1				X										1/1
Interact1							X							1/1

*Vir: C. Batini, C. Cappiello, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:20.*

AIMQ ima več posebnih mer in dimenzij. To je posledica pristopa od zgoraj navzdol, sprejeta v AIMQ pri opredelitvi dimenzij in mer, ki uporablja dve različni razvrstitvi:

1. izdelek v primerjavi s kakovostjo storitev in
2. skladnost s specifikacijo v primerjavi z doseganjem ali preseganjem pričakovanja strank, kar je povzročilo široko razpršeni niz povezanih dimenzij/mer.

Upoštevati je potrebno, da metodologija AIMQ uporablja samo subjektivne metrike za ocenjevanje dimenzij kakovosti. V AIMQ se kakovost podatkov ocenjuje predvsem s pomočjo vprašalnikov, ki vključujejo za vsako dimenzijo kakovosti 4–6 neodvisnih elementov. Elementi imajo naslednjo splošno strukturo: »Ta informacija je (atribut ali fraza)«. Na primer popolnost je povezana s šestimi elementi, vključno z:

1. informacije vsebujejo vse potrebne vrednosti,
2. informacije so dovolj popolne za naše potrebe,
3. informacija izpolnjuje potrebe naših nalog.

Mere, ki jih zagotavlja metodologija DWQ, so tesno povezane s procesi, ki uporabljajo podatke. Vse dimenzije so ovrednotene vzdolž pričakovanja uporabnikov določenega procesa.

3.2.1 Točnost

Wang in Strong (1996) opredeljujeta natančnost kot »v kolikšni meri so podatki točni, zanesljivi in potrjeni«. Ballou in Pazer (1985) določata, da so podatki točni, če vrednosti shranjenih podatkov ustrezajo dejanskim vrednostim v svetu. Po Redmanu (1996) je natančnost določena kot mera bližine vrednosti podatka v do nekatere druge vrednosti v' , ki se šteje kot pravilna. Na splošno je mogoče razlikovati med dvema vrstama natančnosti: sintaktično in semantično. Metodologije kakovosti podatkov upoštevajo samo sintaktične pravilnosti in jih opredelijo kot približek k vrednosti v z elementi, ki ustrezajo opredelitvi domene D (domena ali področje). V sinktatični natančnosti nas ne zanima primerjava v z resnično vrednostjo v' ; zanima nas preverjanje, ali je v katerakoli od vrednosti v domeni D , ali kako blizu je vrednostim v domeni D . Na primer $v = \text{»Jelena«}$ je, obravnavano sintaktično, točno tudi če je $v' = \text{»Helena«}$.

3.2.2 Popolnost

Popolnost je opredeljena kot stopnja, do katere dana podatkovna zbirka podatkov opisuje ustrezen sklop objektov v realnem svetu.

Tabela 5 nam prikazuje prispevke rezultatov raziskav, ki nam pomagajo pri definiciji popolnosti. S primerjanjem pojmov je mogoče ugotoviti, da obstaja precejšen sporazum o abstraktni definiciji popolnosti. Definicije se razlikujejo v kontekstu, na katerega se nanašajo, na primer informacijski sistem po Wand in Wang (1996), podatkovno skladišče po Jarke et al. (1995), entiteta po Bovee et al. (2001).

Tabela 5: Opredelitev dimenzije popolnosti

Referenca	Opredelitev
Wand and Wang 1996	Sposobnost informacijskega sistema, da predstavlja vsako smiselno stanje sistema realnega sveta.
Wang and Wand 1996	Obseg podatkov z dovoljno širino, globino in področjem uporabe za nalogo.
Redman 1996	Stopnja, do katere so vrednosti vključene v zbiranje podatkov
Jarke et al. 1995	Odstotek dejanskih podatkov, vnešenih v vire podatkov in/ali podatkovno skladišče.
Bovee et al. 2001	Informacije, ki imajo vse zahtevane dele opisa entitete.
Naumann 2002	Razmerje med številom vrednosti, ki niso NULL, v viru in velikosti univerzalne relacije.
Liu and Chi 2002	Vse vrednosti, ki naj bi se zbirale, kot definira teorija zbiranja.

Vir: C. Batini in M. Scannapieca, Data quality: concepts, methodologies and techniques, 2006, str. 41

Na raziskovalnem področju relacijskih baz podatkov je popolnost pogosto povezana z smislom null vrednosti. Null vrednost je splošnega pomena za manjkajoče vrednosti – vrednost, ki obstaja v realnem svetu, vendar ni na voljo pri zbiranju podatkov. Da bi karakterizirali popolnost, je pomembno razumeti, zakaj vrednost manjka. Vrednost lahko manjka bodisi zato, ker obstaja, vendar ni znana, ali ker ne obstaja ali ker ni znano da obstaja.

3.2.3 Doslednost

Dimenzija usklajenosti se nanaša na kršitev semantičnih pravil opredeljenih na določenih vrstah podatkov. S sklicevanjem na relacijsko teorijo, so integritetne omejitve vrsta teh semantičnih pravil. V relacijski teoriji poznamo dve temeljni kategoriji integritetne omejitve in sicer: intra-relacijske omejitve in inter-relacijske omejitve. Intra-relacijska omejitve opredeli obseg dopustne vrednosti za atribut domene, inter-relacijska omejitve pa vključuje attribute iz različnih relacij.

3.2.4 Časovne dimenzije: Aktualnost, Volatilnost ali Nihajnost in Pravočasnost

Pomemben vidik podatkov je njihova posodobitev v časovnem obdobju. Glavne časovne dimenzije, predlagane v literaturi, so aktualnost, volatilnost (nihajnost) in pravočasnost. Tabela 6 primerja opredelitve iz literature za te tri dimenzije časa.

Tabela 6: Opredelitev časovnih dimenzij

Referenca	Opredelitev
Wang and Wang 1996	Pravočasnost se nanaša le na zamudo med spremembo stanja v realnem svetu in posledične spremembe stanja v informacijskem sistemu.
Wang and Wang 1996	Pravočasnost je v obsegu, ki je primerna za nalogo pri roki, ob upoštevanju starosti podatkov.
Redman 1996	Aktualnost je stopnja, do katere je podatek aktualen. Vrednost podatka je aktualna, če je ta pravilen kljub možnim razlikam, ki jih povzročajo časovno povezane spremembe njegove vrednosti.
Jarke et al. 1995	Aktualnost opisuje, kdaj so podatki vnešeni v vir in/ali podatkovno skladišče. Volatilnost ali nihajnost opisuje časovno obdobje za katero je informacija veljavna v realnem svetu.
Bovee et al. 2001	Pravočasnost ima dve komponenti: starost in nihajnost. Starost ali aktualnost je mera za starost informacije in temelji na tem, kako dolgo nazaj je bila zabeležena. Nihajnost je mera nestabilnosti informacije, pogostost sprememb vrednost atributa entitete.
Naumann 2002	Pravočasnost je povprečna starost podatkov v viru podatkov
Liu and Chi 2002	Pravočasnost je obseg, v katerem so podatki dovolj aktualni za opravilo.

Vir: C. Batini in M. Scannapieca, Data quality: concepts, methodologies and techniques, 2006, str. 41

Wang in Wang ter Redman (1996) navajajo zelo podobne opredelitve za pravočasnost in aktualnost. Wang in Strong ter Liu in Chi uporabljata za pravočasnost enak pomen, medtem ko Bovee predvideva opredelitev pravočasnosti v aktualnosti in volatilitnosti. Opredelitev aktualnosti, po Boveeju (2001), ustreza pravočasnosti, kot jo opredeljujejo Wang in Strong ter Liu in Chi. Ta primerjava kaže, da ni enotnega mnenja o abstraktni opredelitvi časovne dimenzije, običajno aktualnost in pravočasnost se pogosto uporablja za isti koncept.

3.3 Strategije in tehnike

V korakih izboljšanja vključujejo metodologije dve splošni vrsti strategij in sicer tako imenovano **podatkovno usmerjeno** in **procesno usmerjeno strategijo**. Podatkovno usmerjena strategija za izboljšanje kakovosti podatkov neposredno spreminja vrednost podatkov. Na primer zastareli podatki se osvežujejo iz zbirke podatkov z novejšimi podatki. Procesno usmerjena strategija izboljšuje kakovost s preoblikovanjem procesov, ki ustvarjajo ali spreminjajo podatke. Na primer proces lahko preoblikujemo z vključitvijo aktivnosti, ki nadzoruje format podatka pred shranjevanjem.

Strategije, tako podatkovno kot procesno usmerjene, uporabljajo različne tehnike: algoritme, hevrstike, aktivnosti temeljene na znanju, katerih cilj je izboljšati kakovost podatkov. Tehnike za izboljšanje, ki jih uporabljajo **podatkovno usmerjene strategije**, so (Batini & Cappiello, 2009, str. 5):

- **pridobivanje novih podatkov** izboljšuje obstoječe podatke s pridobivanjem kakovostnejših podatkov z nadomestitvijo vrednosti, zaradi katerih so nastale težave v kakovosti podatkov;
- **standardizacija (ali normalizacija)** nadomešča ali dopolnjuje nestandardne vrednosti podatkov z ustreznimi vrednostimi, ki ustrezajo standardom. Na primer vzdevek nadomestijo z ustreznimi imeni, na primer Megi z Magdalena, oziroma se okrajšave nadomestijo z ustreznimi polnimi imeni, na primer Trebinjska ul. s Trebinjska ulica;
- **povezovanje zapisov** (record linkage) definira prezentacijo podatkov v dveh (ali več) tabelah, ki se nanašajo na isti objekt v realnem svetu;
- **integracija podatkov in shem** opredeljuje enoten pogled na podatke, ki so pridobljeni iz heterogenih virov podatkov. Glavni namen integracije je omogočiti uporabnikom dostop do podatkov, shranjenih v različnih heterogenih virih podatkov, preko enotnega pogleda na te podatke. Značilno za vire podatkov je različna vrsta heterogenosti, ki se lahko na splošno razvrsti v:
 1. tehnološko heterogenost,
 2. schema heterogenost in
 3. heterogenost na ravni instance.

Tehnološka heterogenost se pojavi zaradi uporabe proizvodov različnih proizvajalcev, ki delajo na različnih nivojih informacijske in komunikacijske infrastrukture. Schema heterogenost nastopi predvsem zaradi uporabe

1. različnih modelov podatkov, kot na primer vir, ki sprejme relacijski model podatkov in drugi vir, ki sprejema model podatkov XML, in
2. različne prezentacije za isti cilj.

Heterogenost na ravni instance je provzročena z različnimi konfliktnimi vrednostmi podatkov, nujenih iz različnih virov za isti objekt. Ta tip heterogenosti lahko povzročijo na primer neodvisni in slabo usklajeni procesi, ki dovajajo podatke v različne vire podatkov. Integracija podatkov se sooča z vsemi vrstami naštetih heterogenosti;

- **zanesljivost vira** izbere vire podatkov na podlagi kakovosti njihovih podatkov;
- **lokalizacija napak** in popravki identificirajo in odpravijo napake kakovosti podatkov z odkrivanjem zapisov, ki ne izpolnjujejo dani set kakovostnih pravil. Te tehnike se v glavnem preučujejo na statističnem področju;
- **optimizacija stroškov** določa ukrepe za izboljšanje kakovosti vzdolž določenih dimenzij z zmanjševanjem stroškov.

Dve glavni tehniki, karakteristični za **procesno usmerjene strategije**, sta (Batini & Cappiello, 2009, str. 5):

- **proces nadzora**, ki vnaša postopke preverjanja in nadzora, kadar:
 1. se kreirajo novi podatki,
 2. se podatki posodablajo ali
 3. proces dostopa do novega niza podatkov.

Na ta način se reaktivna strategija uporablja pri dogodkih spremembe podatkov, s čimer se izognemo degradaciji podatkov in razmnoževanju napak;

- **proces preoblikovanja**, ki preoblikuje procese, da bi odstranili vzrok slabe kakovosti in uvedli nove dejavnosti, ki proizvajajo podatke višje kakovosti. Če je proces preoblikovanja radikalen, je ta tehnika kot reinženiring poslovnih procesov (Hammer & Champy, 2001).

Na splošno dolgoročno gledano, procesno usmerjene tehnike prekašajo podatkovno usmerjene tehnike, saj odpravljajo korenski vzrok težav s kakovostjo. Vendar je lahko, gledano kratkoročno, proces preoblikovanja izjemno drag. Podatkovno usmerjene strategije so kratkoročno stroškovno učinkovite, vendar drage gledane dolgoročno. Primerne so za enkratno uporabo in so zato priporočljive za statične podatke.

Tabela 7 prikazuje strategije in tehnike iz različnih metodologij. Metodologija je povezana s strategijo, če ona določa smernice za izbiro in oblikovanje ustrezne tehnike. Stolpec z oznako Procesno usmerjene v zagotavlja enake informacije kot stolpec Kontrola procesa in Preoblikovanje procesa v Tabeli 10. Stolpec označen Podatkovno usmerjene izrecno navaja podatkovno usmerjene tehnike, ki so implicitno obravnavane v naslednjih poglavjih.

Tabela 7 kaže, da pet metodologij kakovosti podatkov sprejema mešane strategije, ki združujejo različne podatkovno in procesno usmerjene tehnike. Metodologija, ki uporablja

širši obseg podatkovno in procesno usmerjenih tehnik, je TIQM. Nasprotno TDQM zagotavlja smernice za uporabo procesno usmerje strategije s pomočjo Information Manufacturing Analysis Matrix (Ballou, 1998), kar nakazuje kdaj in kako izboljšati podatke.

Tabela 7: Metodologije in vrste strategij

Strategija Metod. Akronim	Podatkovno usmerjene	Procesno usmerjene
TDQM		Preoblikovanje procesa
DWQ	Integracija podatkov in shem	
TIQM	Čiščenje podatkov Normalizacija Lokalizacija in korekcija napak	Preoblikovanje procesa
ISTAT	Normalizacija Povezovanje zapisov	Preoblikovanje procesa
COLDQ	Optimizacija stroškov	Kontrola procesa Preoblikovanje procesa
DaQuinCIS	Zanesljivost vira Povezovanje zapisov	
CDQ	Normalizacija Povezovanje zapisov Integracija podatkov in shem Lokalizacija in korekcija napak	Kontrola procesa Preoblikovanje procesa

*Vir: C. Batini, C. Cappiello, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:14.*

Treba je omeniti, da metodologija, ki sprejema izključno bodisi podatkovno (DWQ in metodologija Quality in Cooperative Information Systems, v nadaljevanju DaQuinCIS) ali procesno usmerjene strategije, ne more biti prilagodljiva (fleksibilna) za združbe, ki že imajo vpeljan proces upravljanja s kakovostjo podatkov. Edina metodologija, ki eksplicitno obravnava to vprašanje, je CDQ, ki zbere skupaj podatkovno in procesno usmerjene tehnike. Izbira najbolj primerne strategije in tehnike temelji na domensko (vsebinsko) odvisnih odločitvenih spremenljivkah (Batini, 2006).

Normalizacija, povezovanje zapisov, integracija podatkov in shem predstavljajo podatkovno usmerjene tehnike, ki so najbolj sprejete v metodologijah kakovosti podatkov, medtem ko je preoblikovanje procesov najbolj pomembna v procesno usmerjenih metodologijah.

3.3.1 Podatkovno usmerjene tehnike

Tehnike normalizacije so bile predlagane v več področjih. Tako ISTAT in CDQ zagotavljata tehnike normalizacije za izboljšanje kakovosti podatkov s primerjavo podatkov z look-up tabelami in opredelitev skupnih meta shem. Na primer metodologija ISTAT uporablja nacionalni register ulic kot *look-up* tabelo za ozemeljske podatke.

Povezovanje zapisov je bilo raziskano 50-ih v raziskavah o bazah podatkov in je uporabljeno v številnih področjih, kot so zdravstvo, uprava, aplikacije za popis. Na takšnih področjih je ključnega pomena izdelava procedur za učinkovito računalniško podprto ujemanje, ki lahko zmanjša uporabo administrativnih virov, in ob istem času zmanjša napake ujemanja.

CDQ obravnava tri vrste tehnik povezovanja zapisov (Batini & Cappiello, 2009, str. 15):

- verjetnostna tehnika, ki temelji na široki paleti metod, razvitih v preteklih dveh stoletjih v statistiki in teoriji verjetnosti, ki segajo od Bayesian omrežij do podatkovnega rudarjenja;
- empirične tehnike, ki omogočajo uporabo algoritemskih tehnik, kot so sortiranje, drevo, analiza, primerjava sosed in obrezovanje;
- tehnike, ki temeljijo na znanju (angl. *knowledge-based*), pridobivanju znanja iz datotek in uporabi obrazložitvene strategije.

Projekt DaQuinCIS je razvil specifično tehniko povezovanja zapisov (Bertolazzi, 2003). V platformi DaQuinCIS se povezovanje zapisov izvaja v dveh fazah:

1. povezovanje zapisov usklajuje različne kopije istega subjekta iz različnih virov podatkov,
2. povezovanje zapisov podpira tudi fazo obdelave poizvedb z opredelitvijo istih primerov v rezultatih poizvedb, ki jih vrne vsak vir podatkov.

Metoda povezovanja zapisov temelji na metodi »Sorted Neighborhood« (Hernandez & Stolfo 1998), vendar je uvedeno nekaj novih funkcij:

- ključ ujemanja je izbran samodejno,
- algoritem ujemanja temelji na funkciji, ki normalizira klasično urejanje funkcije razdalje na string (niz) dolžine.

Integracija podatkov in shem (Lenzerini, 2002) je široko področje raziskav, ki se delno prekriva s kakovosti podatkov. Podatkovno usmerjene tehnike izboljšanja, uporabljene v metodologijah, pogosto temeljijo na uporabi novih podatkov za izboljšanje kakovosti zbranih podatkov. Kot posledica, se tehnike izboljšanja kakovosti podatkov osredotočajo predvsem na raven stopnje heterogenosti, da bi prepoznali podobne zapise, zaznali sporne vrednosti in izbrali pravilen primerek.

Metodologije, ki obravnavajo raven stopnje heterogenosti so DWQ, ISTAT, DaQuinCIS in CDQ. Pri metodologiji DWQ, so heterogeni informacijski viri dostopni najprej na enoten način, z ekstrakcijski mehanizmi, ki se imenujejo ovoji (angl. *wrappers*). Nato **mediatorji** prevzamejo nalogo povezovanja informacij in reševanja konfliktov. Nastali standardizirani in integrirani podatki so shranjeni kot materializirani pogledi (angl. *materialized view*) v skladišču podatkov.

3.3.2 Procesno usmerjene tehnike

Metodologije, ki obravnavajo korak preoblikovanja procesa, se nagibajo k izposoji ustrezne tehnike iz literature o poslovnem preoblikovanju procesa (v nadaljevanju BPR) (Muthu, 1999), (Hammer, 1993). TDQM predstavlja izjemo, saj predlaga pristop izvirne kontrole preoblikovanja procesa, ki je naveden kot »informacijski proizvodni sistem za informacijske produkte« (Ballou, 1998). Ta metodologija predlaga model Informacijsko Proizvodne Mape (v nadaljevanju IP-MAP) (Shankaranarayan, Wang & Ziad, 2000), ki se uporablja za modeliranje informacijskih produktov, upravljanih s proizvodnimi procesi. Informacijska proizvodna mapa je grafični model, namenjen za pomoč analitikom vizualizirati proces proizvodnje informacij, identifikacijo lastništva nad fazami postopka, razumeti informacijske in organizacijske meje ter oceniti čas in kakovost mer, povezanih s trenutnim procesom proizvodnje.

Opis procesov je obvezna aktivnost, ki je v skladu s splošno usmeritvijo procesno usmerjenih strategij. Po modeliranju in ocenjevanju podatkov proizvodnega procesa, se identificirajo nove aktivnosti nadzora procesa in/ali se sprejmejo odločitve o preoblikovanju procesa.

Zaradi visokih stroškov ni vedno možno sprejeti kompleksne rešitve, kot je IP-MAP. V nekaterih primerih je praktično neizvedljiv korak temeljitega modeliranja procesa, zato druge metodologije sprejemajo manj formalne, vendar bolj izvedljive rešitve. Na primer CDQ temelji na sklopu matrik, ki opisujejo glavna razmerja med podatki, pretoku informacij, procesi in organizacijskimi enotami.

3.4 Faze in koraki

V najbolj splošnem primeru je zaporedje dejavnosti metodologij kakovosti podatkov sestavljeno iz treh faz (Batini & Cappiello, 2009, str. 3):

1. **Faza rekonstrukcije** je namenjena zbiranju vsebinskih informacij o procesih in storitvah v združbi, zbiranju podatkov in z njimi povezanimi postopki upravljanja, vidikih kakovosti in ustreznih stroških; ta faza se lahko preskoči, če so vsebinske informacije na voljo iz predhodnih analiz.

2. **Ocena/merjenje** meri kakovost zbiranja podatkov ob pomembnih dimenzijah kakovosti; izraz merjenje se uporablja za reševanje vprašanja merjenja vrednosti določenih dimenzij kakovosti podatkov. Izraz ocena se uporablja, kadar so takšne meritve primerljive referenčnim vrednostim, da bi omogočili diagnozo kakovosti. Izraz ocena je sprejet v skladu z večino metodologij, ki poudarjajo pomen vzrokov slabe kakovosti podatkov.
3. **Izboljšanje** se nanaša na izbor ukrepov, strategij in tehnik za doseganje novih ciljev kakovosti podatkov.

Faza rekonstrukcije stanja ni obvezna, če lahko faza ocenjevanja temelji na podlagi obstoječe dokumentacije. Ker imajo metodologije ponavadi take predpostavke, nadalje ne bomo razpravljali o fazi rekonstrukcije.

Čeprav pod različnimi imeni, upoštevajo metodologije faze ocenjevanja in faze izboljšanja kot skupni sklop osnovnih korakov. Koraki faze ocenjevanja so (Batini, Cappielo, Francalanci & Maurino, 2009, str. 16:4):

- **analiza podatkov**, ki obravnava sheme podatkov in opravlja intervjuje za doseganje popolnega razumevanja podatkov in s tem povezanih pravil arhitekture in upravljanja;
- **analiza zahtev po kakovosti podatkov**, ki raziskuje mnenja uporabnikov in skrbnikov podatkov za ugotavljanje problematike kakovosti in postavljanja novih ciljev kakovosti;
- **prepoznavanje kritičnih področij**, ki izbere najbolj ustrezne podatkovne baze in toke podatkov, ki jih je treba oceniti kvantitativno;
- **procesa modeliranja**, ki določa model procesov, ki proizvajajo ali posodablajo podatke;
- **merjenje kakovosti**, ki izbere dimenzijo kakovosti, ki vpliva na kakovost vprašanj, opredeljenih v analizi zahtev DQ korakov in opredeljuje ustrezne meritve; meritve so lahko objektivne, ko temeljijo na kvantitativnih meritvah, ali subjektivne, ko bazirajo na podlagi kvalitativnih ocen administratorjev podatkov in uporabnikov.

Upoštevajmo, da igrajo v vseh fazah ocenjevanja metapodatki pomembno vlogo, saj shranjujejo dopolnilne (komplementarne) informacije o podatkih za različne namene, vključno s kakovostjo podatkov. Metapodatki pogosto zagotovijo potrebne informacije za razumevanje podatkov in/ali ovrednotenje.

Koraki faze izboljšanja so (Batini, Cappielo, Francalanci & Maurino, 2009, str. 16:5):

- **ocenjevanje stroškov**, ki ocenjuje neposredne in posredne stroške kakovosti podatkov;
- **dodeljevanje odgovornosti v procesu**, ki identificira lastnike procesov in opredeljuje odgovornosti za aktivnosti kreiranja in upravljanja s podatki;
- **dodeljevanje odgovornosti za podatke**, ki opredeljuje lastnika podatkov in določa njihovo odgovornost za upravljanje s podatki;

- **ugotavljanje vzrokov napak**, ki opredeljujejo vzroke težav s kakovostjo;
- **izbor strategij in tehnik**, ki opredeljuje vse strategije za izboljšanje podatkov in ustrezne tehnike, ki so skladne s kontekstualnim znanjem, cilji kakovosti in proračunskimi omejitvami;
- **oblikovanje rešitev za izboljšanje podatkov**, ki izbere najbolj uspešno in učinkovito strategijo in sorodni nabor tehnik in orodij za izboljšanje kakovosti podatkov;
- **proces nadzora**, ki opredeljuje kontrolno točko v procesih kreiranja (proizvodnje) podatkov, za spremljanje kakovosti tekom izvajanja procesa;
- **proces preoblikovanja**, ki opredeljuje ukrepe za izboljšanje procesa, ki lahko poda ustrezne izboljšave pri kakovosti podatkov;
- **izboljšanje upravljanja**, ki opredeljuje nova organizacijska pravila za kakovost podatkov;
- **izboljšanje nadzora**, ki uvaja redno spremljanje dejavnosti, ki zagotavljajo povratne informacije o rezultatih izboljšanja procesa in omogoča njeno dinamično izboljšavo.

3.4.1 Faza ocenjevanja

Tabela 8 primerja korake, uporabljene v različnih metodologijah, v fazi ocenjevanja. Metodologije se nanašajo na korake, opisane v poglavju 3.4, čeprav z različnimi imeni.

Tabela 8: Metodologije in koraki ocenjevanja

Korak/ Metod. Akronim	Analiza podatkov	Analiza potrebe po kakovosti podatkov	Identifikacija kritičnih področij	Modeliranje procesa	Ocena kakovosti	Razširljivo za ostale dimenzije in mere
TDQM	+		+	+	+	Fiksno
DWQ	+	+	+		+	Odrpno
TIQM	+	+	+	+	+	Fiksno
AIMQ	+		+		+	Fiksno
CIHI	+		+			Fiksno
DQA	+		+		+	Odrpno
IQM	+				+	Odrpno
ISTAT	+				+	Fiksno
AMEQ	+		+	+	+	Odrpno
COLDQ	+	+	+	+	+	Fiksno
DaQuinCIS	+		+	+	+	Odrpno
QAFD	+	+	+		+	Fiksno
CDQ	+	+	+	+	+	Odrpno

*Vir: C. Batini, C. Cappiello, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:12.*

Najpogosteje obravnavani koraki faze ocenjevanja so **analiza podatkov** in **merjenje kakovosti**, vendar se izvajajo v skladu z različnimi pristopi. Na primer korak merjenja kakovosti se opravlja z vprašalniki v metodi AIMQ, s kombinacijo objektivnih in subjektivnih meritev v metodi DQA, ali s statistično analizo v QAFD. Različni pristopi merjenja izpolnjujejo specifične zahteve različnih organizacijskih kontekstov, procesov, uporabnikov ali storitev. Le nekaj metodologij upošteva korak analize DQ zahtevka, za identificiranje DQ vprašanj in zbiranja nove ravni kakovosti od uporabnikov. Ta korak je posebej pomemben za ocenjevanje in reševanje konfliktov v ciljnih stopnjah DQ iz različnih interesnih skupin. Na primer QAFD priporoča zbiranje ciljne vrednosti kakovosti od različnih vrst strokovnjakov, vključno s poslovnimi strokovnjaki in finančnimi subjekti, vendar ne pomaga pri »pobotu« nezdržljivih ravni DQ. Nekaj metodologij podpira modeliranje procesov. Upoštevajmo, da z izjemo AMEQ, metodologije, ki podpirajo modeliranje procesov, prav tako sprejmejo za fazo izboljšanja procesno usmerjeno strategijo.

V zadnjem stolpcu Tabele 8 je določeno, ali omogoča metodologija razširljivost na nove dimenzije (ali mere), razen na tiste, ki jih izrecno obravnava metodologija. Na primer CDQ izrecno omenja dimenzije, vendar lahko pristop posplošimo tudi z drugimi dimenzijami. Nasprotno zagotavlja ISTAT podrobne mere in postopke za izboljšanje natančnosti, popolnosti in doslednosti; posledično je celotni pristop strogo povezan s takšno dimenzijo. Upoštevajte, da metodologije, ki obravnavajo modeliranje procesov kot ocenjevanje korakov kakovosti, temeljijo na pristopu primernosti za uporabo. Kakovost podatkov ocenjujejo tekom procesa, v katerih se uporabljajo, in s tem opravljajo predvsem subjektivne mere.

3.4.2 Faza izboljšanja

Tabeli 9 in 10 primerjata korake izboljšanja različnih metodologij. Ugotavljanje vzrokov napak je najbolj obravnavan korak izboljšanja. DQA poudarja pomen koraka ugotavljanja vzrokov napak, vendar ne obravnava njegovega izvrševanja. Podobno se DWQ referencira na matematični model, ki temelji na konceptu odvisnosti za podporo koraku ugotavljanja vzrokov napak, vendar je opredelitev modela predstavljena kot tekoče delo in ni zagotovljena (priskrbljena – model ni dan).

Tabela 9: Metodologije in koraki izboljšanja – korak 1

Korak/ Metod. Akronim	Ocena stroškov	Dodeljevanje procesne odgovornosti	Dodelitev podatkovne odgovornosti	Izbira strategij in tehnik	Identifikacija vzroka napake
TDQM	+	+	+	+	+
DWQ	+		+	+	+

se nadaljuje

nadaljevanje

Korak/ Metod. Akronim	Ocena stroškov	Dodeljevanje procesne odgovornosti	Dodelitev podatkovne odgovornosti	Izbira strategij in tehnik	Identifikacija vzroka napake
TIQM	+	+	+	+	+
DQA					+
ISTAT				+	+
AMEQ					+
COLDQ	+			+	+
DaQuinCIS				+	+
CDQ	+	+	+	+	+

*Vir: C. Batini, C. Cappelletti, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:13.*

Samo šest metodologij obravnava več korakov za izboljšanje, kar potrjuje Tabela 10. Izboljšanje aktivnosti (dejavnosti) temelji večinoma na preoblikovanju procesov, z izjemo metodologije DWQ, ki zagotavlja podaljšanje »Goal Question Metric« (Bazilij, 1994), prvotno predlagane na področju inženiringa programske opreme.

Tabela 10: Metodologije in koraki izboljšanja – korak 2

Korak/ Metod. Akronim	Kontrola procesa	Načrtovanje rešitve za izboljšanje podatkov	Preoblikovanje Procesa	Upravljanje z izboljšanjem	Upravljanje s spremljanjem (nadzorom)
TDQM			+	+	+
DWQ		+		+	
TIQM		+	+		+
DQA					
ISTAT		+	+		
AMEQ					+
COLDQ	+	+	+		+
DaQuinCIS					
CDQ	+	+	+		

*Vir: C. Batini, C. Cappelletti, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:13.*

Korak ocene stroškov je ponavadi obvezen v metodologijah o kakovosti podatkov. Ta korak je odločilnega pomena za merjenje ekonomske prednosti rešitve izboljševanja in izbiro najbolj učinkovite tehnike za izboljšanje. V nasprotju s tem se rešitve za upravljanje s korakom izboljšanja izrecno izvajajo samo pri TDQM. Druge metode se nanašajo na širok razpon tehnik za upravljanje in na primere najboljših praks, ki so na voljo pri področju upravljanja sprememb (Kettinger & Grover, 1995). Poleg tega je mogoče

ponoviti fazo ocenjevanja metodologije kot oceno rezultatov faze izboljšanja. Kot primer, DQA izrecno priporoča uporabo prejšnje metodološke faze za oceno učinkovitosti izboljšav.

Razmerja med kakovostjo podatkov, procesi in združbo obravnavajo metodologije TIQM, TDQM in CDQ, ki temeljito razpravljajo o dodelitvi odgovornosti za procese in podatke. Ti koraki so podprti z rezultati faze obnove stanja. CDQ obravnava nabor matrik, ki predstavljajo razmerja med procesi, organizacijskimi enotami in bazami podatkov, ki se proizvajajo med fazo obnove stanja in se nato uporabijo pri koraku dodelitve odgovornosti.

3.5 Stroški

Stroški so pomembna postavka, obravnavana v metodologijah, zaradi učinkov nizke kakovosti podatkov na porabo sredstev. Stroški kakovosti podatkov so vsota **stroškov za ocenjevanje kakovosti podatkov in stroškov dejavnosti izboljšanja podatkov**, imenovani tudi stroški programa kakovosti podatkov in stroški povezani s slabo kakovostjo podatkov. Stroške slabe kakovosti je mogoče zmanjšati z izvajanjem bolj učinkovitih programov kakovosti podatkov, ki so običajno dražji. Zato s povečanjem stroškov za program kakovosti podatkov zmanjšujemo stroške slabe kakovosti podatkov. To znižanje je mogoče obravnavati kot korist programa kakovosti podatkov.

Stroški programa kakovosti podatkov se lahko štejejo kot preventivni stroški, ki so nastali v združbi zaradi zmanjševanja podatkovnih napak. Ta stroškovna kategorija vključuje stroške vseh faz in korakov, ki sestavljajo oceno kakovosti podatkov in postopek izboljšanja.

Stroške slabe kakovosti je mogoče razvrstiti na naslednji način (English, 1999):

- stroški procesa, kot so stroški povezani s ponovno izvedbo celotnega procesa zaradi podatkovnih napak,
- oportunitetni stroški zaradi izgubljenih in zgrešenih prihodkov.

Stroški slabe kakovosti podatkov so močno odvisni od konteksta, v nasprotju s stroški programa kakovosti podatkov. Zaradi tega jih je posebej težko oceniti, ker imata ista vrednost podatka in ustrezna raven kakovosti drugačen vpliv, odvisno od prejemnika (uporabnika podatka). Na primer aktivni trader, ki prejme zastarele podatke o delnicah, lahko povzroči precejšnjo gospodarsko škodo kot posledica napačne naložbene odločitve. V nasprotju časopis, ki mesečno objavlja isto zastarelo poročilo o podatkih trgovanja, ne bo utrpel nobene gospodarske škode.

Dimenzija stroška se upošteva samo v metodologijah TIQM, COLDQ in CDQ. Stroške analiziramo iz dveh različnih vidikov:

1. klasifikacije stroškov in
2. zagotovljenih meril za kvantifikacijo stroškov.

TIQM (English, 1999) in COLDQ (Loshin, 2004) navajata podrobne klasifikacije stroškov, tretjo klasifikacijo pa so ustvarili Eppler in Helfert (2004), kar prikazuje Tabela 11.

Tabela 11: Opredelitev stroškov nekakovostnih podatkov

English	Loshin	Eppler in Helfert
<p>Stroški zaradi neizvedbe procesa Nepovratni stroški Stroški izgube ugleda Stroški ukvarjanja z nezadovoljnimi strankami</p> <p>Stroški popravljanja napak Ukvarjanje s podvojenimi podatki Iskanje manjkajočih podatkov Manjša produktivnost Stroški ponovnega opravljanja dela Stroški preverjanja podatkov Stroški reprogramiranja programske opreme Stroški čiščenja in popravljanja podatkov</p> <p>Izgubljene in zgrešene poslovne priložnosti Stroški izgubljenih poslov Stroški zamujenih poslovnih priložnosti Stroški zmanjšanja vrednosti združbe zaradi predhodnih dejstev</p>	<p>Stroški operativnih procesov Stroški zaznavanja napak Stroški popravljanja napak Stroški ponovnega opravljanja dela Stroški preprečevanja napak Stroški uveljavljanja jamstva Zmanjšanje prihodkov</p> <p>Stroški taktičnih in strateških procesov Stroški zamud Stroški predkupne pravice Stroški nedela Stroški zgrešenih priložnosti Stroški zaradi izgube zaupanja v združbo Stroški vzdrževanja</p>	<p>Neposredni stroški Stroški preverjanja podatkov Stroški vnovičnega vnašanja podatkov Stroški izravnava (kompenzacij)</p> <p>Posredni stroški Stroški izgube ugleda Stroški zaradi napačnih odločitev Stroški izgubljenih vlaganj</p>

Vir: U. Godnov, Vpliv implementacije relacijskega podatkovnega modela na kakovost podatkov v poslovnih informacijskih sistemih, 2009, str. 34.

Po Englishu (1999) ustrezajo stroški kakovosti podatkov stroškom poslovnih procesov in procesov upravljanja s podatki zaradi slabe kakovosti podatkov. Stroški ocenjevanja kakovosti informacij ali inšpekcijska mera dimenzij kakovosti podatkov preverja, ali se procesi izvajajo pravilno. Končno, proces izboljšanja in stroški preprečevanja napak vključujejo dejavnosti za izboljšanje kakovosti podatkov, s ciljem odprave ali zmanjšanja stroškov slabe kakovosti podatkov. Stroški zaradi slabe kakovosti podatkov so analizirani v globino s TIQM in so razdeljeni v tri kategorije:

1. **Stroški neuspešnih procesov** se pojavijo, ko slaba kakovost podatkov povzroča, da se proces ne izvede pravilno. Kot je na primer napačen naslov vzrok, da pošta ne bo dostavljena na pravi naslov.
2. **Ostanki informacij in predelava:** ko so podatki slabe kakovosti, je potrebno vključiti več vrst aktivnosti za upravljanje z napakami, kot so predelava, čiščenje ali zavrnitev.
3. **Stroški izgubljenih in zamujenih priložnosti** ustrezajo izgubi prihodka in dobička zaradi slabe kakovosti podatkov. Na primer zaradi nizke točnosti e-mail naslovov strank, določenega odstotka že pridobljenih strank ni mogoče doseči z rednimi oglaševalskimi kampanjami, kar se odraža v nižjih prihodkih, približno sorazmerno z zmanjšano točnostjo naslovov.

Loshin analizira stroške nizke kakovosti podatkov z razvrščanjem v skupine, glede na njihov vpliv v področja, in sicer:

- **operativno področje**, ki vključuje komponente sistema za obdelavo informacij in stroškov delovanja sistema;
- **taktično področje**, ki poskuša obravnavati in rešiti probleme preden nastanejo;
- **strateško področje**, ki poudarja dolgoročne odločitve.

Uvedeno je več stroškovnih kategorij za operativen in taktično/strateški vpliv. Stroški operativnega vpliva:

- stroški povrnitve v prejšnje stanje nastanejo, ko je treba delo, ki je bilo opravljeno, razveljaviti;
- stroški predelave nastanejo, ko je treba ponoviti procesiranje;
- stroški preprečevanja – ukrepi za preprečevanje operativne okvare zaradi problema s kakovostjo podatkov.

Eppler in Helfert (2004) uporabljata metodo »od spodaj navzgor«. Njuno izhodišče so stroški, ki jih navaja literatura. Te stroške sta razdelila v dve skupini in sicer med neposredne ter posredne stroške nekakovostnih podatkov. Eppler in Helfert ugotavljata, da gre pri zagotavljanju in izboljšanju kakovosti podatkov predvsem za stroške, ki nastanejo zaradi želje po preprečevanju nekakovostnih podatkov, za stroške odkrivanja nekakovostnih podatkov ter za stroške popravljanja, medtem ko se Loshin (2004) in English (1999) usmerjata predvsem na stroške za izboljšanje procesov.

Združbe se za odpravo nekakovostnih podatkov oziroma izboljšavo obstoječe kakovosti podatkov odločajo iz ekonomskih razlogov. Pri določanju ukrepov je treba pretehtati stroške, ki jih bodo ti ukrepi zahtevali, pa tudi koristi, ki jih bodo ukrepi prenesli. Omenjena avtorja English (1999) in Loshin (2004) razvrstita koristi kakovostnih podatkov v tri skupine:

- koristi, ki so neposredno merljive v denarju (na primer boljša kakovost podatkov vpliva na višje prejeme); te koristi lahko poimenujemo tudi **denarne koristi** (angl. *monetizable*);
- **merljive koristi** (angl. *quantifiable*), ki jih sicer ne moremo neposredno izraziti v denarju, lahko pa jih merimo v neki drugi številski merski enoti (na primer boljši podatki davčne uprave omogočajo združbam in posameznikom časovni prihranek, ki ga lahko izmerimo); med merljive koristi spadajo časovno skrajšanje poslovnega ciklusa, večja produktivnost ter večji tržni delež;
- **neotipljive koristi** (angl. *intangible*) so koristi, ki jih ne moremo neposredno izraziti z nobeno številsko mersko enoto (na primer izguba ugleda združbe); neotipljive koristi izboljšanja kakovosti podatkov so večje zadovoljstvo strank, večje zadovoljstvo zaposlenih ter boljša kakovost storitev.

Oba avtorja, English (1999) in Loshin (2004), se strinjata glede denarno ovrednotenih koristi, medtem ko je pri neotipljivih koristih Loshin bolj skromen, English pa bolj obširen, pri čemer poudarja predvsem pomembnost boljše kakovosti storitev.

3.6 Vrste podatkov

Končni cilj metodologij kakovosti podatkov je analiza podatkov, ki na splošno opisujejo realen svet objektov v obliki, da jih je mogoče shraniti, da so dostopni in jih lahko obdeluje software ter so sporočljivi preko omrežja. Na področju kakovosti podatkov večina avtorjev implicitno ali eksplicitno razlikuje tri vrste podatkov (Batini, Cappiello, Francalanci & Maurino, 2009, str. 16:9):

- **strukturirani podatki** so agregacija ali posploševanje predmetov, ki jih opisuje osnovni atributiv domeni. Domena predstavlja razpon vrednosti, ki jim je mogoče dodeliti lastnosti in običajno ustrezajo osnovni vrsti podatkov programskih jezikov, so kot številčne vrednosti ali tekstualni stringi. Relacijske tabele in statistični podatki predstavljajo najpogostejše vrste strukturiranih podatkov;
- **nestrukturirani podatki** so generično zaporedje simbolov, ponavadi kodirane v naravni jezik. Tipičen primer nestrukturiranih podatkov je vprašalnik, ki vsebuje odgovore na vprašanja v prostem – nestrukturiranem formatu ali body e-maila;
- **semistrukturirani (polstrukturirani) podatki** so podatki, ki imajo strukturo z določeno stopnjo prilagodljivosti. Semistrukturirani podatki so prav tako navedeni kot manj shematski ali samoopisujoči.

XML je označevalni (markup) jezik, običajno uporabljan za predstavljanje semistrukturiranih podatkov. Nekaj skupnih značilnosti je:

1. podatki lahko vsebujejo polja, ki niso znana v času načrtovanja, na primer XML datoteka nima povezane XML shemo,
2. iste vrste podatkov so lahko predstavljene na različne načine, na primer datum je lahko predstavljen kot eno polje ali v več poljih, celo znotraj enotnega seta podatkov in

3. med polji znanimi v času načrtovanja, številna polja ne bodo imela vrednosti.

Tehnike kakovosti podatkov postajo bolj kompleksne ko podatki izgubijo strukturo. Na primer predstavljajmo si register, ki vsebuje osebne podatke, kot so ime, priimek, ulico in mesto prebivanja. Tabela 12 prikazuje predstavljanje Nikole Murk, s prebivališčem Trebinjska ulica v Ljubljani, z uporabo strukturiranih nestrukturiranih, in semistrukturiranih vrst podatkov. Iste dimenzije kakovosti imajo drugačno metriko glede na vrsto podatkov.

Tabela 12: Različne predstavitve istega predmeta v realnem svetu

Nikola	Murk	Trebinjska ulica	Ljubljana	Strukturirani podatki
Nikola Murk, s prebivališčem na naslovu Trebinjska ulica, v mestu Ljubljana				Nestrukturirani podatki
<OsebniPodatki> <ImePriimek>Nikola Murk</ImePriimek> <Prebivalisce>Trebinjskaulica, Ljubljana</Prebivalisce> </OsebniPodatki>				Polstrukturirani podatki

Ortogonalna razvrstitev podatkov v literaturi o kakovosti podatkov temelji na ogledu podatkov kot proizvodnega izdelka (Shankaranarayan, 2000). S tega vidika se razlikujejo tri vrste podatkov:

- **neobdelani podatki** so opredeljeni kot podatki, ki niso bili predelani od njihovega nastanka in prvega shranjevanja, lahko so shranjeni za dlje časa;
- **informacijski produkt** je rezultat proizvodne dejavnosti, opravljene na podatkih;
- **sestaven podatek** (komponenta podatka) se generira vsakič, ko je potrebna ustrezna informacija; shranjeni so začasno, dokler končni izdelek ni izdelan – končni podatek – informacija ni kreirana.

Ta razvrstitev omogoča uporabo kakovosti tehnik na podatkih, ki se tradicionalno uporabljajo za zagotavljanje kakovosti v proizvodnih procesih.

Vrste podatkov vplivajo na dimenzije kakovosti podatkov in tehnike za ocenjevanje in izboljšanje. Tabela 13 povezuje vrste podatkov in metodologije kakovosti podatkov. V večini metodologij se govori o strukturiranih podatkih, medtem ko jih samo nekaj obravnava tudi polstrukturirane podatke. V Tabeli 13 je še opomba implicitna obravnava, ko metodologije eksplicitno ne omenjajo vrste podatkov, vendar faze in koraki lahko uporabijo določeno vrsto podatkov, na primer AIMQ uporablja splošen pojem informacije ter opravlja kvalitativne ocene skozi vprašanja, ki se nanašajo na strukturirane podatke, vendar lahko uporabimo kakršnekoli podatke, vključno z nestrukturiranimi podatki.

Tabela 13: Metodologije in vrste podatkov

Metodologija Akronim	Strukturirani	Polstrukturirani
TDQM	X	X
DWQ	X	
TIQM	X	implicitna obravnava
AIMQ	X	implicitna obravnava
CIHI	X	X
DQA	X	
IQM	X	X
ISTAT	X	X
AMEQ	X	implicitna obravnava
COLDQ	X	implicitna obravnava
DaQuinCIS	X	X
QAFD	X	
CDQ	X	X

Vir: C. Batini, C. Capielo, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:23.

3.7 Vrste informacijskih sistemov

Vrsta informacijskega sistema vpliva na metodologijo kakovosti podatkov, tako na ocenjevanje kot na izboljševanje. Različne arhitekture informacijskih sistemov, ali preprosto rečeno vrste informacijskih sistemov, se razlikujejo glede na stopnjo podatkov, procese in upravljanje z integracijami, ki jih podpirajo tehnični sistemi. Ko se stopnja integracije podatkov, procesa in upravljanja zmanjša, postanejo ocenjevanje kakovosti podatkov in tehnike za izboljševanje, ki jih je treba uporabiti, bolj zapletene. Hkrati postane ocenjevanje in izboljšanje kakovosti podatkov večji izziv. Glede na njihovo stopnjo integracije je mogoče razlikovati med naslednjimi vrstami informacijskih sistemov (Batini, Capielo, Francalanci & Maurino, 2009, str. 16:9):

- V **monolitnih informacijskih sistemih** so aplikacije enostopenjske (angl. *single-tier*) in ne zagotavljajo dostopa do podatkov. Čeprav so podatki običajno shranjeni v bazah podatkov, ki se lahko poizvedujejo, ločene aplikacije ne delijo podatkov. To lahko povzroči podvajanje podatkov in morebiten vpliv na vse dimenzije kakovosti.

- **Podatkovno skladišče** je centralizirana zbirka podatkov, vzpostavljena iz več podatkovnih baz. Podatkovna skladišča se redno (periodično) posodablajo – osvežujejo s podatki iz originalnih baz podatkov s procedurami za pridobivanje in uskladitev podatkov. Podatki so vključeni fizično, saj so preoblikovani v skladu s shemo podatkovnega skladišča, združeni in nazadnje shranjeni v podatkovno skladišče.
- **Porazdeljen informacijski sistem** (angl. *distributed information system*) je zbirka aplikacijskih modulov, usklajenih s potekom dela (angl. *workflow*). Aplikacije so običajno razdeljene na stopnje, kot so predstavitev, aplikacijska logika, upravljanje s podatki in izvozne funkcionalnosti za dostop do podatkov na različnih stopnjah. Podatki so lahko shranjeni v različnih podatkovnih bazah, interoperabilnost je zagotovljena z logičnim povezovanjem shem.
- **Sodelujoč informacijski sistem** (angl. *cooperative information system*, v nadaljevanju CIS) opredelimo kot ogromen informativni sistem, ki povezuje več različnih sistemov združbe, ki delijo skupne cilje (De Michelis, 1997). Sodelovanje z drugimi informacijskimi sistemi zahteva zmožnost izmenjave informacij. V CIS-u podatki niso logično povezani, saj so shranjeni v ločenih bazah podatkov v skladu z različnimi shemami. Vendar vključujejo aplikacije transformacije podatkov in postopke za izmenjavo, ki omogočajo interoperabilnost in sodelovanje med skupnimi procesi. Z drugimi besedami, integracija (povezovanje) je realizirana na ravni procesa.
- **Spletni informacijski sistemi** (angl. *web information system*, v nadaljevanju WIS) poimenujemo sisteme, ki uporabljajo spletno tehnologijo za zagotavljanje informacij in storitev uporabnikom ali drugim informacijskim sistemom. S tehničnega vidika je WIS aplikacija odjemalec/strežnik. Takšni sistemi običajno uporabljajo strukturirane, delno strukturirane in nestrukturirane podatke, in so podprti z razvojnimi in upravljalnimi orodji, ki temeljijo na tehnikah, značilnih za posamezne vrste podatkov.
- V **peer-to-peer informacijskih sistemih** (angl. *peer-to-peer information system*, v nadaljevanju P2P) ni nobene razlike med odjemalci (client) in strežniki. Sistem je sestavljen iz sklopa vozlišč, ki delijo podatke in servise, da bi izpolnili zahteve uporabnika. Značilnosti P2P sistema so: brez centralnega usklajevanja, brez centralne baze podatkov, nihče nima splošnega pregleda nad sistemom, klienti so avtonomni in se lahko s sistemom dinamično povežejo ali prekinajo povezavo.

Tabela 14: Metodologije in informacijski sistemi

Metodologija Akronim	Monolitni	Porazdeljen	Podatkovno skladišče	Deloven	Spletni
TDQM	fokusiran	implicitna obravnavna			
DWQ			močno fokusiran		
TIQM	fokusiran	fokusiran			

se nadaljuje

nadaljevanje

Metodologija Akronim	Monolitni	Porazdeljen	Podatkovno skladišče	Deloven	Spletni
AIMQ	fokusiran	implicitna obravnava			
CIHI	fokusiran	fokusiran			
DQA	fokusiran	implicitna obravnava			
IQM					močno fokusiran
ISTAT	fokusiran	fokusiran		močno fokusiran	
AMEQ	fokusiran				
COLDQ	fokusiran				
DaQuinCIS	fokusiran	fokusiran		močno fokusiran	
QAFD	fokusiran				
CDQ	fokusiran	fokusiran		močno fokusiran	

Vir: C. Batini, C. Cappiello, C. Francalanci in A. Maurino,
Methodologies for data quality assessment and improvement, 2009, str. 16:24.

Tabela 14 prikazuje štiri-vrednostno lestvico, kjer:

1. **močno fokusiran** pomeni, da je celotna organizacija metodologije zasnovana in prikrojena ustreznem tipu informacijskega sistema, medtem ko zagotavlja splošne napotke za druge vrste informacijskih sistemov,
2. **fokusiran** pomeni, da določa metodologija podrobna navodila in tehnike za ustrezno vrsto informacijskega sistema,
3. **implicitna obravnava** ima enak pomen kot v Tabeli 13 in
4. polja brez vrednosti pomenijo, da metodologija ali določa splošna navodila ali izrecno ne navaja ustreznega tipa informacijskega sistema.

Spletni informacijski sistem je vključen v tabelo, ker je ena metodologija močno fokusirana nanj. V Tabeli 14 lahko opazimo, da so metodologije AMEQ, COLDQ in QAFD fokusirane na monolitne informacijske sisteme. Običajno obravnavajo strukturirani niz podatkov v enem sistemu in ignorirajo kakovost podatkov pri izmenjavi podatkov med ločenimi aplikacijami ali organizacijami. TIQM je mogoče uporabiti na monolitnih in porazdeljenih sistemih, ker so pri analizi podatkovne strukture monolitni in porazdeljeni sistemi ponujeni kot študije primerov. CIHI je fokusirana na porazdelje sisteme, saj so nacionalne baze podatkov navedene v opisu metodologij. Druge metodologije se lahko uporabljajo na več vrstah sistemov. Vsaka metodologija zagotavlja podrobna navodila za najzahtevnejšo vrsto sistema, delovni informacijski sistema (angl. *cooperative information system*) in se zato lahko uporabi tudi na monolitnih in porazdeljenih sistemih.

Metodologije sledijo različim pristopom pri obravnavanju tokov podatkov in heterogenih bazah podatkov.

4 OPREDELITEV KAKOVOSTI PODATKOV

4.1 Predstavitev poslovne banke X

Poslovna banka X (tako sem jo poimenoval zaradi varovanja poslovne skrivnosti) na slovenskem trgu uspešno deluje že od leta 1993, ko je mednarodna bančna skupina prevzela eno od slovenskih bank.

Omenjeno mednarodno bančno skupino so leta 1922 ustanovile komercialne posojilne zadruge. Že leta 1991 je bila ena od prvih bankčnih skupin, ki je sprejela izziv širjenja na trge Srednje in Vzhodne Evrope. Bančna skupina X danes upravlja uspešno in rastočo bančno mrežo v Srednji in Vzhodni Evropi, ki jo sestavlja več kot 600 poslovalnic v devetih državah Srednje in Vzhodne Evrope (Poslovna banka X, 2009).

Na lokalnem trgu spada Poslovna banka X med manjše banke, igra pa zelo pomembno vlogo na posamičnih segmentih. Povezanost in vpletenost v mednarodni koncern jo istočasno uvršča med večje bančne sisteme. Ta kombinacija pa določa tudi filozofijo njenega dnevnega poslovanja s strankami: fleksibilna in dinamična, strankam pa nudi popolnoma individualno obravnavo finančnih potreb, ob hkratni varnosti in podpori velikega koncerna.

V nadaljnjem tekstu bom uporabljal pojem podružnica za skupek pojmov – hčerinska banka, mednarodna banka, poslovna banka, lizing.

4.2 Razlaga pojmov

Centralna baza podatkov (v nadaljevanju CBP) je dejansko skladišče podatkov. Podatki prihajajo v centralno bazo iz vseh podružnic in v vnaprej dogovorenem formatu oziroma so narejeni po specifikaciji, določeni s strani koncerna. Gre za podatkovni paket – datoteko, ki vsebuje ali vse zapise (mesečno pošiljanje po koncu vseh mesečnih obdelav) ali samo spremenjene oziroma nove zapise prejšnjega dneva (dnevno pošiljanje). Ti podatki se po preverjanju tehnične ustreznosti ter verodostojnosti zapišejo v CBP in so kasneje osnova za vse nadaljnje izračune.

RiWA (angl. *Risk Weighted Assets*) je centralni del izračunavanja minimalnih kapitalskih zahtev, odvisen je od tega, kateri pristop je banka sprejela. Na osnovi podatkov iz CBP, se v RiWi izračunava potrebna višina minimalnih kapitalskih zahtev.

SAP je računovodski del koncernskega Basel II sistema – obravnava stroške, tržna in operativna tveganja. Definiran je čisto posebej, sistem delovanja je enak kot pri CBP.

ZKO je podsistem koncernskega Basel II sistema za oblikovanje informacij o strankah, skupinah povezanih oseb, prevzetih tveganjih, poroštvih, skupni izpostavljenosti, ipd. na nivoju koncerna. Za uspešno poslovanje celotnega sistema je pomembna povratna informacija iz ZKO nazaj v banke. Na podlagi tega imajo vsi subjekti nujne informacije o izpostavljenosti strank na nivoju koncerna in je zato njihovo poslovanje manj tvegano.

Rating je zaključen podsistem koncernskega Basel II sistema, v katerem so stranke razdeljene po skupinah (fizične osebe, pravne osebe...). Strankam se v tem sistemu izdela bonitetna ocena – določi rating, ki je temelj IRB pristopa po Basel II.

IRB pristop je osnovan na lastnih bonitetnih ocenah banke in bankam omogoča določanje višine kapitala, ki bolje odraža dejansko tveganost bančnega portfelja. Obstajata 2 načina IRB pristopa, osnovni in napredni. Osnovni IRB pristop je zgrajen na dveh temeljnih načelih. Prvo načelo je oblikovano na podlagi kakovosti lastnih informacij o sredstvih banke, drugo načelo pa temelji na kakovostnem delu risk managementa in s tem na dobrih analizah tveganja komitenta. Na ta način IRB pristop prilagaja kapital banke resničnemu ekonomskemu stanju poslovnega okolja banke. Ključni elementi IRB pristopa so: **dejavniki tveganja** (opredelijo jih banke in jih nato vključijo v kalkulacijo); **uteži tveganja** (način, s katerim dejavnike tveganja preoblikujemo v kapitalske zahteve) in **minimalne zahteve** (minimalni standardi, ki so določeni za zmanjšanje tveganosti posla). Banke, ki uporabljajo osnovni IRB pristop, morajo zagotavljati svoje ocene **verjetnosti neplačila** (angl. *Probability of default*, v nadaljevanju PD) ter uporabljati ocene ostalih komponent tveganja (**izgube v primeru neplačila** (angl. *Loss Given Default*, v nadaljevanju LGD), prilagojeno **izpostavljenost neplačilu** (angl. *Exposure at Default*, v nadaljevanju EAD)) kot jih določa nacionalni nadzornik. Napredni IRB pristop se razlikuje od osnovnega IRB pristopa v tem, da uporabljajo banke svoje ocene za vse komponente tveganja (PD, LGD in EAD). Omenjeni pristop predvideva zmanjšanje zahtevane višine kapitala za 10–20 % glede na osnovni IRB pristop.

4.3 Osnovni model Basel II sistema

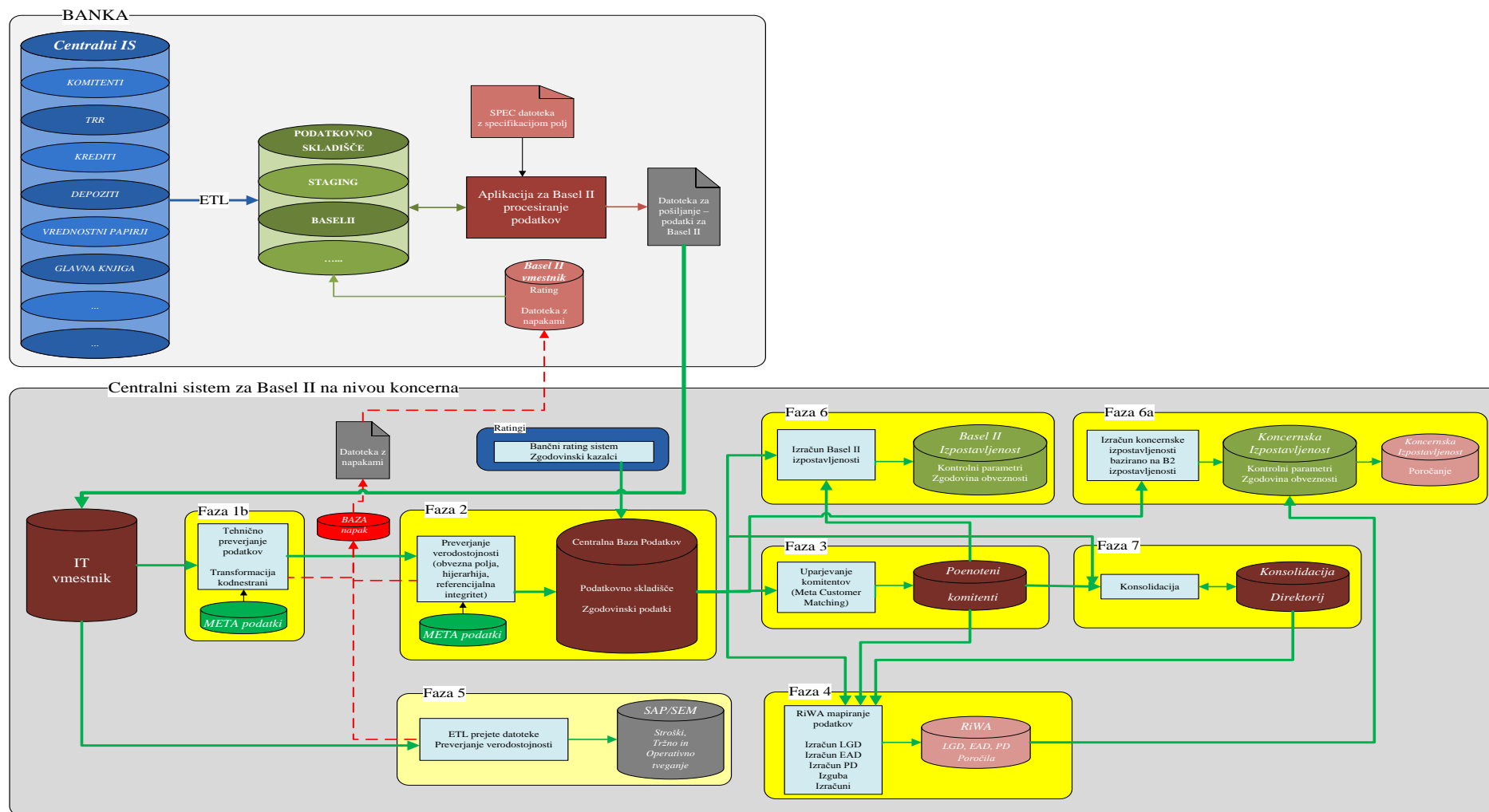
V začetku samo želja bančne skupine, danes pa tudi potreba, je imeti podatke iz podružnic zbrane na enem mestu in sicer v matični državi bančne skupine. Razvita je centralna baza podatkov, naloga podružnic pa je pošiljati kakovostne podatke, prilagojene standardom – specifikaciji na nivoju celotne bančne skupine. S tem pridobimo celoten pregled na nivoju bančne skupine (nad svojimi podružnicami in tudi nad posameznimi komitenti). S tem se poenoti poslovanje finančne združbe in zmanjša tudi tveganja. Če je podružnica pred sprejetjem novega posla z določenim komitentom, lahko hitro ugotovi, za kakšno stopnjo tveganja gre pri novem poslu. Dogajalo se je, da je bil določen komitent pri določeni banki

že močno zadolžen in ni plačeval obveznosti, pa je kljub temu lahko sklenil nove posle pri drugi banki iz iste bančne skupine.

Slika 2 prikazuje osnovni model delovanja informacijskega sistema banke, ki je povezan s centralnim sistemom za Basel II na nivoju bančne skupine oziroma koncerna. Kot vidimo, je slika razdeljena na dva bistvena dela. Zgornji del opredeljuje delovanje in podatke Poslovne banke X, spodnji del pa prikazuje celotno delovanje na nivoju koncerna s CBP. Kot prikazuje zgornji del slike, banke delujejo samostojno in do določenih zneskov same sprejemajo vse poslovne odločitve. Pri večjih poslovnih odločitvah pa se odloča na nivoju koncerna, kjer se odločajo na podlagi podatkov in analiz iz centralne baze podatkov.

Zaposleni v Poslovni banki X vnašajo vse podatke o strankah, sklenjenih poslih, zavarovanjih, itd. v centralni informacijski sistem banke, od koder se podatki, po dnevni ali mesečni obdelavi, prenašajo po določenih poslovnih pravilih v podatkovno skladišče. V podatkovnem skladišču se nad določenimi podatki, na podlagi zahtev specifikacije, zaženejo ETL procedure, ki zapišejo podatke v tabele, definirane s specifikacijo koncerna. Nato se z aplikacijo za Basel II procesiranje podatkov, spet na podlagi specifikacije, kreira datoteka, ki se dnevno/mesečno pošilja v koncernski centralni sistem za Basel II, kje se potem podatki obdelujejo po določenih fazah. Rezultati obdelave so vidni preko poročil, ali kot povratna datoteka v banko – ponavadi samo napake. Na podlagi tega ima koncern celoten pregled nad podatki in poslovanjem podružnic, posledično pa bolj enostavne primerjave med posameznimi podružnicami in tudi med posameznimi državami, v katerih banka posluje.

Slika 2: Osnovni model sistema



4.4 Predstavitev okvira kakovosti podatkov

To poglavje o kakovosti podatkov obravnava kakovostni vidik pošiljanja podatkov v CBP v bančni skupini, od primarnega vnosa v informacijski sistem banke, do prenosa podatkov v RiWa/SAP. Logika izračunavanja in nadaljnjo procesiranje v RiWA/SAP nista tema te naloge.

Za zagotovitev kakovosti podatkov so definirani trije temeljni procesi, ki so podrobneje opisani v nadaljevanju:

1. Proces dostave;
2. Proces spremljanja in eskalacije;
3. Proces upravljanja z specifikacijami.

4.4.1 Temeljni procesi za zagotovitev kakovosti podatkov

4.4.1.1 Proces dostave

Podatki, ki so v informacijskih sistemih posameznih bank in se pozneje posredujejo v CBP in SAP, morajo biti vnešeni ali posodabljeni z ustrežno kakovostjo in popolnostjo. Obstajata dva različna cikla dostave podatkov za CBP in SAP:

- **Dostava ob koncu meseca:** pošiljanje vseh potrebnih podatkov ob koncu meseca. Obvezno morajo biti dostavljeni naslednji podatki:
 - obvezna polja in pogojno obvezna polja opredeljena v specifikaciji;
 - polja, označena kot koncernsko interna (izjeme so še posebej poudarjene);
- **Dnevna dostava:** dnevna dostava novih ali spremenjenih podatkov, zapisov določenih poslov, v skladu z navodili centralnega strokovnjaka.

Potek procesa dostave:

- **Vnos podatkov:** poslovni podatki se nenehno vnašajo v informacijske sisteme bank. Pristojni management ali uprava nosi končno odgovornost za kakovost podatkov.
- **Kreiranje dostavne datoteke:** podobno kot seznam za kontrolo dobave, banka zagotovi ali organizira pridobivanje potrebnih in predvidenih podatkov iz informacijskega sistema banke ter pošlje to datoteko s podatki na podlagi veljavne SPEC. Odgovornost za kakovost podatkov (hitro in dostavo brez napak), je v celoti na dobavnih enotah – posameznih bankah.
- **Sprejem in začasno shranjevanje podatkov:** po sprejemu se dostavljena datoteka procesira. Banke lahko sledijo statusu obdelav preko poročil:

- **tehnični pregled formata (faza 1b):** preverja se, ali format prenešene datoteke ustreza zahtevam SPEC. Tehnične napake se zapišejo v bazo napak.
- **logični pregled (faza 2):** preverja se, ali so poslana vsa obvezna in pogojno obvezna polja ter referencialna integriteta podatkov. Ugotovljene napake se shranijo v bazo napak, in jih je mogoče naknadno pridobiti preko poročil.

Fazi 1b in 2 sta popolnoma avtomatizirani, ocene pregledov se avtomatično prenesejo v bazo napak. Baza napak podatkov je temelj za oceno zagotavljanja kakovosti dobave.

- **Začetek procesiranja:** procesiranje se zažene samodejno. Če banka ni dostavila datoteke, se podatki ne morejo prenesti v CBP in ne pride do nadaljnjega procesiranja. Pri popolnem neuspehu dostave analiziramo vzroke in ustrezno obvestimo odgovorne osebe.
- **SAP podatki:** po tehničnem pregledu podatkov (faza 1b), se podatki, določeni za SAP, prenesejo do SAP-a. CBP procesiranje, logično preverjanje (faza 2) se za SAP ne izvrši. Namesto tega uporablja SAP lastna pravila, rezultati teh pravil so na voljo preko poročila o SAP napakah.

4.4.1.2 Proces spremljanja in eskalacije

Razen *ex-ante* spremljanja v okviru procesa dostav podatkov, je za kakovost podatkov prav tako ključnega pomena *ex-post* spremljanje po procesiranju v RiWa. Ta dva vidika sta združena v en pregled, na ta način je omogočen celovit pregled nad kakovostjo podatkov. Ta pregled zagotavlja celoten pregled kakovosti podatkov posameznih bank na managerski ravni.

Nadzor kakovosti podatkov je razdeljen na spremljanje tehnične in funkcionalne kakovosti ter proces eskalacije, ki je povezan s konceptom nadzora in daje povezavo med kakovostjo podatkov, posredovanjem ustreznih operativnih ukrepov in zahtevano transparentnostjo za management.

4.4.1.2.1 Proces spremljanja

Tehnično spremljanje. V okviru tehničnega spremljanja se oblikujejo poročila, ki podpirajo prepoznavanje virov tehničnih napak. Na voljo so naslednja poročila:

- **Razmerje napak dostavljenih podatkov** (nanaša se na CBP): poročilo prikazuje število dostavljenih zapisov posamezne banke. Vsak nepravilen zapis je predstavljen kot odstotek razmerja odstopanja napačnih zapisov in količine vseh zapisov.
- **Neprenešene transakcije iz CBP v RiWa** (CBP pomembno): poročilo prikazuje število zapisov v CBP za posamezno banko in zapise, ki niso bili prenešeni v nadaljnje

procesiranje v RiWa. Poleg pregleda, ki vsebuje odstotek odstopanja, prikazuje podrobni pogled tudi neprenehane transakcije, razčlenjene do ravni posameznega posla.

Vsebinsko spremljanje. Za identificiranje napak funkcionalne kakovosti se uporabljajo naslednji instrumenti:

- **Celoten test:** pokaže odstopanja (deviacijo) v podatkih pri primerjavi podatkov iz RiWa in bilance banke, poslani s strani kontrolinga banke na koncern. Podatki se preverijo glede popolnosti z uporabo celotnega testa na ravni poslov.
- **Rating prepustnost:** poročilo prikazuje delež »strank z veljavnim ratingom«, kot tudi »stranke s kritično ali neveljavno kombinacijo Basel II segmenta in rating orodja«, »stranke s starim ratingom« in »stranke brez ratinga«. Ti deleži so pomembni parametri v procesu eskalacije.
- **Standardizacija ratingov in segmentacija:** preverja se število interno prepisanih ratingov/segmentov. Če interno prepisani ratingi/segmenti presegajo 1 %, se začne proces eskalacije. Možni popravki morajo biti na pobudo lokalne funkcionalne enote.
- **Zavarovanja:** več poročil preverja popolnost in pravilnost podatkov o zavarovanjih od vnosa podatkov do njihove dostave v CBP. Vsebujejo najpomembnejše informacije v zvezi s zavarovanji (hipoteke, ostalo): skupno število in vrednost zavarovanj, trajanje zavarovanj, podroben pregled vseh zavarovanj. Ta poročila nam pomagajo pri ugotavljanju manjkajočih ali preseženih zavarovanj.
- **Segmentacija:** poročila segmentacije strank preverjajo verodostojnost strankine segmentacije. Izvaja se na ravni koncerna.

4.4.1.2.2 Proces eskalacije

Proces eskalacije se začne z odkrivanjem nepopolne dostave ali napake v dostavi in služi za zagotovitev visoke kakovosti podatkov.

Takoj, ko so na voljo podatki iz RiWe, se sprožijo naslednje testne rutine:

- Celoten test – hitri test na ravni posla,
- Rating prepustnost,
- Ostala poročila povezana z razširitvijo testne rutine.

Zadolžene osebe morajo pregledati poročila in komentirati rezultate ter dati svoja strokovna mnenja. Rezultati se prenesejo v skupna poročila. Če en ali več kakovostnih pregledov, opisanih v naslednjih podpoglavljih, odkrije odstopanja za več kot 5 % (razen, če je določeno drugače), zadolžena oseba naroči pripravo akcijskega načrta, da bi se težava odpravila. Če se kakovost podatkov ne izboljša po koncu 2. meseca, je postopek podoben kot ob koncu 1. meseca – eskalacijski ukrepi se ne določijo. Če se kakovost podatkov ne izboljša niti po koncu 3. meseca, je eskalacijski ukrep obveščanje pristojne uprave

podružnice z nekakovostnimi podatki. Če niti po koncu 4. meseca ni opaziti nobenih izboljšanj, je eskalacijski ukrep obveščanje uprave koncerna.

4.4.1.3 Proces upravljanja s specifikacijo

Podatki v dostavljenih datotekah (komitenti, transakcije, posli,...) se dostavljajo na podlagi veljavne različice dostavne specifikacije (v nadaljevanju SPEC). Ta opredeljuje, kateri podatki morajo biti dostavljeni in v kakšnem formatu iz različnih informacijskih sistemov bank v skupno CBP. Zahteve po podatkih se med letom spreminjajo (poročanje, upravljanje s tveganji, itd.), zato je logično da se spreminja tudi dostavna specifikacija. SPEC je razdeljena na vrste zapisov, seznam je v Prilogi 1 (na primer komitenti, krediti, hipoteke, stanja,...), vključujejo podatke o komitentih in produktno-specifične podatke v obliki definiranih podatkovnih polj. Podatkovna polja so natančno specificirana, tako tehnično (dolžina, format, itd.) kot funkcionalno (vsebina, obvezno polje, itd.). Vsako podatkovno polje je povezano vsaj z enim strokovnim področjem, za katerega sta določena centralni strokovnjak in skrbnik.

4.4.1.3.1 Upravljanje z izdajami SPEC

Cilj upravljanja z izdajami SPEC je balansiranje dveh različnih zahtev: po eni strani se v dinamičnem okolju zahteve po informacijah spreminjajo, na drugi strani pa je plan dostave zagotoviti stabilnost in visoko kakovost podatkov. V ta namen je predvidena letna izdaja SPEC z jasno definiranimi zadnjimi zahtevki po podatkih, sporočenimi vnaprej. Ta letni cikel se lahko spremeni le v primeru nepričakovanih zahtev regulatorjev.

Da bi se zagotovilo koordinirano procesiranje (obdelovanje) zahtevkov o dodatnih podatkovnih zahtevah, se pošlje zadolženi osebi zahteve za spremembo SPEC. Vsebino distribuiranega zahtevka po spremembah pregledajo (razumnost, podvajanje) poslovni strokovnjaki vseh področj. Centralni poslovni strokovnjaki so odgovorni za to, da se odločitev v zvezi implementacijo (upoštevanjem) zahtevka po spremembi pripelje v najkrajšem možnem času do distribucije. Če eden od strokovnjakov zavrne zahtevek v pisni obliki, se razlaga posreduje vsem drugim predhodno obveščenim osebam. Če je zahtevek po spremembi odobren, se vnese v bazo zahtevkov.

4.4.1.3.2 Zaključevanje nove različice SPEC

Mednarodni IT center (v nadaljevanju MITC) objavi osnutek nove različice SPEC in jo pošlje odgovornim osebam v pregled in odobritev. Ta osnutek vključuje tudi zahteve po spremembah, ki so jih registrirani pred zapiranjem roka. Odgovorne osebe preverjajo popolnost in točnost vsebine registriranih zahtevkov. Po končanem procesu preverjanja se rezultati sporočijo na MITC. Na podlagi vseh poročil popravkov pripravi MITC končno

različico SPEC, upoštevajoč tudi dokumente za spremembo ter jo pošlje vsem udeleženkam procesa Basel II v implementacijo – izvajanje.

V okviru objave nove SPEC zagotavlja koncern organizacijo in realizacijo strokovnih delavnic. Te morajo potekati takoj po distribuciji nove različice SPEC z udeležbo MITC in centralnih poslovnih strokovnjakov. Da bi se pripravile strokovne delavnice, so udeleženke zaprosene, da pošljejo svoja vprašanja o novi različici SPEC vnaprej centralnim poslovnim strokovnjakom. Namen strokovnih delavnic je dati možnost vsem ustreznim enotam da se predčasno spoznajo z novo dobavno specifikacijo. Organizira se forum, v katerem si ustrezni nosilci znanj v koncernu izmenjujejo informacije, povezane z vsebino, ali tehnična vprašanja o dostavi.

Nova različica SPEC lahko zahteva spremembe tudi v lokalnih informacijskih sistemih. Tudi temu je treba nameniti ustrezní časovni okvir v procesu upravljanja s specifikacijami. Ključnega pomena je tudi sprememba aplikacije za Basel II procesiranje podatkov, saj ta aplikacija zbira, kontrolira, konsolidira ter transformira Basel II relevantne podatke iz različnih podatkovnih virov ter na koncu izvozi podatke v datoteko za dostavo, ob tem pa izvaja določene kontrole za zagotovitev pravilnosti podatkov. Implementacija nove različice SPEC pri posamezni udeleženci se zaključi s popolno in, v smislu vsebine, pravilno produkcijsko dostavo na podlagi nove različice SPEC.

4.4.2 Opredelitev delovnih mest

Zagotavljanje kakovosti podatkov je dolgotrajen proces, ki traja skozi celoten življenjski cikel podatkov. V njem so vključeni dejavniki, ki jih je potrebno upoštevati pri sami zagotovitvi uspeha. Poleg tehnologije, ki je vključena v sam proces, pa tu ne smemo pozabiti tudi na procese in ljudi.

V nadaljevanju bom opisal delovna mesta v procesu zagotavljanja kakovosti podatkov pri Basel II podatkih.

Izvajanje zahteva sedem profilov delovnih mest v bančni skupini:

1. Koncern:

- Centralni upravljalec kakovosti podatkov (CDQM),
- Centralni poslovni strokovnjaki – različna poslovna področja;

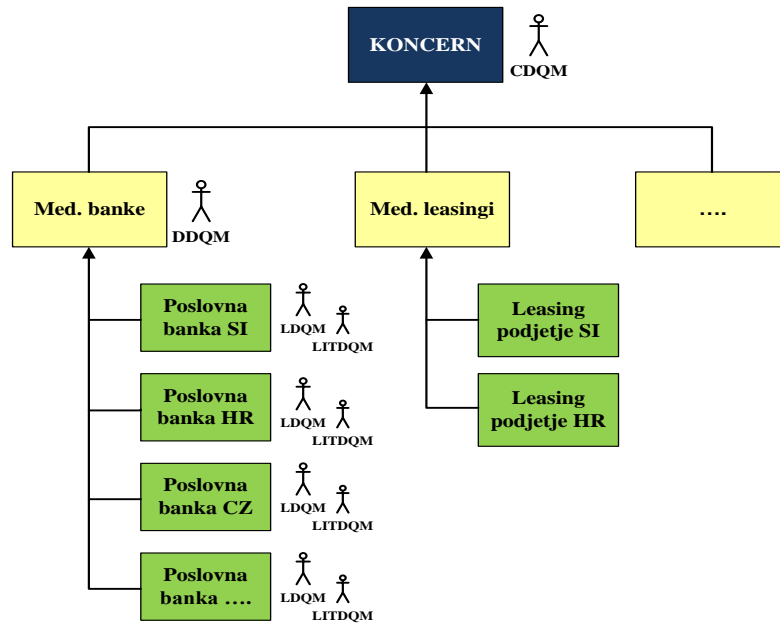
2. Skupina Mednarodne banke:

- Decentralizirani upravljalec kakovosti podatkov (DDQM),
- Decentralizirani poslovni strokovnjaki – različna poslovna področja;

3. Lokalne poslovne banke v raznih državah:

- Lokalni upravljalec kakovosti podatkov (LDQM),
- Lokalni poslovni strokovnjaki – različna poslovna področja,
- Lokalni IT upravljalec kakovosti podatkov ITDQM.

Slika 3: Hierarhična struktura odgovornih za kakovost podatkov



4.4.2.1 Pravila komuniciranja

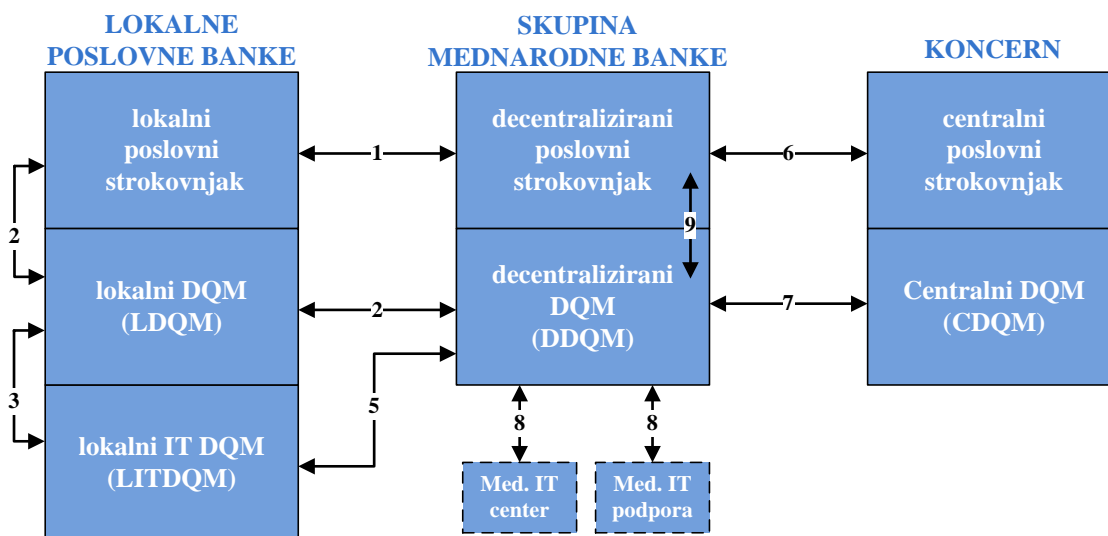
Da bi vzpostavili hierarhičen način komuniciranja in preprečili nenadzorovano obdelavanje vprašanj o kakovosti podatkov med interesnimi skupinami, so definirani komunikacijski tokovi oziroma pravila, ki jih prikazuje Slika 4.

Komunikacijski liniji 1 in 6: Decentralizirani DQM (v nadaljevanju DDQM) je brez funkcionalnega strokovnega znanja. Lokalni poslovni strokovnjaki svoja funkcionalna vprašanja pošiljajo neposredno na decentralizirane poslovne strokovnjake, medtem ko decentralizirani poslovni strokovnjak naveže stike s svojim kolegom, centralnim poslovnim strokovnjakom. Lokalni poslovni strokovnjak se lahko obrne na centralnega poslovnega strokovnjaka neposredno samo, če ni decentraliziranega poslovnega strokovnjaka, ki pokriva specifično poslovno področje. Predpogoj uspešnega komuniciranja je, da lahko lokalni, decentralizirani in centralizirani poslovni strokovnjaki strokovno podpirajo proces.

Komunikacijska linija 2: Lokalni poslovni strokovnjak se obrne na lokalnega DQM (v nadaljevanju LDQM) v zvezi z vsemi vprašanji, vezanimi na lokalno organizacijo in procese DQM. Dvosmerna komunikacija je pogoj za zagotovitev visoko kakovostnih podatkov na lokalni ravni.

Komunikacijski liniji 3 in 5: ITDQM dvosmerno komunicira z LDQM ali DDQM vezano na izvajanje SPEC projekta, lokalnega podatkovnega skladišča, procesa dostave podatkov in odpravljanja napak.

Slika 4: Komunikacijski tokovi



Vir: Interno gradivo Poslovne banke X

Komunikacijski liniji 4 in 7: Glavni kontaktni osebi DDQM sta LDQM in CDQM. Vsi skupaj delujejo kot komunikacijsko središče za vsa organizacijsko procesna vprašanja/zahteve v svojih bankah. DDQM zagotavlja 1. nivo podpore LDQM, CDQM pa 2. nivo podpore (od spodaj-navzgor, funkcionalne in tehnične zahteve). Vsa lokalna DQ (funkcionalna/tehnična) vprašanja, ki jih ni mogoče obravnavati brez podpore DDQM, je treba nasloviti na DDQM v pisni obliki.

Komunikacijska linija 8: MITC in Mednarodna IT podpora (v nadaljevanju MITP) niso hčerinske družbe skupine Mednarodne banke, se lokalni ITDQM dogovarja z DDQM o morebitnih dodatnih IT stroških, vezanih na izvajanje Basel II procesa. V primeru, da stroški izhajajo iz lokalnih zahtev (LDQM in lokalni poslovni strokovnjaki) – zahteva se tehnična ali funkcionalna izboljšava, se pridobi dovoljenje pristojnega decentraliziranega poslovnega strokovnjaka.

Komunikacijska linija 9: Vsa koncernska vprašanja, povezana s kakovostjo podatkov, obravnava DDQM s funkcionalno podporo decentraliziranega poslovnega strokovnjaka.

4.4.2.2 Opredelitev delovnih mest, zadolženih za kakovost podatkov

4.4.2.2.1 Centralni upravljelec kakovosti podatkov

Centralni upravljelec kakovosti podatkov (angl. *central data quality manager*, v nadaljevanju CDQM) je povezava med centralnimi poslovnimi strokovnjaki in DDQM. Jasen poudarek njegovega dela leži na usklajevanju postopkov med ostalimi udeleženci. CDQM ne snosi končne odgovornosti za kakovost dobavljenih podatkov, njegova

odgovornost je predvsem v pravočasnem spremljanju in eskalaciji, kot tudi komuniciranju in usklajevanju med DDQM, različnimi strokovnjaki in MITC glede ključnih nalog, ki so (razdeljeno po procesih):

1. Proces dostave:

- koordinacija temeljnih sprememb v nastavitvah dostave podatkov,
- podpora skupini Mednarodnih bank neposredno pred iztekom roka za dostavo podatkov (zlasti koordinacija z ustreznimi strokovnjaki za odločitve v zvezi z novimi dostavami),
- izvedba odločitve na podlagi pisnega naloga DDQM;

2. Proces spremljanja in eskalacije:

- povzem testnih rezultatov v pregledu statusa kakovosti podatkov in podeljevanje splošnega (celotnega) statusa skupaj s pristojnimi centralnimi poslovnimi strokovnjaki,
- odločitev glede akcijskega načrta, če obstaja znatno odstopanje, in spremljanje tega akcijskega načrta,
- obvešča upravo če ne pride do zboljšanja kakovosti podatkov po večmesečni eskalaciji,
- gonilna sila procesa za nadaljni razvoj procedur nadzora;

3. Proces upravljanja s specifikacijo:

- opazovanje cikla izdaje in zgodnja komunikacija z ostalimi udeleženci,
- zbiranje in posredovanje zahtevkov za spremembo specifikacije v skladu s procesom za upravljanje s specifikacijo, posodobitev baze zahtevkov za spremembe in usklajevanje ob posvetovanju z MITC,
- organiziranje strokovnih delavnic za zgodnje izobraževanje, pred izdajo nove specifikacije,
- usklajevanje in spremljanje pravilne uporabe specifikacije v celotnem koncernu.

4.4.2.2.2 Centralni poslovni strokovnjaki

Centralni poslovni strokovnjaki so odgovorni za vsebinske in metodične opredelitve podatkov, potrebnih za njihova poslovna področja. Zahteve centralnih poslovnih strokovnjakov morajo biti usklajene z vsemi poslovnimi področji zaradi preprečevanja podvajanja podatkovnih polj in nerazumevanja njihovega pomena. Njihove ključne naloge so razdeljene po procesih:

1. Proces dostave:

- posodobitve seznamov udeležencev procesa,
- priporočila za nove dostave v skladu z dostavo ob koncu meseca;

2. Proces spremljanja in eskalacije:

- komentiranje in razlaga ocen kakovosti dostavljenih podatkov poslanih od DDQM,

- podpora CDQM pri odločitvi po potrebi akcijskega načrta,
- tehnična podpora DDQM pri analizi vzrokov in pripravi akcijskega načrta za odpravljanje napak;

3. Proces upravljanja s specifikacijo:

- nadzor, usklajevanje in končna odobritev ali zavrnitev zahteve po spremembah podatkovnih polj iz njihovega poslovnega področja (usklajena odločitev),
- nadzor, usklajevanje in končna odobritev ali zavrnitev osnutka specifikacij,
- naloge povezane s strokovnimi delavnicami,
- kontakt za funkcionalne zadeve,
- podpora implementaciji specifikacije vezano za nova polja za DDQM - podrobna pojasnila in opisi vseh polj,
- posodobitev dokumentacije specifikacije v sodelovanju z CDQM,
- kontaktna oseba za vprašanja v zvezi z ustreznimi polji specifikacije.

4.4.2.2.3 Decentralizirani upravljalec kakovosti podatkov

Decentralizirani upravljalec kakovosti podatkov (angl. *decentralized data quality manager*, v nadaljevanju DDQM) zagotavlja izvajanje procesa kakovosti podatkov in nalog v posamezni podružnici koncerna. Prevzema predvsem odgovornost procesnega usklajevanja, pri čemer je končna odgovornost za zahtevano kakovost podatkov posamezne podružnice na ustreznem managementu. V vsaki podružnici je potrebno na ravni delovnih mest zagotoviti tehnično ter vsebinsko odgovornost za popoln in pravilen vnos podatkov v informacijske sisteme ter pripadajoče odgovornosti dokumentirani v lokalnih smernicah bank. DDQM je ekskluzivni partner LDQM in CDQM vezano na vprašanja o kakovosti podatkov. Njegove ključne naloge, razdeljene po procesih, so:

1. Proces dostave:

- zagotavljanje kakovosti podatkov iz informacijskih sistemov posameznih podružnic – predvsem popolnost in točnost,
- odgovornost za popolno, hitro in točno dostavo podatkov ob koncu meseca ali dneva, predvsem izvajanje *ex-ante* testov (testi pred začetkom/koncem procesiranja podatkov),
- analiza poročila o dostavi, kreiranih s strani MITC,
- pisno naročilo CDQM za implementacijo bančnih odločitev;

2. Proces spremljanja in eskalacija:

- decentralizirana izvedba *ex-post* testov (testi po dostavi in procesiranju podatkov, predvsem rating prepustnost (penetracija), celoten test vseh poglavij) in obveščanje centralnega strokovnjaka s poslovnega področja ali CDQM,
- preiskava vzroka, v primeru nenadnega poslabšanja kakovosti dostavljenih podatkov, v primerjavi s preteklim mesecem,
- oblikovanje akcijskega načrta in zagotovitev decentraliziranega izvajanja v smislu povečanja kakovosti podatkov v procesu dostave podatkov,

- nadzor izvajanja ukrepov v posameznih podružnicah in koncernu, z uporabo poročil o dostavi in poročil o napakah dostave naslednjega meseca,
- distribucija filtriranih informacij o napaki podružnicam/osebam odgovornim za vnos ali naknadno spremembo,
- zbiranje informacij o operativnem tveganju od decentraliziranih strokovnjakov področjih,
- pošiljanje informacij o operativnem tveganju koncerna,
- vzdrževanje/posodobitev statusa kakovosti podatkov,
- informiranje CDQM, če obstaja precejšnja deviacija od dogovorjenega deviacijskega odklona 2 % v dostavi podatkov,
- mesečno poročanje o statusu kakovosti podatkov upravnem odboru koncerna in poslovnim oddelkom koncerna, ki temelji na statusu kakovosti podatkov, ki ga dostavi LDQM;

3. Proces upravljanja s specifikacijo:

- usklajevanje procesa za zahtevo sprememb specifikacije,
- informiranje poslovnih področjih koncerna in vseh LDQM o letnih ciklih izdaje specifikacije in kratkoročna opozorila pred internim iztekom roka,
- uvedba osnutka ali končne različice specifikacije v lastni združbi,
- koordinacija priprave in udeležba na strokovnih delavnicah koncerna z (de)centraliziranimi poslovnimi strokovnjaki, MITC in MITP,
- zagotovitev izvajanja potrebnih sprememb v informacijskih sistemih podružnic, na podlagi zadnje različice specifikacije,
- informiranje CDQM o trenutnem stanju (na primer projektni načrt) znotraj procesa upravljanja s specifikacijo,
- usklajevanje izobraževanja končnih uporabnikov,
- podpora ostalim članom koncerna pri analizi in testiranju, kot del uporabe nove različice specifikacije.

4.4.2.2.4 Decentralizirani poslovni strokovnjaki

Jasna opredelitev funkcionalnih zahtev in algoritmov, potrebnih za zagotovitev pravilnega in nemotenega izvajanja procesa, so splošne zadolžitve decentraliziranega poslovnega strokovnjaka. Decentralizirani poslovni strokovnjaki so odgovorni za:

- opredelitev ključnih dejavnikov uspeha (v nadaljevanju KDU), pomembnih za njihovo poslovno področje,
- vsebinsko in funkcionalno podporo ter metodično opredelitev podatkov, potrebnih za njihovo področje.

Vse zahteve se obravnavajo in dogovarjajo z vsemi decentraliziranimi poslovnimi strokovnjaki. Če decentralizirani poslovni strokovnjak ne prevzame zgoraj navedene

odgovornosti, lahko DDQM ne upošteva slabe kakovosti podatkov in posledičnih vplivov (kolateralne škode). Njihove ključne naloge so razdeljene po procesih:

1. Proces spremljanja in eskalacije:

- komentarji k ocenam kakovosti podatkov po dostavi in procesiranju podatkov,
- podpora DDQM pri odločanju o potrebi lokalnega akcijskega načrta,
- funkcionalna podpora DDQM med analizo glavnega vzroka zmanjšane kakovosti podatkov in oblikovanje akcijskega načrta za čiščenje podatkov,
- poročanje o operativnem tveganju;

2. Proces upravljanja s specifikacijo:

- funkcionalna podpora na strokovnih delavnicah,
- deluje kot enotna kontaktna točka (angl. *single-point-of-contact*, v nadaljevanju SPOC) za funkcionalna vprašanja DDQM in/ali lokalnih poslovnih strokovnjakov,
- pošilja DDQM funkcionalni zahtevek za spremembo specifikacije, ki temelji na interni potrebi podružnice.

4.4.2.2.5 Lokalni upravljelec kakovosti podatkov

Lokalni upravljelec kakovosti podatkov (angl. *local data quality manager*, v nadaljevanju LDQM) zagotavlja izvajanje lokalnih nalog za zagotavljanje kakovostnih podatkov v posamezni podružnici. Prevzema procesno usklajevalno odgovornost, pri čemer je končna odgovornost za trajno zagotavljanje zahtevane kakovosti podatkov na pristojnem managementu. Na ravni delovnih mest je potrebno zagotoviti tehnično ter vsebinsko odgovornost za popoln in pravilen vnos podatkov v informacijske sisteme ter pripadajoče odgovornosti dokumentirani v smernicah. LDQM je ekskluzivni partner DDQM in ITDQM za vprašanja o kakovosti podatkov. Njegove ključne naloge, razdeljene po procesih, so:

1. Proces dostave:

- LDQM je disciplinsko odgovoren za pravočasno, tehnično, vsebinsko pravilno ter popolno dostavo podatkov,
- ročne popravke podatkov, ki so na voljo po koncu meseca in pred vsakdanjo dobavo (na primer ponovno ocenjevanje derivatov, vrednostnih papirjev, itd.),
- pravočasno poročanje DDQM o morebitnih (tehničnih) težavah, ki bi lahko povzročile zamudo pri dostavi podatkov ali znatno negativno odstopanje v kakovost podatkov,
- pravočasne odločitve in informiranje DDQM če je potrebna ponovna dostava podatkov;

2. Proces spremljanja in eskalacije:

- lokalno izvajanje in pripombe na *ex-post* ocene kakovosti podatkov v posameznih podružnicah (predvsem celotni test) in pravočasno obveščanje ustreznega poslovnega oddelka (centralnega) in DDQM,
- oblikovanje akcijskega načrta za lokalni IT in lokalna poslovna področja zaradi popravljanja napak, razdelitev teh napak odgovornim osebam za spremembe/popravek v informacijskem sistemu banke in pozneje spremljanje mesečnega procesa. Pisno poročanje DDQM o napredku,
- spremljanje in analiza statistike dnevnega in mesečnega pošiljanja podatkov s pomočjo ustreznih poročil, ki so na voljo v Business Objectsu (poročila o ratingih, poročila o komitentih, pregled zavarovanj, RiWa podatki),
- mesečno lokalno poročanje o statusu kakovosti podatkov upravi banke in DDQM temelji na statusu kakovosti podatkov,
- deluje kot SPOC za DDQM, ITDQM in lokalne poslovne strokovnjake, vezano na vse lokalne zadeve o kakovosti podatkov,
- zbiranje vseh ugotovljenih operativnih tveganj, vezano na zagotavljanje kakovostnih podatkov,
- poročanje o operativnih tveganjih lokalnem oddelku za upravljanje s tveganji;

3. Proces upravljanja s specifikacijo:

- udeležba na strokovnih delavnicah, ki jih organizira koncern,
- mesečno poročanje DDQM o napredku izvajanja specifikacije,
- pravočasno pošlje DDQM funkcionalni zahtevek za spremembo specifikacije, ki temelji na interni potrebi podružnice.

4.4.2.2.6 Lokalni poslovni strokovnjak

Lokalni poslovni strokovnjaki so lastniki lokalnih podatkov in so implicitno odgovorni za lokalno kakovost podatkov v informacijskih sistemih bank in drugih ustreznih sistemih. Lastnik podatkov je subjekt (entiteta), ki lahko dovoli ali zavrne dostop do nekaterih podatkov in je odgovoren za njihovo točnost, celovitost in pravočasnost. Tako kot obstaja več kot eno poslovno področje, obstaja tudi več lastnikov podatkov. Lokalni poslovni strokovnjaki tesno sodelujejo z LDQM. Lokalni poslovni strokovnjaki podpirajo LDQM in ITDQM na naslednjih področjih:

- opredelitev lokalnih ključnih dejavnikov uspeha – KDU,
- izvedba lokalnega okvira za upravljanje s kakovostjo podatkov,
- proaktivno odkrivanje slabe kakovosti podatkov in informiranje LDQM,
- vzpostavitev lokalnega procesa za tekoče upravljanje s kakovostjo podatkov,
- stalno vrednotenje lokalnega procesa za upravljanje s kakovostjo podatkov,
- podpora LDQM tekom testiranja in prehoda v produkcijsko okolje na projektih, vezanih za kakovost podatkov (specifikacije, ...).

Njihove ključne naloge, razdeljene po procesih, so:

1. Proces spremljanja in eskalacije:

- komentar na *ex-post* ocene kakovosti podatkov,
- podpora LDQM pri odločanju o potrebi lokalnega akcijskega načrta,
- funkcionalna podpora LDQM pri analizi vzroka in oblikovanju akcijskega načrta za popraviljanje napake,
- poročanje LDQM o ugotovljenem operativnem tveganju, vezano na zagotavljanje kakovostnih podatkov;

2. Proces upravljanja s specifikacijo:

- naloge povezane z lokalnimi strokovnimi delavnicami,
- podpora lokalnem IT-u, ki je vključen v lokalno izvajanje SPEC projekta (gap analiza, ETL, zbiranje podatkov, itd.),
- sdeluje kot SPOC za LDQM,
- pravočasno pošlje LDQM funkcionalni zahtevek za SPEC spremembo, ki temelji na interni potrebi podružnice.

4.4.2.2.7 Lokalni IT upravljalec kakovosti podatkov

Lokalni ITDQM je odgovoren za tehnično uspešnost dostave podatkov v CBP. Njegova odgovornost se začne z ekstrakcijo podatkov iz lokalnih informacijskih sistemov (podatkovno skladišče ali centralni informacijski sistem banke) v aplikacijo za procesiranje podatkov za Basel II ter pravočasno dostavo kreirane datoteke na MITC. Da bi nadzirali tok informacij (eskalacija, stroški, vprašanja ali zahtevki, vezani na operativno tveganje), lokalni ITDQM neposredno komunicira z DDQM. Neposredna komunikacijo z MIRC ali MITP je dovoljena s pisnim dovoljenjem DDQM. Njegove ključne naloge, razdeljene po procesih, so:

1. Proces dostave:

- deluje kot SPOC za LDQM, z vsemi tehničnimi zadevami povezanimi z zagotavljanjem kakovostnih podatkov,
- namestitvev, posodobitev, konfiguracija ter delovanje aplikacije za procesiranje podatkov za Basel II,
- pripravljanje datoteke za dostavo,
- spremembe in popravki v tehničnih nastavitvah pošiljanja podatkov,
- pravočasna ali ponovna dostava datoteke s podatki;

2. Proces spremljanja in eskalacije:

- komentiranje *ex-post* ocen tehnične kakovosti podatkov (statistika dostave - Faza 1b & Faza 2),
- popraviljanje napak v sodelovanju z LDQM in lokalnimi poslovnimi strokovnjaki,

- poročanje LDQM o vseh operativnih tveganjih, vezanih na zagotavljanje kakovostnih podatkov;

3. Proces upravljanja s specifikacijo: podpora lokalnem IT oddelku pri lokalnem izvajanju specifikacije.

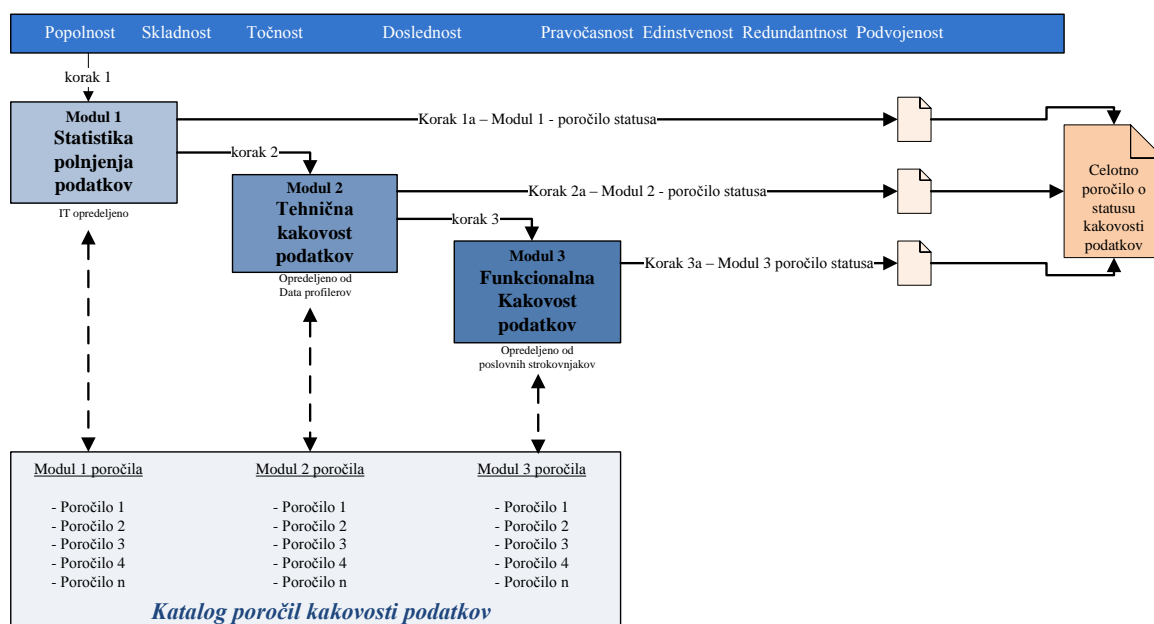
4.5 Poročanje o statusu kakovosti podatkov

Proces spremljanja kakovosti podatkov v definiranem okviru nam razen dejanskega spremljanja rezultatov procesiranja, posledično pomaga pri ovrednotenju uporabnosti okvira oziroma definiranju dodatne poslovne vrednosti uporabe okvira. Na podlagi končnih rezultatov poročil določimo, ali je okvir uporaben, in ali pomeni za banko, z njegovim uveljavljanjem, dodano vrednost.

Najvišji nivo spremljanja kakovosti podatkov je, lahko rečemo managerski pregled, poročanje o statusu kakovosti podatkov, zajem rezultatov vseh poročil, povezanih s kakovostjo podatkov.

Poročanje vsebuje tri module, rezultati modulov so odvisni in zaporedno vplivajo en na drugega. To pomeni, da delujejo z zaporednim vplivom na kakovost podatkov. Vsak modul ima svoja lastna poročila za različne teme, glede na njihovo funkcijo v modulu. Prednost zaporednega toka je, da lahko napovemo funkcionalno kakovost podatkov v zgodnji fazi, ne da bi izgubljali čas s čakanjem na končno poročilo o statusu funkcionalne kakovosti podatkov po čiščenju podatkov.

Slika 5: Poročanje o statusu kakovosti podatkov



Vir: Interno gradivo Poslovne banke X, 2009

Modul 1: poročila kažejo, ali dostavljeni podatki ustrezajo zahtevam algoritmov in funkcij:

- dostavljene datoteke – ali je datoteka dostavljena in koliko zapisov je dostavljeno,
- statistika polnjenja podatkov (Priloga 2):
 - tehnična preverjanja v fazi 1b – preverjanje kodne strani, naziva, kontrolnih zapisov,
 - preverjanje verodostojnost (SPEC obvezna polja, preverjanje hierarhije in ključev) CBP v fazi 2,
- neprenešene transakcije iz CBP v RiWA.





Modul 2: tehnična kakovost podatkov daje vpogled v veljavne, vendar ne nujno pravilne podatke. Če na primer niso pravilno napolnjena obvezna polja, zahtevana za združitev komitentov, ni mogoče uspešno zaključiti poenotenja, kljub razpoložljivosti podatkov. Drug primer je preveriti, ali so privzete (default) vrednosti pravilno dostavljene.

Modul 3: funkcionalna kakovost podatkov daje vpogled v kakovost podatkov in temelji na ključnih dejavnikih uspeha (KDU), opredeljenih s strani poslovnih strokovnjakov, kot:

- rating propusnost – stranke z Basel II izpostavljenostjo, vendar nimajo določenega ratinga, stranke z zastarelimi ratingi, brez ratingov,...
- klasifikacija tveganja podružnice in hčerinskega podjetja mora biti ista,
- segmentacija strank – nepravilna kombinacija segmenta stranke in metode ratinga stranke – $(\text{število strank z nepravilno kombinacijo} / \text{skupno število strank}) * 100 \%$,
- stranke s stanjem 0 ali neaktivne stranke,
- vodilni zapisi strank, preverjanje naslovov, datum rojstva, enolična identifikacija. natančnost je zelo pomembna zaradi uparjevanja meta strank,
- popolnost in pravilnost podatkov o zavarovanjih,
- združevanje meta strank (določanje strank na nivoju koncerna – gre za določanje istih strank, ki se večkrat pojavljajo v bankah); število interno prepisanih ratingov ter število interno prepisanih segmentov,
- celoten test – deviacija v podatkih pri primerjavi podatkov iz RiWa in z bankino bilanco, poslane s strani kontrolinga banke. Podatki se preverijo glede popolnosti z uporabo celotnega testa na ravni posla.

Status kakovosti podatkov je za vsak modul predstavljen z barvami na semaforju. Kot že omenjeno, če je status tehnične kakovosti podatkov rdeč, je funkcionalna kakovost podatkov tudi rdeča, ker ni mogoče meriti nobenega smiselnega kazalnika, če je tehnična kakovost podatkov slaba. Seveda je mogoče, da so nekateri KDU izmerjeni, ki nimajo veze s področjem tehničnega pregleda. Zato je v poročilu o statusu o kakovosti podatkov rezerviran stolpec za ponderiran faktor. Prag in pomen ponderja se opredeli na podlagi vpliva slabe kakovosti podatkov na rezultate RiWA faze.

























Statuse zelen, rumen, rdeč ali siv, opredelim kot:

- če je deviacija med 0 % – < 1 % status je zelen 
- če je deviacija med 1 % – < 2 % status je rumen 
- če je deviacija med 2 % – 100 % status je rdeč 
- če je deviacija med 2 % – 100 % – ampak je poslovno dovoljeno – status je siv 

Kot vidimo v Tabeli 15, je v stolpcu Modul opredeljeno ime ustreznega modula, stolpec Poročilo predstavlja naziv poročila, kot je shranjeno v sistemu poročanja, stolpec Status predstavlja trenutno stanje kakovosti podatkov in stolpec Status-1 predstavlja prejšnje stanje kakovosti podatkov. Stolpec Ponder označuje pomen poročil glede na negativen vpliv kakovosti podatkov za banke. Zadnji stolpec Incident se nanaša na številko incidenta v helpdesk sistemu kakovosti podatkov.

Rdeče stanje samodejno ne pomeni, da je napaka banke: pomeni, da kakovost podatkov te banke ni dovolj verodostojna za uporabo v pomembnih izkazih. Napako lahko povzroči sistem banke (informacijski sistemi bank, napake pri ETL-u,...), napaka v dokumentacija ali napačno mapiranje nekje v MITC med CBP in RiWa. Zato je pomembno zagotoviti pregleden vpogled v razloge in statuse o vprašanih kakovosti podatkov v helpdesk sistemu.

Tabela 15: Poročilo o statusu kakovosti podatkov

Modul	Poročilo	Status	Status-1	Ponder	Incident
Statistika polnjenja	Faza 1b statistika				
	Faza 2b statistika				
	Meta komitenti				INC0005468
	Faza 4 statistika				INC0005469
Tehnična kakovost podatkov	Obvezna polja za združitev komitentov				
	Saldo 0				INC0005458
	Default vrednosti				INC0005448
Funkcionalna kakovost podatkov	Portfolio Analysis				
	Segmentacijska matrika				
	Mirojoči računi				INC0004468
	SKD kode				INC0005498
	...				INC0005461

Vir: Interno gradivo Poslovne banke X, 2009

SKLEP

Dejstvo je, da so združbe prizadete z nizko kakovostjo podatkov, vendar če ta problem ne vpliva direktno na poslovanje (poslovne rezultate), je tipično, da se združba s tem sprijazni in zelo pogosto ne naredi nič v smeri izboljšanja kakovosti podatkov.

V magistrskem delu sem predstavil zagotavljanje kakovosti podatkov pri uvajanju kapitalskega sporazuma Basel II ter definiral okvir kakovosti podatkov, ki pomaga banki pri upravljanju s kakovostjo podatkov, namenjenim za pošiljanje v Basel II sistem.

Kot oseba, ki izhaja iz strogo tehničnih krogov, sem bil na začetku raziskovanja prepričan, da lahko z inštalacijo sofisticiranega softwarea uredim probleme s kakovostjo podatkov. Seveda sem se motil. Upravljanje podatkov zdaj dojemam kot nabor postopkov in politik, ki se aplicirajo med celotnim življenjskim ciklom podatkov, začenši s kreiranjem podatkov, pretvorbo podatkov v informacije, arhiviranje ali uničenje podatkov. Sestavljeno je iz dveh glavnih komponent: podatkov in spremljajoče infrastrukture. Za Basel II vključujemo podatke o komitentih, kreditih, depozitih, ratingih, zavarovanjih, vrednostnih papirjih, rezervacijah,... Infrastruktura vključuje vse procese in tehnologije, ki se uporabljajo za zbiranje, upravljanje in shranjevanje podatkov, kot so pretok podatkov, skladiščenje podatkov, metapodatki, kakovost podatkov, modeliranje podatkov, arhiviranje.

Za doseganje skladnosti z Baslom II, imajo pomembno vlogo zahteve po podatkih. Spomnimo se, da Basel II določa nabor navodil finančnim ustanovam za izračun kapitala, upoštevajoč tveganja. Da bi to lahko upoštevale, morajo banke pokazati, da imajo zahtevane podatke (do sedem let zgodovine) za izračun metrik tveganja, kot sta PD (verjetnost neplačila) in LGD (verjetna izguba ob neplačilu).

Poleg poudarka na podatkih pričakuje Basel II od združbe uveljavljanje robustne prakse za zagotavljanje kakovostnih podatkov. Ustvarjeni okvir za zagotavljanje kakovosti podatkov je politika, ki nam pri konkretni tematiki definira procese, profile delovnih mest, odgovornosti, časovnico, ovrednoti kakovost poslanih podatkov, pomaga pri identificiranju napak.

Banke, ki se pripravljajo na sprejem novih pravil, najbolj zanimajo konkretne številke o tem, kolikšne prihranke naj bi jim prinesla nova pravila v smislu zmanjševanja kapitala in kolikšne stroške bodo povzročili:

- izgradnja ali nadgradnja sistemov za upravljanje s tveganji,
- investicije v informacijsko tehnologijo,
- oblikovanje obsežnih podatkovnih baz,
- izobraževanja zaposlenih,

- preoblikovanja delovnih procesov.

Združbe se za odpravo nekakovostnih podatkov oziroma izboljšavo obstoječe kakovosti podatkov odločajo iz ekonomskih razlogov. Pri določanju ukrepov je treba pretehtati stroške, ki jih bodo ti ukrepi zahtevali, pa tudi koristi, ki jih bodo ukrepi prenesli. Vsaka banka sprejme tisto različico pristopa, ki je za njo najugodnejša, pri kateri so koristi, ki jih prinaša pristop, večje kot stroški.

Z uporabo definiranega okvira kakovosti podatkov se banka pripravlja na prehod iz standardiziranega pristopa izračuna kapitalske zahteve na napredni, IRB pristop. Da bi natančnejše izračunavala kapitalske zahteve kreditnih tveganj, mora biti skladna z IRB pristopom, kar pomeni, da bo zajela dejanska ekonomska tveganja posameznih naložb ali poslov, česar trenutno uporabljen standardizirani pristop ne omogoča. Ker pa IRB pristop temelji na internem rating sistemu razvrščanja komitentov, nam to omogoča diferenciacijo kreditnega tveganja in s tem posledično natančnejši izračun kapitalskih zahtev za kreditna tveganja.

Da bi bila banka skladna z IRB, so podatkovne zahteve naslednje:

- sposobnost kvantitativne ocene kakovosti svojih podatkov,
- učinkovita poslovna strategija čiščenja podatkov,
- ključne spremembe v poslovnih procesih s ciljem vzdrževanja zadovoljive kakovosti podatkov,
- okvir za merjenje in upravljanje s kakovostjo podatkov na trajni osnovi,
- zagotovljeni arhivski podatki.

Mnenja sem, da je namen magistrske naloge dosežen. Uspel sem namreč dvigniti zavedanje o pomembnosti kakovostnih podatkov in pokazati, da lahko problematika direktno ali indirektno vpliva na učinkovitost poslovanja, oziroma na skladnost z Baslom II. Posledično pa bo management seznanjen s stroški pri eventuelni uporabi nekakovostnih podatkov. Lahko zaključim s trditvijo, da sem približal pojem kakovosti podatkov vsakemu sodelujočemu v življenjskem ciklu podatkov ter jim vsadil dozo pomembnosti in odgovornosti.

Danes je kakovost podatkov izredno aktualno, pomembno in zelo hitro razvijajoče se področje. Sicer jo eni dojemajo samo kot trenutno »modno« muho in če se konkurenca s tem ukvarja, zakaj se ne bi tudi oni. Drugi pa z njo res izboljšujejo svoje poslovanje, zmanjšujejo stroške in postajajo bolj konkurenčni – boljši. Izboljšave v kakovosti podatkov so prišle z uvedbo podatkovnih skladišč in poslovne inteligence. Uporaba skladišč podatkov in poslovne inteligence je pripeljala do veliko boljšega razumevanja toka poslovnih podatkov in njihove uporabe. Posledično zdaj vse več združb gleda na kakovost podatkov kot na težavo celotne združbe in ne samo kot na težavo, ki jo je treba rešiti pri

gradnji podatkovnega skladišča, aplikacij poslovne inteligence in negativnih odgovorov regulatorjev pri upoštevanju novih zakonskih zahtev.

LITERATURA IN VIRI

1. Ambler, S. (2006). *Whence data quality?* Najdeno 15. oktobra 2008 na spletnem naslovu <http://www.ddj.com/dept/database/196900212?cid=Ambysoft>
2. Basel Committee on Banking Supervision (2001) *Overview of The New Basel Capital Accord*. Bank for International settlements. Najdeno 25. maj 2008 na spletnem naslovu www.bis.org/publ/bcbsca02.pdf
3. Baselski odbor za bančni nadzor (2000). *Načela upravljanja s kreditnim tveganjem*. Najdeno 11. aprila 2007 na spletnem naslovu <http://www.bsi.si/poslovanje-bank-in-podjetij.asp?MapaId=508>
4. Basel Committee on Banking Supervision (2001) *Working Paper on the Pillar 3 – Market Discipline*. Najdeno 25. maj 2008 na spletnem naslovu http://www.bis.org/publ/bcbs_wp7.pdf
5. Batini, C., & Scannapieca, M. (2006). *Data quality: Concepts, Methodologies and Techniques*. Berlin: Springer.
6. Batini, C., Cappiello, C., Francalanci, C. & Maurino A. (2009, julij). Methodologies for Data Quality Assessment and Improvement. *ACM Computing Surveys*, 41(3), 36-41.
7. Beloglavec, S., & Glogovšek, J. (2002). Basel II, velikost in tržna usmerjenost banke. *8. strokovno posvetovanje o bančništvu 2002(str.75–95)*. Portorož: Zveza ekonomistov Slovenije.
8. Bessis, J. (2002). *Risk Management in Banking*. Chichester: John Wiley & Sons, Ltd.
9. Bauer, K. (2004, december). *The power of metrics: KPIs – not all metrics are created equal*. Najdeno 16. novembra 2008 na spletnem naslovu <http://www.information-management.com/issues/20041201/1014522-1.html>
10. Bovee, M., Srivastava, R. P., & Mak, B. R. (2003, januar). A Conceptual Framework and Belief-Function Approach to Assessing Overall. *Information Quality International Journal of Intelligent Systems*, 18(1),51-74
11. Bowen, P., Fuhrer, D., & Guess, F. (1998, september). *Continuously improving data quality in persistent databases*. Najdeno 12. novembra 2009 na spletnem naslovu <http://www.dataquality.com/998bowen.htm>
12. Burns, L. (2005, maj). The ugly truth about data quality. *DMReview*. Najdeno 17. maja 2009 na spletnem naslovu http://www.dmreview.com/article_sub.cfm?articleId=1028545
13. Chapman, A. (2005, junij). Principles of data quality. *GBIF*. Najdeno 16. junija 2008 na spletnem naslovu <http://www2.gbif.org/DataQuality.pdf>
14. Chung, C. (2002, maj). The top five obstacles to achieving data quality. *InfoManagement direct*. Najdeno 9. decembra 2008 na spletnem naslovu <http://www.information-management.com/infodirect/20020503/5106-1.html>

15. Chung, W., Fisher, C., & Wang, R. (2002). What skills matter in data quality? *MIT*. Najdeno 9. decembra 2008 na spletnem naslovu http://web.mit.edu/tdqm/www/iqc/DQ_GST_at_ICIQ.pdf
16. Deming, W. E. (1986). *Out of the crisis*. Cambridge: Massachusetts Institute of Technology.
17. Dimovski, V., & Gregorič, A. (2000) *Temelji bančništva*. Ljubljana: Ekonomska fakulteta.
18. Dravis, F. (2003). A simple case of ROI. *Information Management Magazine*. Najdeno 23. julija 2007 na spletnem naslovu <http://www.information-management.com/infodirect/20030314/6477-1.html>
19. English, L. (1999). *Improving data warehouse and business information quality: methods for reducing costs and increasing profits*. New York: Wiley.
20. English, L. (2002). The essentials of information quality management. *Information Management Magazine*. Najdeno 12. oktobra 2008 na spletnem naslovu <http://www.information-management.com/issues/20020901/5690-1.html>
21. English, L. (2004). Total Information Quality Management – A Complete Methodology for IQ Management. *Information Management Magazine*. Najdeno 10. junija 2007 na spletnem naslovu <http://www.information-management.com/issues/20030901/7320-1.html>
22. English, L. (2006). Information quality and reference data. *Information Management Magazine*. Najdeno 22. septembra 2008 na spletnem naslovu <http://www.information-management.com/issues/20060301/1048494-1.html>
23. Eppler J. Martin (2003). *Managing Information Quality, Increasing the Value of Information in Knowledge-intensive Products and Processes*. Berlin: Springer.
24. Fisher, C., Lauria, E., & Matheus, C. (2007). *MIT*. In search of an accuracy metric. Najdeno 22. septembra 2008 na spletnem naslovu <http://mitiq.mit.edu/iciq/PDF/IN%20SEARCH%20OF%20AN%20ACCURACY%20METRIC.pdf>
25. Fuchs, G. (2002, november). Practical tips to analyze your data quality. *Information Management Magazine*. Najdeno 11. julija 2007 na spletnem naslovu <http://www.information-management.com/specialreports/20021119/6055-1.html>
26. Gable, G. (1994). Integrating case study and survey research methods: an example in information systems. *European Journal of Information Systems*, 3(2), 112–126.
27. Geiger, J. (2004). Data quality management: the most critical initiative you can implement. *SAS*. Najdeno 19. novebra 2008 na spletnem naslovu <http://www2.sas.com/proceedings/sugi29/098-29.pdf>
28. Goasdoué, V., Nugier, S., Duquennoy, D., & Laboisie, B. (2007). An evaluation framework for data quality tools. *MIT*. Najdeno 11. novebra 2009 na spletnem naslovu <http://mitiq.mit.edu/iciq/PDF/AN%20EVALUATION%20FRAMEWORK%20FOR%20DATA%20QUALITY%20TOOLS.pdf>
29. Godnov, U. (2009). *Vpliv implementacije relacijskega podatkovnega modela na kakovost podatkov v poslovnih informacijskih sistemih* (doktorska disertacija). Ljubljana: Ekonomska fakulteta.

30. Hammer, M., & Champy, J. (1993). *Reengineering the Corporation: A Manifesto for Business Revolution*. New York: Harper Business.
31. Hernandez, M. A., & Stolfo, S. J. (1998). Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. *Journal of Data Mining and Knowledge Discovery 1*.
32. Hipp, J., Guntzer, U., & Grimmer, U. (2001). Data quality mining: making a virtue of necessity. *Cornell*. Najdeno 11. novebra 2009 na spletnem naslovu http://www.cs.cornell.edu/johannes/papers/dmkd2001-papers/p5_hipp.pdf
33. Hudicka, J. (2003, marec). Develop a data quality strategy before implementing a data warehouse. *Information Management Magazine*. Najdeno 30. maja 2009 na spletnem naslovu <http://www.information-management.com/issues/20030301/6404-1.html>
34. Imai, M. (1986). *Kaizen (Ky'zen): the key to Japanese competitive success*. New York: Random House Business Division.
35. Inmon, W. H., Welch, J. D., & Glassey, K. L. (1997). *Managing the data warehouse*. New York: Wiley.
36. Jarke, M., Lenzerini, M., Vassiliou, Y., & Vassiliadis, P. (1995). *Fundamentals of Data Warehouses*. Springer Verlag
37. Jarke, M., Quix, C., Blee, G., Lehmann D., Michalk, G., & Stierl, S. (1999). Improving OLTP Data Quality Using Data Warehouse Mechanisms. *Sigmond Record*, 28(2), 536–537.
38. Karacsony, K. (2006). Proactive data quality. *Information Management Magazine*. Najdeno 13. maja 2009 na spletnem naslovu <http://www.information-management.com/infodirect/20060224/1048865-1.html>
39. Kettinger, W. J., & Grover, V. (1995). Towards and Theory of Business Process Change Management, *Journal of Management Information Systems*, 12(1), 9-30.
40. Lee, Y., Pierce, E., Talburt, J., Wang, R., & Zhu, H. (2007). A curriculum for a master of science in information quality. *Journal of Information Systems Education*, 18(2), 233–242.
41. Lee, Y., Strong, D., Kahn, B., & Wang, R. (2002). AIMQ: a methodology for information quality assessment. *Information&Management*, 40(2), 133–146.
42. Lee, Y. W. (2006). *Journey to data quality*. Cambridge: MIT Press.
43. Loshin, D. (2001). *Enterprise knowledge management: the data quality approach*. San Diego: Morgan Kaufmann.
44. Loshin, D. (2004). Issues and opportunities in data quality management coordination. *Information Management Magazine*. Najdeno 3. avgusta 2008 na spletnem naslovu <http://www.information-management.com/issues/20040401/1000836-1.html>
45. Loshin, D. (2005). Developing information quality metrics. *Information Management Magazine*. Najdeno 10. junija 2006 na spletnem naslovu <http://www.information-management.com/issues/20050501/1026061-1.html>
46. Loshin, D. (2006). The data quality business case: projecting return on investment. *Informatica*. Najdeno 16. februarja 2009 na spletnem naslovu http://i.i.com.com/cnwk.Ld/html/itp/informatica_Data_Quality_Business_Case.pdf

47. Marshall, B. (2007). Data quality and data profiling. *Telusplanet*. Najdeno 13. januarja 2009 na spletnem naslovu <http://www.telusplanet.net/public/bmarshall/dataqual.htm#DATAQUALITY%20PROFILING>
48. Maydanchik, A. (2007). *Data quality assessment*. New Jersey: Technics Publications.
49. McCue, A. (2006, maj). Poor-quality data is biggest CIO headache. *ZDNet*. Najdeno 16. marca 2008 na spletnem naslovu <http://news.zdnet.co.uk/itmanagement/0,100000308,39267945,00.htm>
50. Mullen, N. (2003, januar). Information for innovation: Less Bitter – improving the quality of your data. *Information Management Magazine*. Najdeno 7. julija 2008 na spletnem naslovu <http://www.information-management.com/issues/20030101/6194-1.html>
51. Muthu, S., Withman, L. & Cheraghi, S. (1999). Business Process Reengineering: a Consolidated Methodology. *TWsu*. Najdeno 4. aprila 2007 na spletnem naslovu <http://webs.twsu.edu/whitman/papers/ijii99muthu.pdf>
52. Neely, P. (1998). Data quality tools for data warehousing: A small sample survey. *Center for Technology in Government University at Albany*. Najdeno 17. maja 2007 na spletnem naslovu http://www.ctg.albany.edu/publications/reports/data_quality_tools/data_quality_tools.pdf
53. Ogrinc, P. (2004, november). Ključne ugotovitve slovenske kvantitativne študije učinkov novih kapitalskih pravil. *Banka Slovenije*. Najdeno 20. Februarja 2007 na spletnem naslovu <http://www.bsi.si/library/includes/datoteka.asp?DatotekaId=499>
54. Olson, J. E. (2003). *Data quality the accuracy dimension*. San Francisco: Morgan Kaufmann.
55. Redman, T. (1996). *Data quality for the information age*. Boston: Artech House.
56. Redman, T. (2001). *Data Quality: The field guide*. Boston: Digital Press.
57. Redman, T. (2004a). Barriers to successful data quality management. *Studies in Communication Sciences*, 4(2), 53–68.
58. Redman, T. (2004b). Data: an unfolding quality disaster. *Information Management Magazine*. Najdeno 10. junija 2006 na spletnem naslovu <http://www.information-management.com/issues/20040801/1007211-1.html>
59. Rotovnik T. (2004). Proces nove evropske kapitalske ureditve in njene bistvene novosti. *Bančni vestnik*, 2004(10), 43–47.
60. Rotovnik T. (2006). Odobravanje naprednih modelov za izračun kapitalskih zahtev. *Bančni vestnik*, 2006(1-2), 66-70.
61. Rubin S. (2002, december). Na poti do Basla II. *Bančni vestnik*, 2003(6.), 41–44.
62. Scannapieca, M., Missier, P., & Batini, C. (2005). Data quality at a glance. *DIS*. Najdeno 13. julija 2007 na spletnem naslovu <http://www.dis.uniroma1.it/~monscan/ResearchActivity/Articoli/DS2005.pdf>
63. Secretariat of the Basel Committee on Banking Supervision (2001) The New Basel Capital Accord: an explanatory note. *Bank for International settlements*. Najdeno 25. maja 2008 na spletnem naslovu <http://www.bis.org/publ/bcbsca01.pdf>

64. Shankaranarayan, G., Wang, R.Y., & Ziad, M. (2000). *Modeling the manufacture of an information product with IP-MAP*. In Proceedings of the 6th International Conference on Information Quality (ICIQ 2000). Boston.
65. Silvers, F. (2006a). Deming, data quality and ETL, part 1: point 3 – cease dependence on inspection. *Information Management Magazine*. Najdeno 12. maja 2007 na spletnem naslovu <http://www.information-management.com/infodirect/20060203/1047156-1.html>
66. Silvers, F. (2006b). Deming, data quality and ETL, part 2: point 5 – constant improvement. *Information Management Magazine*. Najdeno 12. maja 2007 na spletnem naslovu <http://www.information-management.com/infodirect/20060210/1047511-1.html>
67. Stordel, H., & Cross, A. (2002, junij). A cost/benefit approach to Basel II. *Risk.net*. Najdeno 19. septembra 2008 na spletnem naslovu <http://www.risk.net/operational-risk-and-regulation/feature/1525329/a-cost-benefit-approach-basel-ii>
68. Tajnikar, E. (2009). *Pripravljenost na Basel II v slovenskem lizinskih sektorju* (magistrsko delo). Ljubljana: Ekonomska fakulteta.
69. Tierstein, L. (2004). A methodology for data cleansing and conversion. *WRSystems*. Najdeno 16. avgusta 2007 na spletnem naslovu <http://www.wrsystems.com/whitepapers/dbclean.pdf>
70. Zavodnik E., & Šušterič, M. (2004). Priprava Slovenskih bank na spremembe kapitalskega sporazuma – Basel II. *Združenja bank Slovenije*. Najdeno 10. aprila 2008 na spletnem naslovu <http://www.bsi.si/library/includes/datoteka.asp?DatotekaId=507>

PRILOGE

KAZALO PRILOG

Priloga 1: SPEC vrste zapisov.....	1
Priloga 2: Statistika polnjenja podatkov – faza 2.....	2
Priloga 3: Seznam uporabljenih kratic	4

Priloga 1: SPEC vrste zapisov

Realiziran prihodek za SAP	BEL
Poroštva, garancije	BU
Cap/floor/collar	CAP
Skupni stroški – SAP	COB
Derivativi	DER
Kreditni konti	DK
Devizne opcije	DVO
Devizni terminski posli	DVT
Lastnik	EIG
Zaključni stavek	ES
Oslabitve / popravki vrednosti	EWB
Denarne vloge	FG
Pogoji denarnih vlog	FGK
Dogovor o obrestni meri forwardov	FRA
Limiti na TRR, revolving	GI
Struktura hipotek	GST
Obveznosti (pasiva) - poroštva, garancije, izvenbilanca	HG
Začetni stavek	HS
Podatki o strankah	KD
Pogoji kreditov	KKD
Lokalni koncern strank	KZL
Obremenjenost nepremičnin	LAS
Podatki o nepremičnini	LIE
Perioda vrednotenja	PW
Perioda vrednotenja za storitve	PWD
Rating	RAT
Bilančni konti (aktiva in pasiva)	SA
Udeležba	SAB
Saldo	SAL
Hranilne vloge z variabilno o.m.	SP
Swapi	SW
Pogoji swap-ov	SWK
Plačila swap-ov	SWZ
Linije zakladništva	TRL
Podatek o prometu	UMS
Vloge - fiksna o.m., obračun obresti ob zapadlosti, bullet payment	VL
Zastava denarja in zlata	VPB
Zastava polic	VPV
Zastava vrednostnih papirjev	VPW
Podatki o emitentih	WPE
Pogoji vrednostnih papirjev	WPK
Nostro vrednostni papirji (v lasti banke) - aktiva in pasiva	WPN
Pozicije vrednostnih papirjev	WPP
Podatki o emitentih	WPT
Tečaj vrednostnih papirjev	WPU
Porazdelitev hipoteke (1 hipoteka za več kreditov)	WST
Odstop terjatev	ZE

Priloga 2: Statistika polnjenja podatkov – faza 2

=====
SAP - SENTENCES (not Stage-2 processed!!!)
=====

Sentence	sum sent
KKD	22977
UMS	215
WPK	7
WPU	11
CAP	13
COB	35
DVO	16
DVT	4
PW	38451
PWD	27124
SW	10
SWK	20
SWZ	163
BEL	7135

=====

This is the result of stage 2:

SENTENCETYPE: WPN

sum sent:	16
sum OK:	16
sum NOK:	0
difference to last month	4
percentage NOK:	0,00
percentage difference:	33,33

SENTENCETYPE: WPE

sum sent:	80
sum OK:	80
sum NOK:	0
difference to last month	2-
percentage NOK:	0,00
percentage difference:	2,43-

..

SENTENCETYPE: UKU

sum sent:	33
sum OK:	33
sum NOK:	0
difference to last month	0
percentage NOK:	0,00
percentage difference:	0,00

===== S U M P E R F I L E T Y P E =====

Sum Filetype: 1

sum sent:	216301
sum OK:	215719
sum NOK:	582
difference to last month	3247
percentage NOK:	0,26
percentage difference:	1,52

Natančen opis napak v podatkih

Datum in ura imp.;Oznaka banke;Datum
podatkov;DBFaza;Verzija;Faza;VrstaDatoteke;VrstaZapisa;PovezanaVrstaZapisa;Polje;Pravilo;Instanca;VrstaN
apake;Stevilo;Napaka;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;KD;;DATBETRGR;21;33;ORDINARY;1;14532;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;KD;;DATUMSATZKMU;21;34;ORDINARY;1;14532;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;KD;;KUNDNr;6;1;ORDINARY;1;14532;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;KD;;NAME1;2;140;ORDINARY;8;14532;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;ED;;KONTOAUSL;2;44;ORDINARY;5;14532;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;EVF;;KONTOAUSL;2;70;ORDINARY;11;14532;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;AS;;KUNDNr_KR;2;8;ORDINARY;210;14532;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;GI;;KUNDNr;6;11;ORDINARY;1;14532;

06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SA;SA;;0;0;DUPLICATE;1;14530;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;EWB;EWB;;0;0;DUPLICATE;1;14530;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;EWB;;EWBPRODTYP;5;267;ORDINARY;7;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SAL;;UEZZAEHLER;15;1;ORDINARY;29;14542;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SAL;HG;;1;1;PARENT;2;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SAL;FG;;1;1;PARENT;6;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SAL;GI;;1;1;PARENT;12;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SAL;SP;;1;1;PARENT;2;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SAL;SA;;1;1;PARENT;6;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;SAL;SAL;;0;0;DUPLICATE;1;14530;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;TRE;DK;;1;1;PARENT;27;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;TRE;GI;;1;1;PARENT;11;14500;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;VPV;DK;;31;6;PARENT;4;14507;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;VPV;GI;;31;6;PARENT;4;14507;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;VPV;VL;;31;6;PARENT;4;14507;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;VPV;LE;;31;6;PARENT;4;14507;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;VPV;WE;;31;6;PARENT;4;14507;
06-12-2010 22:49:37;97;2010-11-30-23.59.59.000000;P;11.00;2;1;VPV;HG;;31;6;PARENT;4;14507;

Opis napak

Številka napake	Tekst napake
14500	Obvezna polja nimajo vrednosti, ali je vrednost neustrezna.
14504	Napačen datum.
14505	Polje ni numerično.
14506	Napaka v hierarhiji! Niso poslane vrste zapisa (GI, DK, VL, WE, HG, LE).
14507	Napaka v hierarhiji! Nadrejeni zapis ni vključen.
14508	Napaka v hierarhiji! Niso poslane vrste zapisa (GI, DK,VL,WE, HG, LE, SA).
14529	Vrsta zapisa (na primer DK, KKD in SAL) ni popolna.
14530	Podvojen zapis – primarni ključ že obstaja!
14531	Napaka v hierarhiji! SIZ=>BU,VP,VPB,SS,VGH,VPV,ZE,VPW,EV,GBST.
14532	Obvezno polje je napačno.
14537	Neobvezno polje ni napolnjeno, ali je vrednost neustrezna.
14538	Neobvezno polje je neustrešno napolnjeno.
14539	Obvezna koda valute je napačna ali neustrezna.
14540	Neobvezna koda države je napačna ali neustrezna.
14542	Napačna vrednost v polju XXXXXXXX v SAL vrsti zapisa.
14543	Napaka v hierarhiji! Vsa obvezna polja niso izpolnjena.
14544	PRODID ni vključen ali KDU ni pomemben za CBP.
14545	Zapis za brisanje je poslan brez da je prej poslan zapis za kreiranje.

Priloga 3: Seznam uporabljenih kratic

AIMQ	A methodology for Information Quality assessment
AMEQ	Activity-based Measuring and Evaluating of product information Quality
BCBS	Basel Committee on Banking Supervision
BPR	Business Process Reengineering
CBP	Centralna Baza Podatkov
CDQ	Comprehensive methodology for Data Quality management
CDQM	Central Data Quality Manager
CIHI	Canadian Institute for Health Information methodology
CIS	Cooperative Information System
COLDQ	Cost-effect Of Low Data Quality
DaQuinCIS	Quality in Cooperative Information Systems
DDQM	Decentralizirani DQM
DDQM	Decentralized Data Quality Manager
DQA	Data Quality Assessment
DWQ	Datawarehouse Quality Methodology
EAD	Exposure At Default
EU	Evropska unija
IPMAP	Information Product Map
IQM	Information Quality Measurement
IRB	Internal ratings-based
IT	Information technology
KDU	Ključni Dejavnik Uspeha
LDQM	Local Data Quality Manager
LGD	Loss Given Default
MITC	Mednarodni IT Center
MITP	Mednarodna IT Podpora
OLTP	OnLine Transaction Processing
OLAP	OnLine Transaction Processing
P2P	Peer-to-Peer information system
PD	Probability of Default
RIWA	Risk Weighted Assets
QAFD	Quality Assessment of Financial Data
SPEC	Dostavna specifikacija
SPOC	Single Point Of Contact
TDQM	Total Data Quality Management
TIQM	Total Information Quality Management
WIS	Web Information System