

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

DIPLOMSKO DELO

**ZAGOTAVLJANJE KAKOVOSTI
INFORMACIJ PRI PODATKOVNEM
SKLADIŠČENJU**

Ljubljana, julij 2009

ERIKA ŽEFRAN

IZJAVA O AVTORSTVU

Študentka Erika Žefran izjavljam, da sem avtorica tega diplomskega dela, ki sem ga napisala pod mentorstvom izr. prof. dr. Jurij Jaklič in dovolim objavo diplomskega dela na fakultetnih spletnih straneh.

V Ljubljani, junij 2009

Podpis:

Kazalo

Uvod.....	1
1 Podatkovno skladišče	2
1.1 Opredelitev podatkovnega skladišča.....	2
1.2 Izgradnja podatkovnega skladišča.....	4
1.2.1 Glavne pridobitve podjetja z uvedbo podatkovnega skladišča.....	4
1.2.2 Arhitektura podatkovnega skladišča.....	6
1.3 Razlika med analitično in operativno uporabo podatkov iz podatkovnega skladišča..	9
1.4 Cilji, ki naj bi jih dosegalo podatkovno skladišče.....	10
1.5 Koristi in pomanjkljivosti podatkovnega skladišča	11
1.5.1 Koristi podatkovnega skladišča	11
1.5.2 Pomanjkljivosti podatkovnega skladišča	12
1.6 Vrednost podatkovnega skladišča	12
1.7 Polnjenje podatkovnega skladišča.....	13
1.8 Orodje za analizo podatkov v podatkovnem skladišču	15
2 Kakovost podatkov in informacije	17
2.1 Opredelitev podatka in informacije.....	17
2.2 Opredelitev kakovosti podatkov in informacije	18
2.3 Sodila kakovosti podatkov in informacij	19
2.3.1 Sodila kakovosti podatkov	19
2.3.2 Sodila kakovosti informacije.....	21
3 Zagotavljanje kakovosti	25
3.1 Dejavniki, ki vplivajo na kakovost podatkov.....	25
3.2 Napake v podatkih, ki se pojavljajo v podatkovnem skladišču	25
3.3 Priprava in čiščenje podatkov	27
3.4 Zagotavljanje kakovosti informacije.....	29
3.5 Proces celotnega upravljanja s kakovostjo informacije	29
3.5.1 Proces 6: Oblikovanje okolja za kakovostne informacije.....	30
3.5.2 Proces 1: Ocenjevanje kvalitete lastnosti pod. in informacijske arhitekture	31
3.5.3 Proces 2: Ocenjevanje kakovosti informacije.....	31
3.5.4 Proces 3: Ocenjevanje str., ki se pojavijo zaradi nekovostnih informacij	32
3.5.5 Proces 4: Reorganizacija in prečiščevanje podatkov	32
3.5.6 Proces 5: Izboljšava procesa pridobivanja kakovostnih informacij	33
3.6 Zaposleni in kakovost informacije	33
4 Vpliv kakovostnih in nekovostnih podatkov na poslovanje podjetja	34
4.1 Vplivi nekovostnih podatkov na poslovanje	35
4.2 Koristi kakovostnih podatkov	37
4.3 Vpliv kakovostnih informacij na odločanje v podjetju in njegovo uspešnost	37
Sklep.....	38
Literatura in viri	40

Kazalo slik

<i>Slika 1: Podatkovna skladišča, 2009</i>	2
<i>Slika 2: Centralizirana arhitektura</i>	7
<i>Slika 3: Distributivna arhitektura</i>	8
<i>Slika 4: Federativna arhitektura</i>	9
<i>Slika 5: Kocka - dimenzije kraja, časa in izdelka ter dejstva</i>	16
<i>Slika 6: Primerjanje izidov subjektivne in objektivne ocene</i>	21

Uvod

Danes je velik del uspeha podjetja odvisen od tega, kako kakovostne informacije ima na voljo. Zaradi tega je ta tema zelo aktualna, vendar pa je bila do sedaj nekoliko zanemarjena in postavljena na drugi tir. Kakovost informacije je tako zelo pomembno in hitro razvijajoče se področje, na katerem gradi vsako sodobno podjetje, ki hoče konkurirati in dosegati dobre rezultate.

Velika količina podatkov, s katerimi se soočajo podjetja, postane izziv. Tako se za povečanje uspešnosti v poslovanju začnejo graditi podatkovna skladišča, ki dajejo enostaven pregled nad podatki posameznega podjetja. Tako urejeni podatki pa omogočajo izdelavo analiz, opazovanje trendov, predvidevanja... Vendar pa je pri podatkovnem skladišču velik problem nekonsistentnost podatkov, saj ti prihajajo iz različnih sistemov (Jaklič, 2001, str. 3).

Izraza podatek in informacija se danes pogosto uporabljata kot sinonima. Vendar pa med njima obstaja precejšna pomenska razlika (Mohorič, 1999, str. 445-448). Podatek postane informacija šele takrat, ko ga postavimo v nek določen okvir, kjer dobi svoj smisel (Islovar, 2009). Vendar pa kakovost podatkov v podatkovnem skladišču sama po sebi ne zadošča. Iz podatkov moramo poiskati odgovore na manj ali bolj zapletena poslovna vprašanja, ki naj pripeljejo do boljših poslovnih odločitev. Z drugimi besedami, iz zbranih podatkov skušamo izluščiti informacije in znanje, ki nam pomagajo, da smo uspešni in konkurenčni.

Temeljni cilj moje diplomske naloge je predstavitev pomena in načina zagotavljanja kakovostne informacije pri podatkovnem skladiščenju.

V prvem poglavju diplomske naloge tako predstavim opredelitev podatkovnega skladišča, cilje, vrednosti, polnjenje podatkovnega skladišča ter njegove prednosti in slabosti. Podrobno opisujem tudi orodje OLAP (sprotno analitično procesiranje) za analizo podatkov v podatkovnem skladišču.

Drugo poglavje je namenjeno kakovosti podatkov in informacije, v katerem opisujem razliko med njima. Opredeljujem tudi kakovost podatka in informacije, sodila informacije, ocena in merjenje ter problemi, ki se pojavljajo pri kakovosti le-te.

V tretjem poglavju opisujem zagotavljanje kakovosti, v katerem je predstavljam način zagotovitve kakovosti informacije, dejavnike, ki vplivajo na kakovost informacije ter napake in čiščenje podatkov. Podrobno opisujem proces za izboljšanje in zagotavljanje kakovostnih informacij (TIQM) in kako zaposleni vplivajo na kakovost podatkov v podjetju.

Četrto poglavje namenjam opisu vplivanja kakovostnih in nekovostnih podatkov na poslovanje podjetja. Podrobno opisujem vpliv in stroški nekovostnih podatkov, koristi kakovostnih podatkov ter kako kakovostni podatki vplivajo na odločanje v podjetju in njegovo uspešnost.

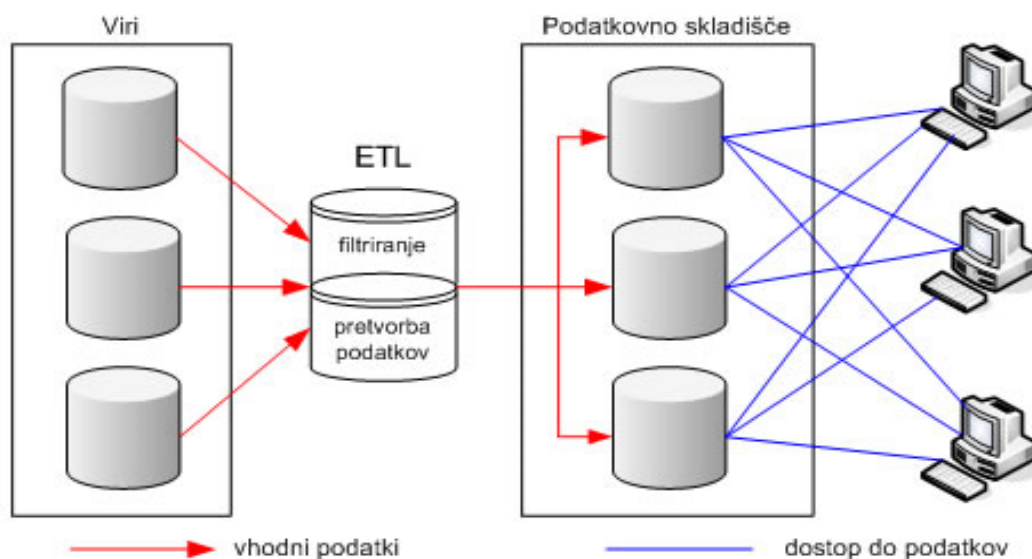
1 PODATKOVNO SKLADIŠČE

1.1 Opredelitev podatkovnega skladišča

Pri pregledovanju literature na temo podatkovnega skladišča (ang. *Data Warehouse*) sem zasledila veliko različnih definicij, ki ga opredeljujejo. Nekaterih od osnovnih definicij podatkovnega skladišča danes ne moremo več imeti za veljavna, saj se predvsem zaradi napredkov v strojni in programski opremi pojavljajo vedno nove arhitekture in rešitve. Ko so se pojavila podatkovna skladišča, so lahko ljudje zaposleni v podjetju končno videli podatke in informacije skupaj zbrane na enem mestu. Jaklič (2002, str. 18) pravi takole: »Zmotno je mnenje, da je podatkovno skladišče neko odlagališče neuporabnih podatkov, temveč rešuje problem, ki ga ima danes mnogo organizacij: goro podatkov, ki jih ne morejo uporabljati«.

Ralph Kimball (1996a, str. xxiii-xxv), eden izmed največjih strokovnjakov, ki se ukvarja s temo podatkovnega skladiščenja, je podal enostavno definicijo, ki pravi, da je podatkovno skladišče posnetek podatkov o transakcijah, ki so strukturirane posebej za povpraševanje in primerne za analizo podatkov. Tukaj gre za zelo drugačno podatkovno strukturo, ki je v svojem namenu posvečena popolnoma drugim ciljem, kot si običajno predstavljamo, kaj je naloga informacijskega sistema.

Slika 1: Arhitektura podatkovna skladišča



Vir: Podatkovna skladišča, Želite prehiteti konkurenco in povečati dobiček, 2008.

Podatkovno skladišče in njegova arhitektura, kot je prikazana na sliki 1, je samo po sebi velika zbirka časovno določenih podatkov za podporo poslovnim odločitvam. Vanj se zlivajo podatki iz vseh podatkovnih virov podjetja. Ko podatke organiziramo, lahko z njimi delamo razne analize in poizvedbe, ki drugače na lokalnih bazah niso mogoče.

Največkrat uporabljena in omenjena definicija pa je definicija W. H. Inmona, ki ga velika večina šteje za samega »izumitelja podatkovnega skladišča«. Inmon (1996, str. 33) podatkovno skladišče opredeljuje kot zbirko podatkov, ki je namenjena za podporo upravljavcem podatkovnih skladišč pri odločanju ter zanjo veljajo naslednja načela:

- predmetna usmerjenost oziroma organiziranost po posameznih področjih (*ang. subject oriented*),
- integriranost (*ang. integrated*),
- obstojnost (*ang. non volatile*),
- časovna spremenljivost (*ang. time variant*).

Predmetna orientiranost

Ponavadi transakcijski sistemi vsebujejo podatke o procesu, ki jih podpira. Uporabniki pa bi radi gledali podatke, ne glede na to iz katerih virov prihajajo, ampak gledano s poslovnega področja. Taka organiziranost omogoča, da imajo uporabniki boljši vpogled in razumevanje vsebine na poslovno področje.

Integriranost

Od vseh značilnosti podatkovnega skladišča je integriranost najpomembnejša, saj predstavlja enakost njihovega vsebinskega pomena in enakost oblike, čeprav pridejo podatki iz različnih aplikacij in različnih oddelkov. Podatki so iz različnih virov tako obdelani na način, da ustrezajo določenim pravilom. Zaradi tega pa, ker podatkovno skladišče nastaja v podjetju skozi daljše obdobje in ga tako oblikujejo različni ljudje, lahko pride do tega, da so isti podatki različno šifrirani, tako po vsebini kot tudi po obliki:

- Oblika: kadar imata dva podatka dva različna zapisa in ureditev (npr. telefonska številka, ki je lahko zapisana kot 040 444 444 ali (040) 444-444), moramo to združiti v en opis.
- Funkcija: kadar dva različna podatka poimenujeta isto zadevo z dvema različnima imenoma, ju združimo v eno poimenovanje.

Obstojnost

To pomeni, da ko je enkrat podatek shranjen v podatkovnem skladišču, je nespremenljiv in naj se ne bi več spreminjal.

Časovna spremenljivost

Podatkovno skladišče bi bilo neuporabno, če bi vsebovalo samo tekoče podatke in ne tudi zgodovine. Veliko analiz in poročil se dela ravno na tem, da se primerja med seboj prejšnje in sedanje stanje. Tako lahko na podlagi primerjave zgodovinskih in trenutnih podatkov delamo trende za vnaprej. Koliko časa se lahko podatki v podatkovnem skladišču hranijo, je odvisno od potreb, zadnje čase pa predvsem zakonodaje - področje varstva osebnih podatkov, ki na trgu strogo ureja, koliko časa se lahko kakšni podatki beležijo v podatkovnem skladišču.

Skladišče podatkov je tako osrednja zbirka podatkov za celotno podjetje, ki je namenjena shranjevanju in dostopanju podatkov za informiranje in odločanje, ter je ločena od operativnih sistemov. Rešuje probleme dostopanja do podatkov na ravni celotnega podjetja in odpravlja probleme neskladnosti med različnimi vrstami zapisov podatkov. Tehnično je osnovni cilj vseh podatkovnih skladišč **konsistentnost podatkov**. Velikokrat je problem ravno v tem, da podatki prihajajo iz različnih sistemov in so zaradi tega težko povezljivi, zelo pogosto pa tudi nekonsistentni (različne vrednosti pomensko enakih atributov v različnih sistemih). Zaradi tega je čiščenje podatkov najzahtevnejši in hkrati tudi najpomembnejši proces pri gradnji podatkovnega skladišča (Jaklič, 2001, str. 3).

Podatkovno skladišče tako predstavlja informacijsko središče celotnega podjetja za dostopanje do najrazličnejših vrst podatkov, pomembnih v procesu odločanja in izmenjave sporočil, ki so nujna za usklajeno odvijanje poslovnih procesov. Zagotavlja zanesljive in zapletene strukture podatkov, hkrati pa omogoča nosilec poslovnih odločitev hiter dostop do informacij. Podatkovna skladišča imajo potencial, ki omogoča podjetjem razumeti in koristno uporabiti podatke, ki so nastali ob izvajanju njihovega posla.

Tehnologija podatkovnih skladišč skrbno sledi sodobnim organizacijskim trendom. Glede na trend sploščevanja organizacije lahko tudi podatkovno skladišče omogoča globalno dostopnost do podatkov za vse ravni. Omogoča zbiranje znanja o poslovnih procesih, ter daje ključno podlago za prilagajanje poslovnega sistema vedno novim zahtevam globalne konkurence oziroma za novo nastale pritiske v poslovnem okolju (Debelak, 2002, str. 61).

1.2 Izgradnja podatkovnega skladišča

1.2.1 Glavne pridobitve podjetja z uvedbo podatkovnega skladišča

Ko načrtujemo podatkovno skladišče je odločilnega pomena prav pravilna presoja o tem, kateri podatki so zares potrebni za analiziranje in management. Pomembno je tudi, kako te podatke najbolje predstavimo. Zato je najbolje, če se manager in podjetnik aktivno vključita v projekt izdelave skladišča. Bistvo podatkovnega skladišča je, da imamo fleksibilen dostop do podatkov. Ko imamo tako vse podatke zbrane na enem mestu se nam čas, ki ga potrebujemo za dostop do podatkov, bistveno skrajša. Lahek in hiter dostop do podatkov pa vodi do večje produktivnosti in učinkovitosti podjetja (Jaklič, 2001, str. 3):

Podatki v podatkovnem skladišču so združeni iz različnih transakcijskih sistemov na enem mestu, očiščeni, poenoteni in po potrebi dopolnjeni.

Zato so lahko poročila in analize hitreje pripravljene, možnosti napak pa so manjše, kar je rezultat v višji kakovosti in usklajenosti analiz. Pogosto so podatki v posameznih transakcijskih sistemih neočiščeni, zato moramo pri polnjenju iz posameznega izvora v podatkovno skladišče predvideti tudi postopke zagotavljanja kakovosti podatkov. 'Umazani' podatki so zavajajoči, nepravilni, neveljavni, nepopolni ter neusklajeni podatki med različnimi izvornimi sistemi podjetja.

V podatkovnem skladišču lahko hranimo tudi podatke, pridobljene od drugih organizacij, in ne le podatkov iz lastnih transakcijskih sistemov.

Dodatne podatke lahko uporabimo za različne analize oz. poročanje (npr. zunanje bonitetne ocene vrednostnih papirjev oz. komitentov). Podatkovno skladišče nam omogoča, da se odločamo na podlagi bolj pravih in celostnih informacij.

Podatkovno skladišče hrani zgodovino podatkov.

Podatkovno skladišče hrani zgodovinske podatke, ki jih potrebujemo za analize, medtem ko transakcijski sistemi zgodovine ne hranijo ali pa je težje dostopna. Zato lahko iz podatkovnega skladišča pripravimo poročila takšna, kot smo jih v preteklosti posredovali regulatorju. Potrebno se je odločiti, za koliko let bomo hranili zgodovino podatkov. Za inicialno polnjenje se moramo odločiti, ali bomo zgodovino gradili od danes naprej ali bomo polnili tudi podatke za nazaj za določeno število let. Pri tem lahko naletimo na ovire zaradi sprememb v času kot so sprememba formata določenih podatkov, sprememba pomena določenih veljavnih vrednosti, sprememba organizacijske sheme idr.

Za podatkovno skladišče se uporabljajo podatkovni modeli (npr. zvezdna shema) in tehnologije, ki omogočajo hitre in lažje poizvedbe ter pripravo poročil.

Poizvedbe in priprava poročil iz podatkovnega skladišča ne obremenjujejo transakcijskega sistema. Z modeli podatke organiziramo tako, da so poslovnim uporabnikom razumljivi in intuitivni za uporabo, poizvedbe in priprava poročil pa je relativno enostavna. S tem je lahko tudi zahtevano znanje tehničnih skrbnikov podatkovnega skladišča nižje oziroma priprava poizvedb in poročil zahteva manj časa.

Poleg tega pa si lahko uporabniki sami skrajšajo čas, ki je potreben, da dobijo ustrezne informacije in se zato bolj ukvarjajo s samo vsebino in s problemom kako informacijo koristno uporabiti. Sprotno osveževanje podatkovnega skladišča pa tudi omogoča skrajšanje časa, ki poteče od nekega kritičnega dogodka do časa, ko je o tem obveščen manager in tako

lahko na dogodek ustrezno reagira, poleg tega pa lahko podatke in informacije za odločanje sedaj spremlja na dnevni ravni. Tukaj izkoristijo dobro priložnost, ki bi jim ob pomanjkanju informacij ušla.

Tudi v primeru, da podjetje zamenja transakcijski sistem, podatki v podatkovnem skladišču ostajajo (Milavec, 2008).

1.2.2 Arhitektura podatkovnega skladišča

Od tega, kako bomo zgradili naše podatkovno skladišče v podjetju, je zelo pomembna odločitev za katero arhitekturo se bomo odločili, saj ta določa podatkovni model, vlogo področnih skladišč ter sosledje korakov v razvojnem ciklu. Čeprav so podatkovna skladišča danes široko razširjena, ni enotnega mnenja katera razvojna metodologija je najboljša. Uporabljata se dva pristopa. Prvi je povezan z Billom Inmonom, ki je pionir na področju podatkovnega skladiščenja in predlaga uporabo pristopa od zgoraj navzdol (*ang. top – down*). Drugi pa je povezan z Ralphom Kimballom, visoko cenjenim svetovalcem in avtorjem številnih knjig s področja podatkovnega skladiščenja. Kimball predlaga uporabo pristopa od spodaj navzgor (*ang. bottom – up*). Oba, tako Kimball kot Inmon, sta za podatkovno skladišče določila ista načela, vendar pa se zelo razhajata pri tem, kako naj bo podatkovno skladišče zgrajeno. Največji zagovornik distributivne arhitekture je Kimball, medtem ko Inmon zagovarja centralizirano arhitekturo. Inmonov pristop je zaradi posledičnega daljšega časa vzpostavitve precej bolj tvegan.

Golob je arhitekture podatkovnih skladišč razvrstil v **tri glavne sklope** (Golob, 2001):

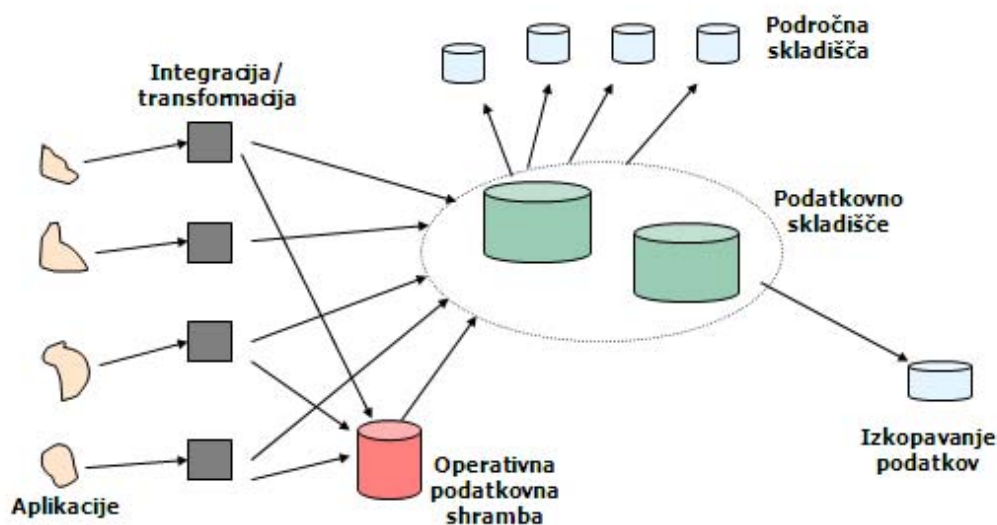
- centralizirana arhitektura ali pristop od zgoraj navzdol (Inmonov pristop),
- distributivna arhitektura ali pristop od spodaj navzgor (Kimballov pristop),
- federativna arhitektura ali hibridni pristop med obema zgoraj navedenima pristopoma.

Centralizirana arhitektura ali Inmonov pristop

V večini velikih organizacij je razvoj celovitega podatkovnega skladišča najbolj željen končni cilj. Centralizirana arhitektura (*ang. Corporate Information Factory*), kot prikazuje slika 2, je v začetku 80-ih let svojo pot začela z arhitekturo, ki je v središče postavila podatkovna skladišča in njim pripadajoča področna podatkovna skladišča. Kmalu zatem so v sistem vstopili meta podatki, ki so prevzeli vidno vlogo kot povezovalni faktor vseh komponent te arhitekture. V sredini 90-ih let se je v arhitekturo vključila operativna shramba podatkov, kar je predstavljalo začetek podpore taktičnim odločitvam. Pri Inmonovem pristopu se uporablja usklajeni zajem podatkov iz izvornih sistemov. V središču centralizirane arhitekture podatkovnega skladišča je podatkovno skladišče zaključenega organiziranega sistema, ki polni področna skladišča, samo pa se polni iz operativnih podatkovnih baz ter operativnega podatkovnega skladišča. Inmonova metodologija načrtovanja podatkovnega skladišča

generira visoko konsistentne dimenzijske poglede na podatke v področnih podatkovnih skladiščih, saj so ta napolnjena iz istega izvora – normaliziranega podatkovnega skladišča. Poleg tega je tako načrtovanje tudi odporno na poslovne spremembe. Glavna pomanjkljivost pristopa je, da predstavlja izredno obsežen projekt na obsežnem področju. Zahteva tudi velike denarne in časovne stroške preden imajo uporabniki koristi od podatkovnega skladišča (Inmon, 2009).

Slika 2: Centralizirana arhitektura



Vir: Jaklič, Upravljanje podatkov. Građiva pri predmetu Baze podatkov in njihova uporaba, 2003.

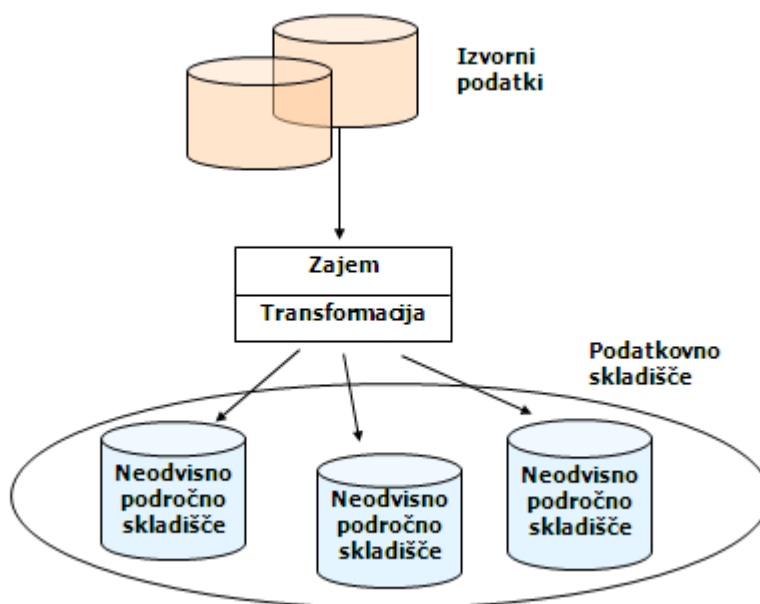
Distributivna arhitektura ali Kimballov pristop

Kimball v enem od svojih intervjujev poudari (Ferle, 2005, str. 5-7), da je Bill Inmon ustvaril ime podatkovno skladiščenje, njegov arhitekturni pristop pa se imenuje Corporate Information Factory (CIF). Na najvišjem nivoju sta si CIF in Kimballov pristop z enotnim vodilom (ang. *Data Warehouse Bus Architecture*) zelo podobna. Oba pridobivata, preoblikujeta in nalagata podatke v končne baze podatkov, do katerih dostopajo končni uporabniki. Vendar pa pod tem nivojem, Inmon ne zagotavlja tehničnih podrobnosti za CIF pristop. V tem se razlikuje njuno delo. Načrtovalcem podatkovnih skladišč Kimball zagotavlja specifične in podrobne smernice za reševanje najpogostejših težav podatkovnega skladiščenja vključno s čiščenjem podatkov, integracijo raznolikih virov podatkov ter ravnanjem s časovnim spreminjanjem značilnih entitet podatkovnega skladiščenja kot so uporabniki in izdelki.

Kimball priporoča pri graditvi podatkovnega skladišča razdelitev naloge na obvladljive podnaloge, katerih rešitve potem združujemo v celoto, kar je prikazano na sliki 3. Pri

planiranju razvoja podatkovnega skladišča moramo imeti v mislih arhitekturo, ki omogoča razvoj in implementacijo celovitega skladišča, vendar pa se moramo osredotočiti tudi na izpolnjevanje manjših, lažje rešljivih nalog, se pravi na razvoj posameznih področnih podatkovnih skladišč. Pri Kimballovem pristopu najprej zgradimo področno podatkovno skladišče, ki omogoča poročanje in analiziranje za specifične poslovne procese. Začetno podatkovno skladišče črpa podatke iz majhnega števila izvornih sistemov. Zaradi svojega omejenega obsega se lahko področno skladišče hitro zgradi po razmeroma nizki ceni in preskrbi hitro vračilo vloška. Če se izkaže za uspešno, dodajamo nova ali pa nadgrajujemo obstoječa. Proces izgradnje podatkovnega skladišča zajema postopno izgradnjo posameznih neodvisnih področnih skladišč, ki jih na koncu povežemo v logično podatkovno skladišče celotne organizacije. Taka skladišča so med seboj povezana preko medsebojno skladnih skupnih dimenzij in merljivih dejstev. Z uporabo le teh uporabniki lahko povprašujejo po vseh skladiščih hkrati. Priprava podatkov se začne z usklajenimi zajemi podatkov iz izvornih sistemov, ki jih prečistimo, medsebojno uskladimo, preoblikujemo in združimo. Tako pripravljene podatke napolnimo v predstavitveno področje.

Slika 3: Distributivna arhitektura



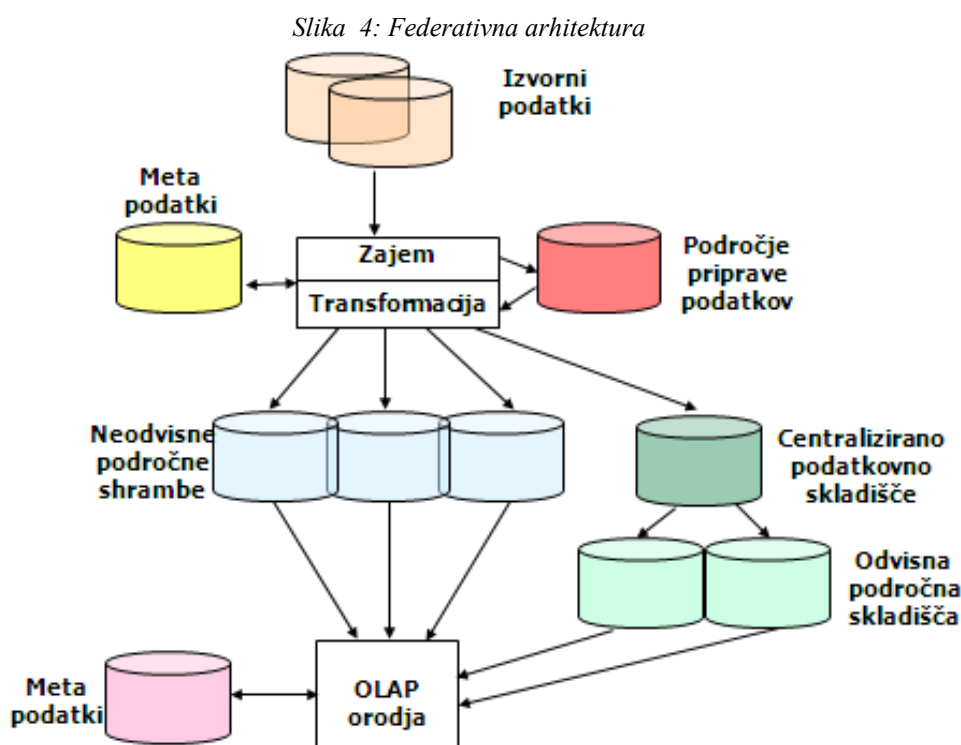
Vir: Jaklič, Upravljanje podatkov. Gradiva pri predmetu Baze podatkov in njihova uporaba, 2003.

Federativna arhitektura

Podatkovno skladišče, ki ima federativno arhitekturno strukturo je nekakšna hibridna rešitev med zgornjim dvema opisanimi arhitekturama in je prikazana na sliki 4. Centralizirano podatkovno skladišče, ki vsebuje podrejeno področna podatkovna skladišča in distributivna arhitektura, ki ima neodvisna področna skladišča.

Ta arhitektura odpravlja vse slabosti, ki naj bi jih imeli prejšnji arhitekturi. Med to spadajo velika vlaganja. Poleg tega pa tudi zagotavlja hitro povrnitev vloženih sredstev z uporabo neodvisnih področnih skladišč (Kimball & Ross, 2002).

Ena izmed največjih pomanjkljivosti te arhitekture je, da je še v fazi uveljavitve in da še ni sprejeta s strani proizvajalcev sistemov za podatkovno skladiščenje.



Vir: Jaklič, Upravljanje podatkov. Građiva pri predmetu Baze podatkov in njihova uporaba, 2003.

1.3 Razlika med analitično in operativno uporabo podatkov iz podatkovnega skladišča

Potrebe po podatkih na operativnem in na analitičnem nivoju se povsem razlikujejo. Za podatke, ki jih uporabljamo na operativnem nivoju je tako značilno (Jaklič, 2002, str. 16-17):

- da se podatki ne bi smeli podvajati, s čimer se izognemo težavam pri njihovem posodabljanju,
- uporabnike zanimajo predvsem trenutne vrednosti podatkov,
- dostop do podatkov je pogost, količina podatkov, s katerimi delamo je ponavadi majhna,
- način uporabe podatkov se le redko spreminja (ponavljajoči se vzorci uporabe),
- zanesljivost mora biti visoka, čas dostopa pa kratek.

Po drugi strani pa za uporabo na analitičnem nivoju velja naslednje:

- da podatkov večinoma ne posodabljam, zato podvajanje ne predstavlja problema,
- uporabnike zanimajo tudi zgodovinski podatki,
- dostop do podatkov je neenakomeren, količina podatkov, s katerimi ponavadi delamo, je velika, čeprav analitik dobi kot rezultat le nekaj sumarnih vrednosti,
- način uporabe podatkov je zelo raznolik,
- zanesljivost ni pomembna kot na operativnem nivoju, za zahtevnejše analize pa smo pripravljeni počakati tudi več časa.

Ker je za operativni nivo pomembno, da se podatki ne podvajajo, je struktura v operativnih bazah zapletena. Uporabniki tega sicer ne vidijo, ker do podatkov dostopajo skozi vnaprej pripravljene programe, ki strukturo skrijejo. Tako imamo različna orodja za analizo podatkov v podatkovnem skladišču, ki omogočajo uporabniku izdelavo poljubnih analiz podatkov.

1.4 Cilji, ki naj bi jih dosegalo podatkovno skladišče

Glavni cilj v podatkovnem skladišču mora biti odpravljanje težav v podjetju (Adelman, 2000, str. 37-47).

Cilje najlažje opredelimo na dolgoročne in kratkoročne. Kratkoročni morajo biti uresničeni takoj, ko se gradnja podatkovnega skladišča konča. Odpravljati pa morajo tiste najbolj pogoste napake, ki so veljali za tradicionalne sisteme za podporo odločanju. Medtem ko lahko dolgoročne cilje uresničimo šele takrat, ko so uresničeni kratkoročni cilji. Podatkovno skladišče nadgrajujemo tako, da sledi potrebam uporabnikov.

Nekateri izmed osnovni ciljev, ki jih mora izpolnjevati podatkovno skladišče so (Kimball, 1992a):

- Informacija mora biti vedno lahko dostopna uporabnikom podatkovnega skladišča. Njegova vsebina mora biti kar se da razumljiva in čitljiva. Prav tako morajo biti za uporabo enostavna tudi orodja za dostop do podatkovnega skladišča, ter morajo vračati rezultate povpraševanj v minimalnem času.
- Podatkovno skladišče mora predstavljati informacije dosledno. Podatki v njem morajo biti verodostojni in kakovostni preden jih predstavimo uporabnikom, da le –ti lahko začnejo delati na njih.

- Podatkovno skladišče mora biti prilagodljivo in prožno. Načrtovano je po navadi tako, da se lahko prilagaja spremenjenim uporabniškim potrebam, poslovnim okoliščinam, podatkom in tehnologijam, ki se jih uporablja.
- Podatkovno skladišče mora informacije ščititi. Imeti mora učinkovito kontrolo dostopov do zaupnih informacij.
- Služiti mora kot nekakšna osnova za izboljššan proces pri odločanju.
- Uporabniki morajo sprejeti podatkovno skladišče, saj le tako lahko dobro deluje. Uspešnost skladišča merimo glede na sprejemanje s strani uporabnikov.

Podatkovno skladišče mora biti predvsem tudi prilagojeno za uporabnike, ki potrebujejo podatke iz njega. Skladišče naj bi bilo dostopno čim večjemu številu končnih uporabnikov. Je pa res, da od njih ne moremo pričakovati, da bodo vsi ustrezno ravnali z njim. Tisti, ki so tehnično bolj podkovani, bodo uporabljali bolj razvita orodja za analizo, medtem ko bodo nekateri uporabniki zadovoljni že z uporabo razpredelnice in preprostih poizvedb, ki jih bodo lahko delali. Vse je odvisno tudi od tega, kolikšno količino podatkov in informacij bodo potrebovali.

1.5 Koristi in pomanjkljivosti podatkovnega skladišča

1.5.1 Koristi podatkovnega skladišča

Uporaba podatkovnih skladišč nam prinaša določene koristi:

- **Hitrejši in cenejši proces odločanja**

Za pripravo dokumentacije, poročil in statistik je potrebno manj ljudi in opreme, kompleksne poizvedbe pa ne obremenjujejo informacijskega sistema. Poveča se hitrost dostopa do vseh potrebnih informacij, hkrati pa ne trpi ostalo delo v podjetju.

- **Izboljšana kakovost informacij**

Podatkovna skladišča, ki vključujejo vse podatke, od transakcijskega do strateškega nivoja, zagotavljajo boljšo kvaliteto in večjo prilagodljivost analiz. Visoka natančnost pridobljenih informacij in zanesljivost rezultatov sta posledica kakovostnih podatkov, ki se skozi proces integracije in prečiščevanja v rednih časovnih presledkih prenašajo iz raznih baz podatkov v podatkovno skladišče podjetja.

- **Boljši odnosi s strankami**

S povezavo podatkov v podatkovnem skladišču je možno ugotoviti nove trende, zahteve in želje strank in jih s pravočasnim ukrepanjem obdržati ali pridobiti nove. Nadgradnja

podatkovnega skladišča se imenuje sistem za urejanje odnosov s strankami ali CRM. V tujini se veliko vloga v tovrstne sisteme, začetki takšnega razmišljanja pa se pojavljajo tudi pri nas. Osnovno podatkovno skladišče nudi sumarne podatke, medtem ko sistem za urejanje odnosov s strankami nudi tudi detajlni pogled v celotno komunikacijo s posamezno stranko.

- **Hitro ugotavljanje novih trendov**

Na osnovi obdelav podatkov in hitrih poizvedb je možno hitro priti do vseh kritičnih podatkov in ukrepati povsod tam, kjer poslovanje ni dobro. Z dnevno analizo je možno hitro ugotoviti nove trende, jim prilagoditi poslovno politiko in tako povečati dobiček ter prehiteti konkurenco. Podjetje se na tak način lahko izogne mnogim neljubim presenečenjem ob analizi poslovanja ob koncu poslovnega leta (Podatkovna skladišča, Želite prehiteti konkurenco in povečati dobiček, 2008).

1.5.2 Pomanjkljivosti podatkovnega skladišča

Pri uporabi podatkovnih skladišč pa se srečujemo tudi z naslednjimi pomanjkljivostmi (Berry, 2000, str. 145):

- lahko, da vseh virov, ki jih potrebujemo, ni na voljo,
- lahko niso uporabljeni vsi podatki,
- uporabljeni so lahko samo določeni nabori podatkov, za katere smo nekoč mislili, da so koristni, sedaj pa temu ni tako,
- podatki, ki zasedejo največ pomnilniškega prostora, morda niso skladiščeni.

1.6 Vrednost podatkovnega skladišča

Investicija v podatkovna skladišča se mora ponavadi povrniti v obdobju od treh do petih let. Ker je finančni vložek na začetku zelo velik, je to precej kratko obdobje, vendar pa izkušnje kažejo, da se učinek kakovostnih informacij pri vseh poslovnih informacijah pokaže skoraj takoj.

Posledice vpeljevanja podatkovnega skladišča so predvsem:

- boljše poznavanje trga, njegovih značilnosti in trendov,
- boljše znanje o navadah in potrebah strank ter
- pravi vpogled v dogajanje v podjetju.

Raziskovalno podjetje International Data Corporation je že leta 1996 naredilo raziskavo o povrnitvi investicij v podatkovna skladišča. V študiji so obravnavali 62 podatkovnih skladišč v ZDA, ugotovitve pa so presenetile celo same raziskovalce, saj je bila povprečna povrnitev investicije v obdobju dveh do treh let skoraj 400 odstotna. Drugo raziskavo je naredilo

podjetje NCR Corporation, ki je pri več kot 20 podatkovnih skladiščih ugotovilo od 300 do celo 1000 odstotno povrnitev investicije (Podatkovna skladišča, 2008).

1.7 Polnjenje podatkovnega skladišča

Poslovne potrebe so informacijske zahteve končnih uporabnikov, ki hočejo v podatkovnem skladišču uporabne in prave podatke za kasnejše odločanje pri pomembnih odločitvah (Kimball, 2004b, str. 4).

Podatkovnega skladišča s podatki ne polnimo ročno, ampak se jih prenaša iz različnih virov, ki pa so lahko poslovne aplikacije, zasnovane na različnih bazah podatkov, pisni dokumenti, tabele v Excelu ter različni zunanji podatki. Ko združujemo vse te podatke, moramo za vsak vir posebej izdelati vmesnike. Njihova vloga je, da pravilno zajemajo podatke, jih obdelujejo in preverjajo njihovo pravilnost ter podatke shranjujejo v skladišču. To se mora storiti v naprej določeni obliki, ki je enotna za vse podatke, ki se nahajajo v podatkovnem skladišču.

V podatkovno skladišče se podatki prenašajo iz dveh razlogov: da nam omogočajo lažje iskanje in analiziranje združenih podatkov iz različnih virov ter da nam zagotovijo kvalitetne rezultate povpraševanj in poročil. V podatkovno skladišče so podatki prineseni iz zunanjih virov (ne nastajajo v samem podatkovnem skladišču). Zato obstaja možnost, da se podatkom med prenosom v podatkovno skladišče zmanjša njihova kakovost.

Polnjenje podatkovnega skladišča je bistvenega pomena. Pri tem moramo podatke:

- **prečistiti** (pridobivanje podatkov iz zunanjih virov),
- **pretvoriti med različnimi podatkovnimi tipi** (pretvoriti znak v številko, besedilo v datum),
- **povezati skupaj z različnimi OLTP** (nalaganje podatkov v podatkovno skladišče).

Ta proces imenujemo **ETL** – Extract(ion), Transform(ation) & Load (pridobi, pretvori in naloži oziroma prenesi).

Pridobivanje podatkov iz zunanjih virov in njihovo čiščenje

Takrat, ko začnemo zbirati podatke iz sistemskih virov, ki so med seboj različni, pravimo da gre za prvo fazo ETL-a. Ker so sistemski viri ponavadi ločeni in vsak operira z različno strukturo in formatom podatkov, tako obstaja veliko podatkov, ki imajo različne formate, najpogosteje pa so v uporabi relacijske podatkovne baze in nepovezane datoteke. To delamo vsakodnevno, pri tem pa pazimo, da izvirnega sistema ne preobremenimo v času, ko je ta že tako ali tako preobremenjen z izvajanjem tekočih opravil.

Preoblikovanje podatkov

V okviru postopka za preoblikovanje podatkov v podatkovnem skladišču se moramo zavedati, da ko podatke dobimo iz izvornih sistemov, le-ti še niso primerni za nalaganje v podatkovno skladišče. Vse podatke, ki so zajeti, je potrebno preoblikovati na tak način, da bodo ustrezali tako poslovnim zahtevam kot tudi zahtevam skladišča. Nekatere podatke moramo pred nalaganjem obogatiti in oplemenititi, saj je kvaliteta teh podatkov zelo vprašljiva v starejših sistemih.

Preden se podatki v podatkovnem skladišču začnejo uporabljati, morajo iti skozi mnogo preoblikovanj, kjer nam veliko dela povzroči sam proces izboljševanja same kvalitete podatkov. Kako dobra oziroma slaba je kakovost podatkov, je vidno kasneje pri analizi in posledično sprejetih slabih ali dobrih poslovnih odločitvah. Vsako podjetje se mora zavedati kompleksnosti in zahtevnosti, čiščenja in preoblikovanja podatkov. Seveda nekateri podatkovni viri potrebujejo več obdelave podatkov, nekateri pa manj.

Glavne naloge preoblikovanja podatkov v podatkovnem skladišču so:

- selekcija,
- razcepitev/združevanje,
- pretvorba,
- seštevanje,
- oplemenitenje.

Nalaganje podatkov v podatkovno skladišče

Zadnja faza je nalaganje preoblikovanih in očiščenih podatkov v podatkovno skladišče. Kakšen je obseg tega procesa, je zelo odvisno od velikosti samega podjetja oziroma organizacije. Samo polnjenje v večini primerih zahteva veliko časa. Dobiti pa je potrebno tudi primerne časovne termine, ker mora biti podatkovno skladišče v času polnjenja nedostopno za ostale uporabnike. Zato to ponavadi to delamo v nočnih urah, ko uporabniki ne uporabljajo podatkovnega skladišča. Ravno tako pa moramo, zaradi časovno zamudnega postopka, dobro razmisliti o razdelitvi procesa. Polnimo lahko paralelno, tako da v podatkovno skladišče pošiljamo več manjših paketov podatkov. Del podatkovnega skladišča, ki je že napolnjen, pa lahko uporabljamo medtem, ko se nad ostalim delom izvaja proces polnjenja.

Proces je pozneje ključnega pomena zaradi uporabe tehnologije OLAP, saj se nad podatki, ki se v tem procesu prenašajo iz OLTP-ja v podatkovno skladišče, izvajajo analize, ki jih pripravijo uporabniki. Skratka, na tem mestu določimo, katere podatke bomo analizirali pozneje z OLAP-om, saj podatkov, ki jih v procesu ETL ne prenašamo, ne moremo vključevati pozneje v nobene analize. Najtežje delo je ravno na tem mestu, saj odločitev kaj,

kako in zakaj prenašati, velikokrat odloča o »usodi« podatkovnega skladišča. Podatki, še zlasti če vstopajo v proces ETL iz različnih virov, morajo biti po končanem procesu med seboj povezani oziroma se morajo med seboj pravilno povezovati, sicer nimajo pomena. Znano je, da ravno povezovanje podatkov iz različnih virov obsega približno 80 % celotnega projekta izdelave tipičnega podatkovnega skladišča, kar je pravzaprav logično. Veliko večino dela in časa porabimo ravno pri odločanju, urejanju in oblikovanju, katere podatke bomo prenašali in analizirali (Jagarinec, 2005, str. 78).

Sodobna ETL orodja nam omogočajo pomoč pri naslednjih fazah (ETL, 2008):

- priključitev na poljuben vir ne glede na izvor podatkov (poljubne podatkovne zbirke podatkov, poljubne poslovne aplikacije, nestrukturirani izvori – razne datoteke),
 - preverjanje kvalitete podatkov in združevanje podatkov iz večih sistemov (npr. šifrant kupcev poenotiti preko več sistemov z vgrajenimi matematičnimi algoritmi za odkrivanje ujemanja podatkov),
 - hitrost prenosa podatkov (zmožnost paralelnega delovanja omogoča prenose ogromnih količin podatkov v kratkem obdobju),
 - pregled izvora in celotnega tokokroga podatkov, ki bo postal v prihodnosti nujen.
- Te vrste orodij pa so zelo primerne za prihodnost, ker se ni več potrebno spraševati, ali bo mogoče dobiti podatke, ali lahko ugotovimo izvor podatkov, ali bomo dovolj hitro napolnili podatkovno skladišče, kako bomo implementirali kakovost podatkov ipd.

1.8 Orodje za analizo podatkov v podatkovnem skladišču

OLAP je kratica za On Line Analytical Processing, kar pomeni sprotno analitično procesiranje in je način podpore managementa na osnovi večdimenzionalnih podatkovnih kock. OLAP pripomočki omogočajo managerju vpogled v podatkovne kocke. Deluje tako, da je potrebno podatke najprej prenesti iz podatkovnega skladišča v OLAP podatkovno bazo, kjer jih lahko analiziramo. Vpogled in analiziranje tako pripravljenih podatkov omogoča uporabnikom preprosto pregledovanje podatkov in vrtanje v globino (*ang. drill down*). Na ta način bolje razumemo poslovanje podjetja, iščemo vzroke odstopanj, izvajamo "kaj-če" analize za določen poslovni problem in podobno (Marolt Šmid, 2008, str. 20-23).

Funkcionalnost OLAP opredeljuje predvsem dinamična večrazsežnostna analiza zgoščenih podatkov, pridobljenih iz podatkovnega skladišča, ki nudi končnemu uporabniku podporo pri analizi in odločanju. Podpora pri analiziranju vključuje naslednje aktivnosti (Krsnik, 2008):

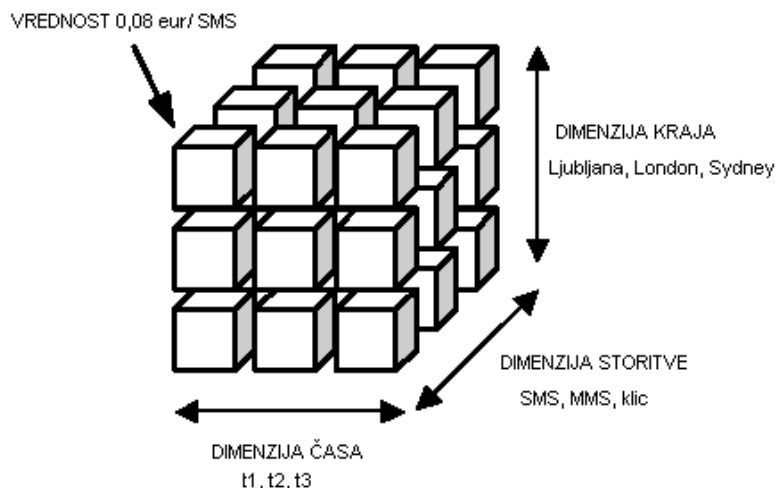
- modeliranje in kalkulacije prek več dimenzij ali več hierarhičnih nivojev,
- analiza trendov v različnih časovnih obdobjih,
- prikaz poljubnega dela podatkov,
- vrtanje v globino,

- enostavno medsebojno primerjanje različnih vidikov podatkov.

Ko z ETL-om napolnimo podatkovno skladišče v določenih časovnih intervalih, štejemo, da imamo v podatkovnem skladišču vse podatke kakovostne in urejene, tako da je naše delo le še analiziranje. Pri tem je tehnologija OLAP kot nalašč za to delo, saj v osnovi ponuja pogled na agregirane podatke z večih dimenzij.

Za primer sem si izbrala poslovanje telekomunikacijskega podjetja po svetu. Če hočemo analizirati poslovanje tega podjetja, moramo najprej vedeti, za kateri kraj bomo gledali (celina, država, mesto), naziv trgovine (kratko ime, dolgo ime), čas nakupa (leto, mesec, dan, ura), stranko (poštna številka, priimek) in izdelek/storitev (kategorija, tip, kakovost). Torej lahko na prodani izdelek/storitev gledamo z večih dimenzij, kot so dimenzije časa, kraja, stranke in podobno. Temu se strokovno pravi, da gledamo na podatke v podatkovnem skladišču **večdimenzionalno** – z večih vidikov naenkrat. To še ne pomeni, da moramo vedno za vsak izdelek upoštevati vse dimenzije (vidike), ravno nasprotno, v analize vključimo tiste dimenzije, ki nas trenutno zanimajo in so zanimive za analiziranje. Ponavadi preko dimenzij gledamo neko opravljeno transakcijo (nakup, prodaja, najem in podobno), za vsako transakcijo pa določimo, kaj od tega bo merljivo oziroma števno, kar pa je seveda različno od predmeta storitve.

Slika 5: Kocka - dimenzije kraja, časa in izdelka ter dejstva



Pri nakupu je to lahko neka cena izdelka, pri telefonskem klicu je to dolžina klica, vrednost klica in podobno. Vedno moramo vnaprej določiti glede na katero merljivo količino bomo pregledovali podatke prek več dimenzij. Primer na telekomunikacijskem podjetju je prikazan zgodaj na sliki, kjer je prikazana prodaja storitve glede na kraj in čas. Za vsako storitev (SMS, MMS, klic,...), ki jo opravimo v omrežju, lahko določimo, kje je bila opravljena (Ljubljana, London, Sydney ...) in v katerem času je bila opravljena ($t_1, t_2, t_3, \dots$). Tako gledamo nakup

storitve skozi tri dimenzije – storitev, kraj in čas. Zanima nas, kje se skriva cena SMS-a, ki je bil poslan iz Ljubljane nekam v Slovenijo v začetku meseca oktobra. Ta vrednost je merljiva vrednost in se skriva v presečišču dimenzije kraja, časa in cene storitve (0,08 EUR/SMS).

Tabeli, v kateri shranjujemo podatke, pravimo tabela dejstev (*ang. fact table*), v kateri merimo količino in ceno poslanih SMS-ov, ki so dejstva. Časovno dimenzijo lahko s hierarhijo razdelimo na leto, mesece, dneve, del dneva, uro, minuto; dimenzijo kraja lahko na celino, državo, mesto, kraj; izdelek lahko glede na tip izdelka, kategorijo izdelka in podobno. Dimenzije in tabelo dejstev tako poimenujemo kocka, s pomočjo katere lahko na primer analiziramo, katera storitev je bila najbolje prodajana konec oktobra 2007 in v katerem kraju.

Tako preidemo na bistveno lastnost uporabe tehnologije OLAP – glede na podatke, ki jih imamo v podatkovnem skladišču, lahko izvajamo različne analize. Če je dimenzij še več, so analize lahko toliko zanimivejše in kompleksnejše. Zelo preproste so tudi analize primerjave poslanih SMS-ov po istih mesecih v več letih, ko hočemo primerjati med seboj poslane SMS-e januarja v letih 2000–2005. Tehnologija OLAP na tem mestu upraviči svoj obstoj in uporabnost.

Vse to bi se sicer dalo z nekaj truda narediti tudi z uporabo stavkov SQL, vendar ne smemo pozabiti, da je lahko jezik SQL razmeroma težaven za uporabo že pri na videz enostavnih analizah. S tehnologijo OLAP pa to pomeni nekaj klikanja pri uporabi čarovnika in analiza je že pripravljena. Ravno tako za OLAP velja, da je praviloma izveden v večuporabniškem okolju po principu strežnik/odjemalec. Bistveno zanj pa je, da ima vseskozi kratek odzivni čas, ne glede na število uporabnikov, velikost in zapletenost podatkovne baze (Jagarinec, 2005, str.79).

2 KAKOVOST PODATKOV IN INFORMACIJE

2.1 Opredelitev podatka in informacije

Za sam uvod o kakovosti bom najprej predstavila razliko med podatkom in informacijo. V literaturi lahko dobimo veliko različnih definicij, ki govore o tem, kaj je podatek in kaj informacija. Izraza podatek in informacija se danes pogosto uporabljata kot sinonima. Vendar pa med njima obstaja precejšna pomenska razlika. Podobno se zadnje čase dogaja tudi z izrazom znanje, ki ga ponekod uporabljajo v pomenu podatka (Mohorič, 1999, str. 445-448).

V internetnem Slovarju informatike je **podatek** znano dejstvo o določeni stvari, predmetu, pojavu, ki je temelj za sklepanje, medtem po ko je **informacija** podatek, obdelan in prikazan tako, da je razumljiv uporabniku. Informacija je tako znanje pridobljeno iz podatkov (Islovar, 2009).

Preden oblikujemo informacijo, moramo najprej zbrati podatke, na podlagi katerih se potem odločimo in tvorimo informacijo. Viri podatkov so lahko zelo različni, kar je odvisno od dejavnosti podjetja. Vire lahko na grobo razdelimo na zunanje in interne. Zunanji viri so lahko različne podatkovne baze, ki nahajajo izven organizacije in se prenašajo ali vnašajo v interne sisteme organizacije preko različnih kanalov, uporabljajo se za različne namene in v različnih okoljih. Enako velja za interne vire podatkov, le s to razliko, da se zajemajo interno preko normalnega delovanja organizacije. Eden od pomembnih internih virov je operativni CRM, ki poleg zbiranja internih informacij deluje tudi kot skupni imenovalec vseh relevantnih informacij iz drugih internih sistemov (Tegeltija, 2008, str. 21-23).

Pomembno pri vsem tem pa je spoznanje, da povečanje kakovosti informacije temelji na povečanju kakovosti podatkov, kar je tudi usmeritev, ki ji sledi današnja informacijska tehnologija. Pri vse tem pa ne smemo pozabiti na uveljavljeno pa čeprav napačno razmišljanje, da so podatki v podatkovnem skladišču shranjeni in obdelani pravilni in s tem tudi kakovostni oziroma da je za morebitne napake in nekakovost 'kriva' prav informacijska tehnologija (Welzer, 1998, str. 72).

2.2 Opredelitev kakovosti podatkov in informacije

V literaturi se pojavljajo številne opredelitve kakovosti podatkov. Ena najbolj znanih pa je definicija Englisha (1999a), ki postavlja definicijo, v katerih loči naravno (*ang. inherent*) in stvarno (*ang. pragmatic*) kakovost. Pod naravno kakovost spada pravilnost in natančnost podatkov. Pod stvarno kakovost pa štejemo vrednost informacije, ki jo imajo natančni podatki v podpori uporabnikom pri sprejemanju poslovnih odločitev. Podatki, ki uporabnikom ne pomagajo pri doseganju ciljev, nimajo kakovosti, ne glede na to, koliko točni so.

Kakšen je uspeh podatkovnega skladišča, je v veliki meri odvisen od tega kako kakovostni so sami podatki v njem. Ko iz podatkovnega skladišča zagotovimo neke podatke, moramo biti prepričani, da so ustrezne kakovosti, saj ima to lahko v nasprotnem primeru zelo negativne posledice. Zaposleni lahko delajo analize na nekvalitetnih podatkih, pa se tega sploh ne zavedajo. Takrat govorimo o tem, da je podatkovno skladišče neuporabno in nezanesljivo.

Kakovost informacije je določena z razliko med pričakovano oziroma potrebno informacijo, ki jo določa naš trenutni cilj ter namen in pridobljeno informacijo. (Gerkeš, 2009). V idealnem primeru ne bo razlike med potrebno in pridobljeno informacijo. Merilo za kakovost informacije je torej: manjša je razlika med potrebno in pridobljeno informacijo, večja je kakovost informacije. Kakovost informacije se kaže tudi v tem, kako spodbuja prejemnika k dejanjem oziroma kako prispeva k temu, da se posameznik bolje odloča.

Kakovost podatka, in s tem posledično tudi informacije, je odvisna tudi od same arhitekture podatkovnega skladišča. Ne moremo točno reči, katera arhitektura je primernejša, saj je tu

veliko različnih mnenj. Če gledamo zgradbo centraliziranega podatkovnega skladišča je osrednje podatkovno skladišče edini vir podatkov za področno skladišče. Podatkovno skladišče tako vsebuje veliko količino zelo podrobnih podatkov, ki so iz daljšega obdobja, medtem ko so podatki v področnih skladiščih agregirani in sumirani omejene zgodovine v veliko bolj zapletenih strukturah (Golob, Welzer, Brumen, 2003, str. 13). Iz tega lahko sklepamo, da če so podatki v osrednjem skladišču napačni, se napačno agregirajo in sumirajo tudi v poročnih skladiščih. Prav tako pa je težavno tudi vzdrževanje velikega števila strukturiranih podatkov. Distributivna arhitektura podatkovnega skladišča pa, drugače kot centralizirana, ne predvideva centralnega podatkovnega skladišča, temveč le unijo neodvisnih področnih, ki vsebujejo veliko količino zelo podrobnih podatkov iz daljšega obdobja in ne le agregirane podatke omejenega časovnega obdobja. Vsako področno skladišče mora biti predstavljeno z dimenzijskim modelom, ki mora biti znotraj skladišča skladno, drugače podatkovno skladišče ne more delovati kot integrirana celota (Lalič, 2005, str. 12).

Informacijska tehnologija zajema področja zajemanja, obdelovanja, shranjevanja in prenašanja vseh vrst informacij. Pri doseganju zadovoljstva strank ima pomembno vlogo kakovost podatkov, ki neposredno vpliva tudi na učinkovitost poslovnih procesov. Kakovost informacij je tako konsistentno zadovoljevanje informacijskih potreb analitikov in strank (English, 1999, str. 24).

V zadnjem času znanje podjetja predstavlja vedno večjo prednost pred ostalimi tekmeci. Zaradi tega je vedno večje tekmovanje med tem kako najbolj zajeti podatke in reagirati na informacije, ki so v danem trenutku na voljo, ter z njimi ravnati na tak način, da nam prinesejo največjo korist.

2.3 Sodila kakovosti podatkov in informacij

2.3.1 Sodila kakovosti podatkov

Kakovosti podatkov v podatkovnem skladišču ne moremo natančno izmeriti, temveč jo lahko samo ovrednotimo oziroma ocenimo. Pri tem pa moramo vedeti za pomen meritve kakovosti. Veliko različnih metod merjenja kakovosti temelji na tem, da se postavijo pravila, nato pa pride do ocenjevanja primerov podatkov, ki postavljena pravila kršijo in ti tako predstavljajo odstotek nekakovostnih podatkov.

Pri natančnosti podatkov moramo upoštevati lastnosti, ki se lahko nanašajo na vsebino ali pa na strukturo podatkov. Temu pravimo *sintaktična natančnost*, ki opisuje pravilnost podatka, drugi vidik pa je *vsebinska natančnost*. Kot primer: v bazi podatkov imamo vrednost podatka Ime = Miha, ki je sintaktično natančna, vendar pa je vsebinsko napačno, če gre tukaj v tem primeru dejansko za osebo z imenom Janez (Godnov, 2008, str. 10).

Tako se lahko pojavijo slabi podatki, ki so posledica napak, ki se pojavijo pri vnosu podatkov, napačno določena poslovna pravila, podvojeni podatki, manjkajoče ali napačne vrednosti. Omeniti pa moramo še to zahtevo, da podatki sledijo poslovnim pravilom (npr. datum rojstva mora biti starejši vsaj za šest let od datuma vpisa v šolo). Podatki, tudi če so brez napak, nimajo vrednosti, če so poslovnemu uporabniku nerazumljivi ali nedosegljivi. V tem primeru je najbolje, da jih določimo z intervjuji uporabnikov, pregledom in analizo poslovnih pravil.

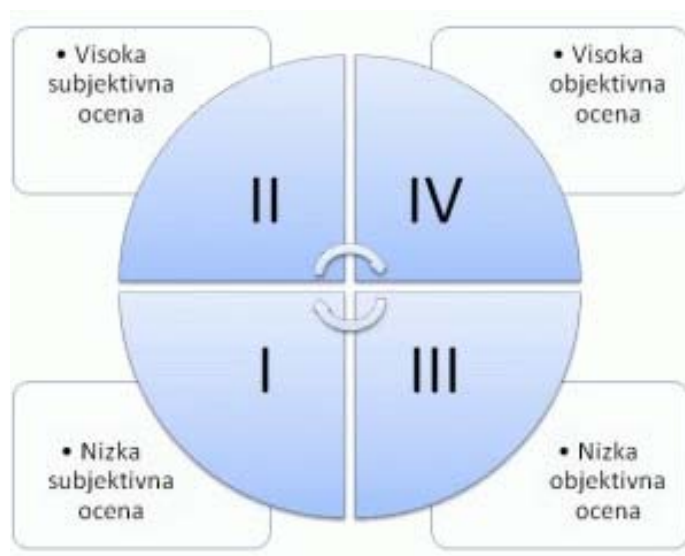
Poleg tega pa je potrebno pri vsakem sodilu upoštevati tudi vidik namenske uporabe, ki je pri ocenjevanju kakovosti bistvenega pomena. Vsak sektor v podjetju potrebuje tudi svoj nivo natančnosti, popolnosti in skladnosti pri zadovoljevanju svojih potreb. Takrat, ko pričakovanja in zahteve niso zadovoljene, uporabniki prenehajo zaupati podatkovnemu skladišču in iščejo nadomestitev potreb v drugih alternativah.

Kakovost podatkov, ki prihajajo v podatkovno skladišče, je potrebno nadzirati. Ko vhodne podatke večkrat zmerimo, nam meritev pokaže ali se res približujemo kakovosti podatkov, ki smo si jo zastavili. Vzdrževanje pa predstavlja tudi oceno uporabe novih podatkov in ugotavljanje vzrokov za slabo kakovost podatkov in njihovo odpravljanje (Rizman Žalik, 2004, str. 19-29).

Ocena kakovosti podatkov, še posebej, če je to prva ocena v podjetju, je zelo pomembna, saj podjetju zagotavlja osnovo za spremljanje različnih dejavnosti in njihov vpliv na kakovost podatkov v prihodnosti, in sicer izboljšanje ali poslabšanje kakovosti podatkov.

Godnov (2008, str. 10-11) v članku pojasni, da lahko oceno kakovosti podatkov razčlenimo na subjektivno in objektivno oceno. Subjektivna ocena kakovosti je močno zaznamovana z znanjem in potrebami udeležencev v procesu. Kako le-ti upravljajo s podatki in dojemajo kakovost podatkov močno vpliva na njihovo nadaljnje ravnanje. Subjektivno ocenjevanje ponavadi preverjajo s pomočjo vprašalnika, kamor anketirani zaposleni podajo svoje videnje kakovosti podatkov. Objektivno ocenjevanje lahko izvedemo z upoštevanjem poslovnih pravil ter omejitev, organizacijskih in državnih predpisov ali pa navedenih dejavnikov ne upoštevamo. Ko sta obe oceni določeni, je zelo pomembno preverjanje. Najboljše za podjetje je, da sta obe oceni enaki. Kljub temu, da sta lahko slabi, je to lahko dobro, saj pomeni da se zaposleni nekakovosti tudi dejansko zavedajo. Posebej pozorni pa moramo biti takrat, kadar se ti dve oceni med seboj razlikujeta. V primeru, da je objektivna ocena kakovosti dobra, subjektivna pa slaba, to pomeni, da zaposleni podatkom, ki jih imajo na voljo, ne zaupajo. Slabo oceno lahko pripišemo preteklim slabim izkušnjam s kakovostjo podatkov. Če se zaposleni soočijo s tem, potem v veliko primerih zaradi tega opravljajo dodatna, lastna preverjanja. Podobno je tudi, če je objektivna ocena slaba, subjektivna pa dobra. To pomeni, da zaposleni nekakovost podatkov sprejmejo kot nekaj normalnega in vsakodnevnega.

Slika 6: Primerjanje izidov subjektivne in objektivne ocene



Vir: Zakaj je kakovost podatkov tako pomembna?

Cilj analize kakovosti podatkov, kot jih prikazuje slika 6, je umestitev izidov v posamezno rubriko, pri čemer je najboljša novica za združbo, če analiza zapolni rubriki II in IV, medtem ko rubriki I in III zahtevata nadaljnjo obravnavo.

2.3.2 Sodila kakovosti informacije

Obstaja več različnih definicij sodil kakovosti informacij. Njihova opredelitev pa ni še popolnoma dokončana. Kakovost tako presojamo z različnimi sodili, ki so (Gradišar & Resinovič, 1998, str. 48-51):

- dostopnost,
- točnost,
- pravočasnost,
- popolnost,
- zgoščenost,
- ustreznost,
- razumljivost,
- objektivnost.

Dostopnost

Hitra dostopnost uporabnika do informacije, ko jo potrebuje, je izredno pomembna za učinkovito odločanje, saj se s čakanjem na informacijo, vrednost le-te že preveč zmanjša. Dostopnost merimo s časom, od trenutka zahteve določene informacije do trenutka, ko lahko razpolagamo z informacijo, ki lahko znaša samo nekaj mikrosekund ali pa nekaj ur ali celo

dni in je odvisen od metod, ki jih uporabljamo pri iskanju informacij, ter od sredstev, ki so pri tem na voljo.

Točnost

Zaradi nepravilnega ali napačnega delovanja informacijskega sistema in motenj v komunikacijskih kanalih ter neznanja, napačne razlage in površnosti človeka lahko pride do napačne in nezanesljive informacije, ki ne odraža dejanskega stanja sistema in tako napeljuje uporabnika k sprejemanju napačnih sklepov in odločitev ter akcij. Zanesljivost informacije je možno povečati z vgraditvijo nadzornih mehanizmov, ki omogočajo odkrivanje in popravljanje napak, v informacijski proces.

Pravočasnost

Da bi uporabnik informacijo koristno uporabil pri sprejemanju ustreznih odločitev in akcij, mora informacijo dobiti pravočasno, na kar poleg časa ustvarjanja informacije vpliva tudi dostopnost. Pravočasnost informacije je funkcija odzivnosti informacijskega sistema. Čim večja je odzivnost sistema, tem prej bo informacija na voljo uporabniku.

Popolnost

Pogosto se dogaja, da uporabnik dobi pravočasno in točno, vendar nepopolno informacijo, ki mu ne daje vseh potrebnih elementov za sprejemanje ustreznih odločitev in akcij. Ker pa absolutno popolne informacije seveda ni, jo opredelimo le kot bolj ali manj popolno.

Zgoščenost

Zaradi težnje k čim večji popolnosti informacije obstaja nevarnost zasičenosti uporabnika s preveliko količino podrobnosti, ki jih ne more niti pregledati, kaj šele uporabiti. Zato mora biti informacija ravno prav zgoščena oziroma kratka in jedrnata, da nudi uporabniku le tiste podrobnosti, ki jih nujno potrebuje.

Ustreznost

S tem sodelom ugotavljamo mero prilagojenosti informacije informacijskim zahtevam in potrebam uporabnika, saj je zanj učinkovita le tista informacija, ki odraža dejansko stanje hotenega področja in omogoča učinkovite odločitve in akcije. Seveda pa je ustreznost informacije za vsakega uporabnika drugačna, saj je za nekoga neka informacija odločujoča, za nekoga drugega pa neustrezna. Prav tako pa se ustreznost informacije spreminja tudi s časom, saj v nekem času ustrezna informacija zaradi spremenjenih pogojev lahko postane neustrezna.

Razumljivost

Informacija mora biti uporabniku razumljiva, saj drugače izgubi svojo vrednost in se ne more učinkovito uporabiti.

Objektivnost

Informacija mora čim bolj stvarna in objektivna ter oblikovana čim bolj nepristransko.

Mnenje, katero sodilo je med vsemi najpomembnejše, do danes še ni poenoteno, med vsemi pa najbolj izstopata sodili natančnosti in zaupanja, ki sta uporabniku informacije tudi najbližji.

Vzrokov, zakaj pride do nekovostnih podatkov v podatkovnem skladišču, je zelo veliko. Velik delež problemov s kakovostjo podatkov pade na sam izvor podatkov ter pri njihovem vnosu v podatkovno skladišče. Pri napakah pa so posredno ali pa neposredno povezani tudi skrbniki sistemov, sami uporabniki, netočno ali celo napačno določena poslovna pravila, izpadi samega delovanja skladišča in tako podobno.

Nekatere raziskave so pokazale, da so najpogostejši vzroki za nekovostne podatke neskladne opredelitve entitet, napačen vnos podatkov zaposlenih, različni prehodi in obdelave podatkov, različna pričakovanja uporabnikov, zunanji podatki, napačen vnos pri strankah, sistemske napake, spremembe vhodnih postopkov in drugo (Godnov, 2008, str. 10-11).

Nepopolni, nekovostni, umazani, netočni podatki in tako dalje (*ang.dirty data*) so problem vsakega podjetja že od nekdaj, vendar se dandanes daje temu veliko večji pomen kakor včasih. Zaradi tega je ena od osnovnih nalog podatkovnega skladišča, da vsebuje čiste, integrirane, zgodovinske podatke v kar se da najkrajšem času in z čim nižjimi stroški (Adelman & Moss, 2000, str. 255).

Problem kakovosti informacije lahko definiramo kot situacijo, v kateri vsebina in medij informacije ne zadovoljujeta potreb proizvajalcev, skrbnikov ali uporabnikov informacij (Eppler, 2003, str. 14). Eppler (2003, str. 29) opredeli osem problemov kakovosti informacij.

1. Mnogovrstni viri

Ista informacija izvira iz večih virov, kar povzroči različne vrednosti, vodi k zmedi ter k manjši kredibilnosti in sprejemljivosti. Posledica tega je, da raba informacije upada, uporabniki vanjo dvomijo in ji ne zaupajo. Poleg tega pa je vzdrževanje dražje in bolj težavno. Problem se lahko reši tako, da se razvije skupne definicije in skladne postopke.

2. Subjektivno pridobivanje

Informacija je pridobljena s subjektivnim mnenjem, kar vodi k pristranskosti. S tem se objektivnost informacije zmanjšuje, kar vodi do tega, da jo je težko vrednotiti. Rešitev se vidi v večjem izobraževanju poslovnih uporabnikov, boljših pravilih, kvalificiranih sistemih.

3. Napake pri pridobivanju

Sistemske napake pri pridobivanju informacij vodijo do izgubljanja informacij. Posledica tega je, da se poveča težnja po iskanju in popravljanju informacij. Problem lahko rešimo z izboljšavo postopka pridobivanja, spodbudo in kontrolo.

4. Prevelika količina informacij

Velik obseg nakopičenih informacij oteži dostop do informacij v razumnem času. Posledica tega je zahteva, da se presežek časa skrajša in informacije naredijo bolj jedrnat in zgoščene. Rešitev je analiza informacijskih potreb ter razvoj običajnih nizov relevantnih informacij.

5. Porazdeljeni sistemi

Porazdeljeni in heterogeni sistemi vodijo do protislovnih definicij, formatov in vrednosti. Zaradi tega informacije ne morejo biti združene in sestavljene. Rešitev dobimo v tem, da integriramo sisteme v eno platformo ali pa zmanjšamo število sistemov v enega.

6. Spreminjajoče se zahteve

Tako, kakor se spreminjajo zahteve uporabnikov informacij in okolje, tako se spreminja tudi relevantnost in uporabnost informacije v podjetju. Pri tem je najbolje predvidevati kakšne bodo spremembe, sistemi in zahteve v prihodnosti, preden se le ti pojavijo.

7. Varnost in zasebnost

Lahek dostop do informacij bo lahko vodil v konflikt z željo in potrebo po zasebnosti, varnosti in zaupnosti. Zaradi tega so se razvili mehanizmi za zavarovanje in omejen dostop do podatkov.

8. Pomanjkljivosti računalniške opreme

IT infrastruktura je v nekaterih podjetjih nezadostna in omejena z dostopi do informacij. Zaradi tega so uporabniki opreme nemotivirani in ne morejo delati s svojo največjo produktivnostjo. To težavo lahko rešimo tako, da nadgradimo sisteme, kar pa seveda zahteva tudi večje stroške, ki so povezani z nadgradnjo.

Iz zgornjih točk vidimo, da problem kakovosti informacije ni samo na koncu, ko informacijo že dobimo, vendar se začne že veliko prej v samem procesu pridobivanja informacij. Problem kakovosti informacije so posledica posredovanja neustreznih ali nepopolnih informacij, neskladnih ali nepravilnih informacij ali pa posredovanja informacij v napačnem smislu.

3 ZAGOTAVLJANJE KAKOVOSTI PODATKOV IN INFORMCIJE

3.1 Dejavniki, ki vplivajo na kakovost podatkov

Na kakovost podatkov v podatkovnem skladišču lahko vpliva mnogo dejavnikov. Le ti so razdeljeni na dve skupini: človeški dejavniki ali tehnološki dejavniki.

Človeški dejavniki so lahko posredni (lastnosti posameznika, npr. prizadevnost, znanje, izkušnje, sposobnost itd. in se razlikujejo od človeka do človeka) ali neposredni (sistem zagotavljanja kakovosti v podjetju, motiviranje ljudi za kakovostno delo, načrtno usposabljanje in izobraževanje zaposlenih, urejenost reševanja problemov kakovosti ipd).

Tehnološki dejavniki pa so sestavljeni iz delovnih sredstev, finančnih sredstev za investicije, opreme za merjenje in preskušanje, zahtevanih standardov in predpisov, surovin in materialov ipd. Kakovost podatkov je odvisna od kombinacije obeh skupin dejavnikov, ki se med seboj tesno prepletajo (Marolt, 1994, str. 10).

3.2 Napake v podatkih, ki se pojavljajo v podatkovnem skladišču

V vsakem podatkovnem skladišču se najdejo nekakovostni in napačni podatki. Vendar pa v podatkovnem skladišču ne smemo imeti takih podatkov, ki so sprti s poslovnimi pravili ali so nekonsistentni. Tako podatkovno skladišče s takimi podatki je neuporabno in lahko podjetju celo škoduje.

V podatkovnem skladišču lahko obstaja možnost, da se napake ponavljajo in jih ne moremo odpraviti, ali pa je vpliv napak tako velik, da je zanje potrebno napraviti postopke, ki bodo napake odpravljali. V skrajnem primeru pa je potrebna celo sprememba arhitektura podatkovnega skladišča (Golob & Welzer & Brumen, 2003, str. 12-22).

Greenfield je tipe napak, ki se najpogosteje pojavljajo, **razdelil na štiri sklope** (Greenfield, 2009):

1. Nepopoln podatek:

- manjkajoči zapisi; to pomeni zapis, ki bi moral biti v izvornem sistemu, pa ga tam ni,

- manjkajoča vsebina polj,
- manjkajoča polja.

2. Nepravilen podatek:

- napačno kodiranje,
- podvojeni zapisi; v enem polju se lahko pojavi več kot eden zapis,
- napačni podatki, ki so vstopili iz izvirnega sistema v bazo,
- napačni seštevki.

3. Nerazumljiv podatek:

- več vsebin v enem polju,
- kode, ki so nepoznane in jih ne znamo dešifrirati,
- povezave med podatki, ki se ne ujemajo.

4. Neskladen podatek:

- neuskrajena uporaba različnih kod; nekje lahko za določitev spola uporabljamo F (*ang. female*) oziroma M (*ang. male*), medtem ko nekje drugje uporabljamo 1 za žensko in 0 za moškega,
- neuskrajena uporaba različnih kod; to je običajno problem pri opredelitvi subjekta, ki se s časom spreminja. Na primer, recimo v letu 2006 ste kupci A, B, C in D. Leta 2007 stranka A kupi kupca B. Leta 2008 stranka A proda dela, ki so pod kupcem A in C kupcu D. Ko zgradite svoje skladišče v letu 2009, ki temelji na vrsti poslovanja in opravimo analize, se lahko srečujejo z dilemo, kako opredeliti prodaje kupcem A, B, C in D v prejšnjih letih,
- spreminjanje objektov v času,
- večkratne kode za isti objekt,
- neuskrajena imena in naslovi,
- različne oznake z enakim pomenom; npr navajanje viola in škrlatne barve,
- neuskrajnost poslovnih pravil.

Ko se ukvarjamo z odpravljanjem napak moramo biti pozorni na to, da nam vse to povzroči veliko dolgoveznega dela. Zato je za reševanje teh težav najboljše stalno preverjanje in vzdrževanje podatkov. Zaradi tega lahko tudi več časa porabimo za preverjanje kot čiščenje. Najtežje je ravnanje in odkriti napake neskladja.

3.3 Priprava in čiščenje podatkov

Definicija procesa čiščenja podatkov je, da se ukvarja z odstranjevanjem napačnih podatkov in podatkov, ki so v podatkovnem skladišču nepotrebni. Čiščenje je nekakšna kurativa in ne preventiva, saj čistimo podatke, ki so v podatkovnem skladišču že nekakovostni. Vedno je na samo čiščenje podatkov potrebno gledati kot poslovni problem. Seveda ne moremo vedno očistiti vseh podatkov v nekem razumnem času, hkrati pa tudi ignoriranje kakovosti podatkov in s tem izogibanje čiščenju podatkov ni rešitev. Strokovnjaki priporočajo neko srednjo pot, do katere je potrebno priti s sklepanjem kompromisov, kar pa je težji del odločitve pri zagotavljanju kakovostnih podatkov za podatkovno skladišče. Potrebno se je odločiti, katere prednosti pridobi podjetje s čistimi podatki in določiti stroške čiščenja podatkov ter narediti analizo. Na koncu primerjamo stroške čiščenja podatkov v primerjavi s posledicami, če podatkov ne očistimo. Če poslovna izguba zaradi napačnega podatka presega stroške čiščenja, se odločimo za čiščenje te vrste podatka. Čiščenje podatkov izvedemo pri polnjenju v podatkovno skladišče, popravke podatkov pa lahko izvedemo tudi v transakcijskih sistemih. Včasih popravke izvedemo v podatkovnem skladišču in jih nato iz podatkovnega skladišča prenesemo nazaj v transakcijski sistem. Poleg čiščenja podatkov lahko podatke v podatkovnem skladišču tudi dopolnimo (Jaklič, 2001, str. 3)

Pri čiščenju naj vedno sodelujejo tudi končni uporabniki, ne pa samo informatiki, saj ti prej zaznajo nepravilnosti in najboljše poznajo procese, v katerih nastajajo podatki in so odgovorni za kakovost podatkov.

Zgodi se lahko, da vodja projekta nima dovolj časa in drugih virov za izvedbo določenih aktivnosti in čiščenje podatkov. Takrat ne more zagotoviti kakovostnih podatkov, ki jih potrebujejo in zahtevajo uporabniki. Adelman in Moos (2000, str. 38 - 40) menita, da se je v takih primerih zelo prepričljivo držati naslednjih navodil:

1. Nikoli ne poizkušaj očistiti vseh podatkov.

Vsakdo bi rad imel vse podatke popolnoma očiščene, vendar ni noben za to pripravljen veliko plačati ali čakati veliko časa. Čiščenje vsega vzame veliko časa.

2. Nikoli ne pusti vseh podatkov neočiščenih.

Vedno očisti vsaj nekatere podatke. Konec koncev je namen podatkovnega skladišča, da zagotavlja čim bolj čiste in zanesljive podatke.

3. Določi prednosti, ki jih imajo čisti podatki.

Zastavi si vprašanja, ki ti pomagajo pri tem, da ugotoviš vzroke za neskladnosti (ali so vzroki slabi podatki ali programska oprema, kakšne so izgube zaradi slabih podatkov in kateri podatki so slabi).

4. Določi ceno, ki jo imamo s čiščenjem podatkov.

Preden se lotimo čiščenja podatkov moramo narediti oceno stroškov, ki jih imamo s čiščenjem podatkov. Potrebno je določiti in oceniti, koliko časa je potrebno za izvedbo naslednjih opravil:

- analizo podatkov,
- določitev pravil pri popravkih podatkov,
- napisati in izdelati program za čiščenje podatkov,
- popraviti stare formate in podatkovne zbirke.

5. Primerjaj ceno čiščenja podatkov z izgubo, če bi podatke pustil take kot so bili.

Vsako dejavnost v poslovnem svetu je potrebno oceniti z vidika stroškov, kar pa velja tudi za čiščenje podatkov. Pri vsakem podatkovnem elementu je potrebno primerjati med seboj stroške čiščenja z izgubo, ki jo prinašajo podatki, če ostanejo umazani. Vendar je določitev poslovne škode zaradi umazanih podatkov zelo težka, kakor je tudi izguba ugleda podjetja pri dobaviteljih in kupcih.

6. Postavi prioritete slabih podatkov nasproti očiščevalnim ciljem.

Težji del je usklajevanje časa, ki ga imamo pri izvedbi celotnega projekta, s cilji ki jih je potrebno na koncu doseči. Ker imamo na seznamu veliko točk in se vseh ne bo dalo narediti v načrtovanem času, je potrebno določiti vrstni red čiščenja in prioritete.

7. Za vsako točko očiščevanja se vprašaj, ali se da očistiti.

Preveriti je potrebno, kje so viri za čiste podatke: v starih podatkovnih zbirkah, datotekah, papirnih arhivih ali drugje. Lahko pa z analizo pridemo do ugotovitve, da se želenih podatkov enostavno ne more očistiti.

Velikokrat pa samo čiščenje ni dovolj, saj s tem, da očistimo podatke, zagotovimo kakovostne podatke samo za določen trenutek. Čez neko obdobje bomo imeli v podatkovnem skladišču spet nekakovostne podatke, saj s samim čiščenjem ne odpravimo vzrokov za nekakovost, poleg tega pa se podatki skozi čas spreminjajo in s staranjem izgubljajo na svoji kakovosti, ki so jo imeli od začetka.

3.4 Zagotavljanje kakovosti informacije

Ko hočemo zagotoviti kakovost podatkov, tega ne moremo narediti samo z enim posegom ali samo z enim projektom. Zagotavljanje kakovosti je dolgotrajen proces, ki traja skozi celoten življenjski cikel podatkovnega skladišča. 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 ljudi in procese. Po mojem mnenju pa je v samem procesu zagotavljanja kakovostnih podatkov v podatkovnem skladišču kakovost podatkov odvisna od tega, kakšni so postopki polnjenja skladišča.

Pri zagotavljanju kakovosti informacije moramo uspešno in natančno izvajati kontrolo pregledov informacij, ki morajo biti natančno določeni, vedeti moramo kaj in zakaj to pregledujemo, v kakšni obliki bomo podali rezultate kontrolnega pregleda in na koncu kakšni bodo ukrepi na osnovi pridobljenih rezultatov. Kakovost podatkov pač ni samoumevna in brezplačna. Zanj se je potrebno boriti in jo ohranjati.

Vedno pa je potrebno upoštevati tudi namen uporabe podatkov in informacije v podatkovnem skladišču. Pri zagotavljanju kakovosti moramo vedeti, da če ima organizacija v svoji zbirki podatkov na primer 10 odstotkov netočnih naslovov svojih strank, so lahko ti podatki zadostne kakovosti v primeru, če te podatke uporabljamo za neko določeno trženjsko akcijo. Če pa te podatke uporabljamo takrat, ko okrog razpošiljamo vabila za volitve, pa so ti podatki nekakovostni. Torej odvisno je za kaj in kdaj potrebujemo podatke (Godnov, 2008, str. 10).

Uspeh organizacije v kateri delujemo, je močno odvisen od sprejemanja pravilnih in pravočasnih odločitev. Sprejemanje odločitev pa zahteva popolne in ažurne informacije, kajti samo take omogočajo strateško in taktično odločanje na vseh ravneh v podjetju in merjenje učinkovitosti teh odločitev.

3.5 Proces celotnega upravljanja s kakovostjo informacije

Kot sem omenila že zgoraj, se problem kakovosti informacije pojavi že pri samem procesu pridobivanja informacij. Večina organizacij ne razume dejanskega načela zagotavljanja kakovosti podatkov. Kvaliteta podatkov ni le čiščenje podatkov. Predvsem pa je kakovost in kvaliteta podatkov v podatkovnih skladiščih zelo odvisna od kakovosti in kvalitete podatkov na samem viru.

Proces celotnega upravljanja s kakovostjo informacije (*ang. Total Information Quality Management - TIQM*) ni projekt, ampak proces. Lahko bi rekli, da je koncept, ki upravlja podatke v podjetju na podlagi katerega se s pomočjo podprocesov oblikujejo nove informacije, katere služijo nadaljnjim upravljavcem in samemu managementu podjetja. Vendar pa TIQM ni samo to, ampak veliko več, saj ni le program ali preprost proces s katerim ocenjujemo in prečiščujemo podatke.

TIQM je bil prej poznan pod imenom TQdM (*ang. Total Quality data Management*). Vendar pa je English s časoma ugotovil, da se poslovni ljudje bolj poistovetijo z izrazom informacija kot pa podatek. Kljub temu pa še veliko ljudi uporablja stari izraz (TIQM Quality System, 2009).

Pri vsem skupaj gre za zavedanje informacijske soodvisnosti, ki jo imamo v podjetju ter zavedanje, da informacije predstavljajo ključno vrednost podjetja oziroma dajejo konkurenčno prednost podjetju. Zavedati se moramo, kako pomembno je zadovoljstvo stranke in nenazadnje, kako pomembna je kultura v podjetju, katere del so vsi zaposleni in s katero se trudijo za izboljšavo delovnih procesov, zadovoljstvo stranke in minimizacijo stroškov.

TIQM je posledica nepretrgane izboljšave procesa v katerem pridobivamo informacije in sicer v dveh sestavinah (TIQM Quality System, 2009):

- Razvoj informacijskih sistemov: tukaj gre za razvojne sestavine, s pomočjo katerih definiramo informacije, razvijamo in izboljšujemo delovne procese, informacijske sisteme, informacijsko infrastrukturo in baze podatkov.
- Poslovni in proizvodni sistemi: tukaj kreiramo, osvežujemo, brišemo podatke in jih dostavljamo uporabnikom, ki jih potrebujejo.

Pri pregledu celotne metodologije TIQM bomo spoznali, zakaj je kakovost informacije in njeno nenehno izboljševanje tako zelo pomembno. Celoten proces TIQM lahko zapišemo v šestih korakih (English, 1999, str. 69):

1. ocenjevanje kvalitete lastnosti podatkov in informacijske arhitekture,
2. ocenjevanje kakovosti informacije,
3. ocenjevanje stroškov, ki se pojavijo zaradi nekovostnih informacij,
4. reorganizacija in prečiščevanje podatkov,
5. izboljšava procesa pridobivanja kakovostnih informacij,
6. oblikovanje okolja za kakovostne informacije.

Proces upravljanja s kakovostjo informacije je sestavljen iz petih delov. Nad vsem tem pa visi oblikovanje okolja in kulture v podjetju, s katerim se zavedamo pomembnosti nenehnega izboljševanja kvalitete informacije, ki jo management potem uporablja pri svojih odločitvah.

3.5.1 Proces 6: Oblikovanje okolja za kakovostne informacije

Proces 6 je obravnavan kot prvi zato, ker je bistvenega pomena za dolgoročno izboljšanje kakovosti informacije. Če v podjetju obstaja slaba kakovost informacije, lahko pride do tega, da je stalnica kulture v podjetju to, da je nivo kvalitete informacije nizek in slab. Management

ima rajši hitrost, kakor kvaliteto. Naj si bo to zaradi nepoznavanja prave vrednosti, ki jo ima informacija, ali pa zaradi tega, ker imajo management in drugi uporabniki informacij tako mišljenje, da se kvaliteta informacij kaže v tem, kako hitro je opravljanje procesa in ne kakšna je njegova kakovost.

Čeprav je težko, se morajo v organizaciji vsi zavedati, da so pomemben del, ki sestavlja celoto. Zelo pomembno je, da se razume vrednost informacije, ki poteka po poti od ustvarjalcev informacije do njenih končnih uporabnikov. Pri tem se morajo ljudje v podjetju tudi zavedati soodvisnosti sodelavcev pri izmenjavi in posredovanju informacije in tega kaj uporabniki informacij od samih informacij pričakujejo. Veliko k vsemu temu pripomore stalno izobraževanje in izpopolnjevanje, saj le tako lahko sledijo vsem spremembam in pripomorejo k večji učinkovitosti.

3.5.2 Proces 1: Ocenjevanje kvalitete lastnosti podatkov in informacijske arhitekture

Proces pred ocenjevanjem kakovosti informacije je proces ocenjevanja kvalitete lastnosti podatkov in informacijske arhitekture. Kakovosti informacije ne moremo meriti brez tega, da poznamo lastnosti podatkov in natančnih lastnosti, kakšne naj bi informacije bile. Z analizo lastnosti podatkov v naših bazah ocenjujemo kakovost informacije v podatkovni bazi ali izven procesa.

Osnovni korak tega procesa je definiranje bistvene in kritične lastnosti podatkov in informacijske arhitekture ter definiranje skupine informacij, v katerih se lahko pojavljajo informacije slabe kvalitete. Po tem koraku sledi proces, kjer se določi skupine uporabnikov, ki informacije uporabljajo. Konča pa se s procesom v katerem se oblikuje tehnična ocena podatkov in preveri minimalno ustreznost, ki je zahtevana.

3.5.3 Proces 2: Ocenjevanje kakovosti informacije

Proces ocenjevanja kakovosti informacije je zelo podoben procesu, ki ga uporabljamo za ocenjevanje kakovosti v proizvodnji. Obstajajo ocene tehničnih skladnosti s specifikacijami in zmanjševanje spremenljivosti končnih izdelkov, v našem primeru informacij. Na samem koncu pa ocenjujemo odstopanje med stopnjo zadovoljstva kupca in njegovimi pričakovanji. Proces ocenjevanja kakovosti poteka v osmih korakih. Prvi korak je, da določimo skupino, ki jo bom ocenjevali ter lastnosti, ki jim bomo namenili največ pozornosti. Po tem določimo katero izmed baz podatkov oziroma proces bomo ocenjevali v trenutku. Korak, kjer po naključnem izboru izberemo vzorce podatkov za ocenjevanje, je ključnega pomena. Da bi zmanjšali stroške analiziranja, izberemo reprezentativni vzorec podatkov na podlagi katerega se naredi ocena stanja. Kljub temu, da je to samo vzorec, ki ne zajema celotne populacije, dobimo neko približno vzorčno oceno, ki pa mora biti ob dobri statistični analizi dokaj blizu resničnemu stanju. Ko imamo to narejeno, sledi razlaga dobljenih podatkov in podajanje

ocene na podlagi opravljene analize. Končni rezultat je podano poročilo o kakovosti informacije.

3.5.4 Proces 3: Ocenjevanje stroškov, ki se pojavijo zaradi nekovostnih informacij

Nekoč je bilo pri izboljševanju kakovosti informacije težko določati stroške, ki so bili povezani s slabo kakovostjo informacije. Proces ocenjevanja stroškov, ki se pojavijo zaradi nekovostnih informacij, nam predstavlja rešitve odpravljanja stroškov, povezanih s slabo kvaliteto informacij. Da lahko ocenimo stroške kvalitete informacije, moramo imeti določene neke poslovne cilje (povečanje dobička, zmanjševanje stroškov itd...), da lahko ocenimo primerjavo.

Prvo na kar moramo biti pozorni je, da poznamo in določimo cilje. Predvsem pazljivi moramo biti pri kvaliteti podatkov, saj je to vrednost zadovoljnega kupca oziroma uporabnika, ki jo merimo z merilom zapravljenih in izgubljenih priložnosti. Sledi ocena stroškov pri pridobivanju informacij, kje in v kaj smo vložili največ:

- Razvoj informacijske infrastrukture. Osnove za ohranjanje in vzdrževanje informacij.
- Vrednost in razvoj posredovanja informacij. Z razvojem uporabe in izkoriščanja informacij se večja dodana vrednost proizvodom, storitvam oziroma podatkom.
- Razvoj v nepotrebne izboljšave. Sem štejemo predvsem izboljšave, ki povečujejo stroške samemu podjetju, ne pa povečujejo vrednosti proizvodom in storitvam. Taki primeri so lahko nepotreben razvoj podatkovnih baz, odvečne aplikacije, ki zajemajo nekatere informacije, ki jih zajemajo že nekatere obstoječe, razvoj v vmesne faze, ki pa lahko samo podvajajo in preoblikujejo informacije iz ene podatkovne baze v drugo.

3.5.5 Proces 4: Reorganizacija in prečiščevanje podatkov

Reorganizacija in prečiščevanje podatkov je proces izboljševanja proizvodov, storitev oziroma podatkov. Bistvo tega procesa je, da se poišče vse neustrezne in pomanjkljive podatke ter se jih popravi na nivo kvalitete, ki nam ustreza. Ko so podatki očiščeni, se ponovno obdelajo, tiste podatke, ki pa se jih ne da obdelati, vzamemo iz nadaljnje obdelave in jih zavržemo. Proces nam prikazuje, kako je potrebno podatke preoblikovati do nivoja kvalitete, ki nam je sprejemljiv. Ko preoblikujemo podatke želimo:

- Oblikovati kakovostne informacije znotraj podatkovnih baz.
- Preoblikovati podatke za prilagajanje novemu sistemu ali novo uvedenim aplikacijam.
- Preoblikovati, prečistiti in vse to poslati v podatkovna skladišča.

Proces poteka tako, da najprej poiščemo podatke, ki jih je potrebno reorganizirati in prečistiti. Nato sledi analiza vira podatkov v kateri ugotovljamo, kakšno je odstopanje od pravil in kakšni so vzorci podatkov. Z analizo poiščemo ključen pomen ali vzorec podatkov, ki ga želimo standardizirati. Ko je standardizacija podatkov narejena, jih lahko med seboj primerjamo, povežemo in tako najdemo podvojene podatke. Če so podatki pravilni, lahko vstopijo v nadaljnji proces. Če pa podatki niso pravilni, na primer podvojeni, se jih s pomočjo določenih algoritmov združi. Podatke lahko popravljamo na dva načina: avtomatično popravljanje podatkov ali po sproti popravljanje skrbnikov informacij. Tako so ustrezni podatki poslani v podatkovno skladišče in podatkovne baze. Preden pa podatki tja prispejo, se še enkrat preveri, ali je bila izpeljava, preoblikovanje in povzemanje podatkov v skladu s pravili. S tem se izognemo podvajanju podatkov in njihovi pristnosti.

3.5.6 Proces 5: Izboljšava procesa pridobivanja kakovostnih informacij

Izboljšava procesa pridobivanja kakovostnih informacij je zadnja stopnja v sklopu procesov izboljševanja kakovosti informacij. Tukaj gre za nekakšno združitev vseh prej opisanih procesov, saj bi iz prejšnjih analiz procesov težko govorili o neki bistveni izboljšavi kvalitete informacij. Zato je to zadnja faza v procesu za podjetja, ki veliko vlagajo v svoja znanja in razvijajo svoja znanja. Izboljšava procesa pridobivanja kakovostnih informacij bi moral biti za podjetje najbolj pogosto uporabljen proces med vsemi šestimi procesi. Tu gre za uporabo znanih problemov, analizo njihovega izvora ter implementacijo procesa izboljšave kvalitete informacij, ki nas ščitijo pred slabo kvaliteto informacij in z njimi povezanimi nepotrebni stroški.

Proces začnemo tako, da najprej definiramo problem in določimo skupine ljudi, ki se bodo soočile z izboljšavo kakovosti informacije. Nato razvijemo plan izboljšave kakovosti in plan, kako bomo te izboljšave izvedli. Kot je v navadi se to dela v testnem obdobju, v katerem ocenjujemo, beležimo in spremljamo spremembe. Če rezultati v testnem obdobju ne kažejo pravih in zadovoljivih rezultatov proces vrnemo v predhodno stopnjo nazaj, seveda odvisno od tega kje se skrivajo glavni vzroki za take rezultate. Če pa smo z rezultati zadovoljni sledi zadnji korak, v katerem uspešno opravljen proces za pridobivanje kakovostnejših informacij vgradimo v procese v podjetju in stalno z nadzorom pridobivamo kakovostnejše informacije.

3.6 Zaposleni in kakovost informacije

Kakovost informacije je tudi na splošno v današnji dobi tako zelo pomembna, da njeno zanemarjanje lahko pripelje podjetje v propad. S kakovostjo in temu primerno miselnostjo mora biti seznanjen vsak, ki je zaposlen v podjetju. Zaposleni imajo na kakovost podatkov zelo velik pomen in vpliv. Sodelujejo pri celotnem procesu pridobivanja podatkov, polnjenju podatkovnega skladišča s podatki, njihovem urejanju in na koncu analiziranju.

Skrbniki podatkovnih skladišč morajo pri gradnji in nastajanju le-tega z ekipo načrtovalcev sodelovati od samega začetka. Podjetja danes ne morejo poslovati učinkovito in s tem konkurirati na globalnem področju, če svojim zaposlenim ne nudijo pravočasnega dostopa do točnih in skladnih informacij.

Na začetku uvedbe informacijske podpore odločanju je bil uporabnik podatkovnega skladišča precej pasiven. Znal je povedati le kaj si želi, če pa to ni bilo na prioritetni listi programerja, je na podatke lahko čakal tudi več mesecev. Dobljeno poročilo je bilo statično in uporabnik se je moral z njim zadovoljiti. Današnja vloga je bolj aktivna, saj si nekateri uporabniki že sami priskrbijo informacije, ki jih želijo.

Pri zagotavljanju kakovosti je zelo pomembno, da uporabniki vzamejo podatkovno skladišče "za svoje" in redno namenjajo nekaj časa kakovosti podatkov v podatkovnem skladišču (npr. primerjava s podatki v transakcijskih sistemih, tako da se hitro ugotovijo napake). Še posebej pomembno je predvideti, upoštevati in preveriti vpliv sprememb na kakovost podatkov v podatkovnem skladišču ob vseh spremembah poslovnih procesov. Na ta način zagotavljamo ažurnost in točnost podatkov v skladišču (Milavec, 2006, str. 9-11).

Pri izdelavi podatkovnega skladišča pa morajo ravno tako sodelovati tudi končni uporabniki, saj bodo le na ta način zbrani pravi podatki, ki bodo odgovarjali na poslovna vprašanja, podatkovno skladišče pa bo hitro vrnilo prave odgovore. Sodelovanje poslovnih analitikov, ki poznajo poslovna pravila, je za organizacijo ključnega pomena. Oni ugotavljajo vsebinsko pravilnost podatkov. Kljub temu, da je pregledovanje usklajenosti podatkov s poslovnimi pravili mogoče avtomatizirati, se ročnemu pregledovanju podatkov skoraj nikoli ne bomo mogli izogniti (Godnov, 2008, str. 10). Nekakovost podatkov je, ko se pojavi, potrebno odpraviti, kar povzroča določene stroške in nezadovoljstvo med zaposlenimi, ki morajo podatke popravljati. Poleg tega, da imajo delo s popravljanjem podatkov, jih to tudi odvrne od pomembnejšega dela.

4 VPLIV KAKOVOSTNIH IN NEKAKOVOSTNIH PODATKOV NA POSLOVANJE PODJETJA

Za informacijo lahko rečemo, da je kontekst, v katerem so vzeti in interpretirani podatki. Če imamo samo informacijo, ki jo vzamemo iz večjega dela 'sestavljenske', je sama po sebi nekoristna in ničvredna. Informaciji moramo tako vedno dodati še znanje in izkušnje sporočevalca, možnost uporabe, primerno komunikacijsko orodje, komunikacijski kanal in ustrezno podporo. Le-ta poskrbi, da sporočevalec dobi informacijo v pravilni obliki in seveda ob pravem času. Takrat postane informacija pomembno sredstvo v vsakodnevnih poslovnih izzivih. S pravo informacijo pridemo do novih strank, zadržimo obstoječe, ustvarimo konkurenčno prednost in jo tudi vzdržujemo, povečujemo in izvajamo navzkrižne in dopolnitvene prodaje itd. Izpolnjen pogoj, da informacija postane najpomembnejše sredstvo

je, da jo prejme ustrezna oseba, ki jo bo znala uporabiti na tak način, da ji bo to prinašalo določeno prednost (Tegeltija, 2008, str. 21- 23).

Informacije so vedno bolj pomemben dejavnik uspešnega delovanja podjetja, zato moramo vzpostaviti tudi ustrezne sisteme varovanja informacij na temelju varnostnih standardov, zakonodaje in razpoložljive informacijske tehnologije. Le-ta pa koristi svojemu namenu, če jo uporabljamo na primeren ter pravilen način, da nam prinaša koristi (zmanjšuje stroške in povečuje produktivnost) (Horjak, 2009).

4.1 Vplivi nekovostnih podatkov na poslovanje

Po različnih raziskavah, ki so jih opravljali na temo nekovostnih podatkov, je odstotek podatkov, ki bi jih lahko kvalificirali kot nekovostne, v različnih sodobnih zbirkah podatkov med 1 in 10 odstotki, medtem ko je odstotek nekovostnih podatkov v transakcijskih zbirkah podatkov pogosto kar med 60 in 90 odstotki. Kljub takemu stanju, pa odgovorni v podjetju še vedno ne namenijo dovolj pozornosti, čeprav to lahko zelo negativno vpliva na dobiček podjetja (Godnov, 2008, str. 10-11).

Nekovostni podatki v podjetju imajo zelo velik vpliv na samo poslovanje podjetja. Po ugotovitvah Burns-a (2005) nekovostni podatki:

1. Povzročajo stroške.

Delo oseb, ki se ukvarjajo z odpravo vseh teh nekovostnih podatkov. Nekatera podjetja imajo posebne oddelke, ki se ukvarjajo s strankami (npr. poslan napačen izdelek, izdelek poslan na napačen naslov), zaradi česar ima podjetje dodatne stroške.

Če ima podjetje v svojem podatkovnem skladišču nekovostne podatke, ima z njimi tudi določene stroške. Lahko se zgodi, da določena organizacija ne more izvršiti določenega poslovnega procesa, ker ima netočne podatke o naslovih v svojem podatkovnem skladišču. Tako lahko organizacija celo izgubi ali zamudi poslovne priložnosti, ki se ji ponujajo, ker zaradi nekovostnih podatkov in informacij ni vložila določenih investicij. Ravno tako pa lahko organizacija izgubi ugled na tržišču, saj ji ljudje ne zaupajo več tako močno.

English (1999, str. 209-210) opredeli stroške nekovosti informacije na tri glavne kategorije:

- **Stroški procesov, ki niso uspeli.**

To so stroški, ki nastanejo, ko se proces ne izvede pravilno. Kot primer bi navedla, da se zaradi nepravilne cene stranki zaračuna napačen znesek.

- **Stroški popravljanja napak.**

Stroški, ki se pojavijo pri preverjanju in popravljanju napak, ki se pojavijo skozi sam proces. Ko na primer zaposleni ne zaupajo v pridobljene podatke in jih gredo še enkrat preverjati in popravljati.

- **Stroški zgrešenih in zamujenih poslovnih priložnosti.**

Le-ti so prihodki in dobiček, ki ni realiziran zaradi slabe kakovosti informacije. Podjetje lahko zaradi napačnih oziroma pomanjkljivih informacij ne opravi določenih vlaganj. Lahko pa pripelje celo do te mere, da že dobljene posle lahko izgubi.

2. Vplivajo na razmerja s strankami.

Isto stranko imamo lahko v podatkovnem skladišču večkrat vneseno, zaradi česar večkrat prejme našo pošto. Poleg tega ima lahko stranka napačno vpisan naslov, zato ne dobi blaga, ki ga je naročila. Vse to lahko na koncu vodi do izgube strank, saj lahko le-te pri ponavljanju teh napak rajši odidejo h konkurenci. Pri obravnavanju strank moramo predvsem upoštevati naslednja dejstva:

- Podjetja vedno porabijo veliko več denarja za pridobitev novega kupca kot za ohranitev že obstoječega kupca, ki je podjetja navajen.
- Veliko dražje je ponovno pridobiti kupca, ki je odšel k tekmecu, kot pa skrbeti, da ostane zadovoljen.
- Izdelek je mnogo lažje prodati obstoječemu, ki nam zaupa, kot pa novemu kupcu.
- Nekateri kupci podjetju prinašajo več denarja kot drugi.

3. Vplivajo na poslovne odločitve.

Slabi podatki so slaba osnova za sprejemanje izvedbenih, taktičnih in strateških odločitev.

4. Vplivajo na zamujene priložnosti.

Obvladovanje razmerij s kupci in poslovno obveščanje, kamor sodi podatkovno rudarjenje, ne prinesejo zelenega učinka. S podatkovnim rudarjenjem lahko pridemo do vzorcev, ki ne ustrezajo dejanskemu stanju trga, združba pa zamudi svojo priložnost.

5. Negativno vplivajo na celotno kakovost poslovanja podjetja.

6. So življenjskega pomena.

Na primer v bolnišnicah je lahko napačen zapis krvne skupine pacienta usoden in ima za posledico smrt. Zavedati se moramo, da so nekateri podatki življenjskega pomena.

Izguba časa in veliko energije zaposlenih v organizaciji so posledica zgornjih nekovostnih podatkov, ko se le-ti ukvarjajo z odpravo njihovih posledic. Zaposlenim to počasi postane vsakdanje opravilo, posledično s tem pa se med njimi uveljavi mnenje, da je to čisto nekaj normalnega (Burns, 2005).

4.2 Koristi kakovostnih podatkov

Podjetje se zavzema za to, da ima v svojem podatkovnem skladišču čim manj nekovostnih podatkov. Ko dela selekcijo in čistko, je pomembno, da pretehta stroške, ki jih bodo ti ukrepi potegnili za seboj in koristi, ki jih bodo ti ukrepi prinesli. Koristi, ki jih imamo od kakovostnih podatkov, so naslednje in razdeljene na štiri dele:

1. **Ekonomske koristi**, ki so popolnoma merljive v denarju (npr. boljša kakovost podatkov vpliva na višje prejeme).
2. **Finančne koristi**, pod katere lahko štejemo zvišanje prihodkov in znižanje stroškov.
3. **Merljive koristi**, ki jih sicer ne moremo izraziti v denarju, jih pa lahko merimo v neki drugi merski enoti (npr. boljši podatki davčne uprave omogočajo časovni prihranek podjetij in posameznikov, ki ga lahko izmerimo).
4. **Nemerljive koristi** so koristi, ki jih ne moremo izraziti z mersko enoto kot je denar (npr. izguba ugleda podjetja). Če izboljšamo kakovost podatkov, pridemo do večjega zadovoljstva strank, večjega zadovoljstva zaposlenih v podjetju ter boljše kakovosti storitve ali izdelka (Zakaj je kakovost podatkov tako pomembna?, 2009).

4.3 Vpliv kakovostnih informacij na odločanje v podjetju in njegovo uspešnost

Skupna točka vseh danes uspešnih organizacij je hitrost njihovega zaznavanja, odzivanja ter prilagajanja na spremembe. Podlaga današnji uspešnosti organizacij pa je kakovost in hitrost odločanja ter posledično dobro izvajanje odločitev. V podatkovnih skladiščih hranimo veliko število podatkov, ki se uporabljajo pri sprejemanju poslovnih odločitev. Vendar kakovost podatkov sama po sebi ne zadošča. Iz podatkov moramo poiskati odgovore na manj ali bolj zapletena poslovna vprašanja, ki naj pripeljejo do boljših poslovnih odločitev. Z drugimi besedami, iz zbranih podatkov skušamo izluščiti informacije in znanje, ki jih potem potrebujejo vodje pri vodenju in odločanju. Bistvena funkcija manarerskega informacijskega sistema je oskrbovanje vodij z informacijami.

Pridobiti kakovostne informacije je v vsakem podjetju, če se odločamo na podlagi podatkov iz podatkovnega skladišča, zelo pomembno za sprejemanje kvalitetnih odločitev. Kvalitetne odločitve pa so ključ do uspešnosti in učinkovitosti. Sprejemanje odločitev je najpomembnejša aktivnost organizatorja in vodje procesa, v katerem se ustvarja neka dodana vrednost, poklicno se s tem ukvarjajo predvsem managerji. Ko se odloča, morajo njegove odločitve voditi k uspešnosti in učinkovitosti podjetja, se pravi da mora stremeti k doseganju želenih ciljev in to s čim manjšini stroški. Pri sprejemanju odločitev pa imajo informacije ključno in pomembno vlogo.

Če med seboj primerjamo dva managerja, ki sprejemata odločitve, bo tisti, ki ima na voljo kakovostne informacije lažje in boljše sprejel odločitev kot tisti, ki se odloča le na podlagi intuicije in brez kakovostnih informacij. Čim bolj realno sliko nam predstavljajo informacije, bolj so kakovostne in bolj jim lahko zaupamo. Kakovostne informacije so vedno dražje od manj kakovostnih, zato se v organizacijah srečujejo z dilemo, ali je korist od informacij večja, kot pa so stroški zanje. Pogosto je ceneje kupiti bolj kakovostne informacije kot pa sprejeti napačne odločitve (Pregled ponudbe poslovnih informacij, 2009).

Danes ni več problem v tem, da informacij ni na voljo. Pojavlja se druga težava: podatkov in informacij je preprosto preveč, z njimi smo tako preplavljeni, da moramo biti previdni in izbirčni pri njihovem zbiranju in obdelavi (Informacijska podpora odločanju, 2009).

Za uspešno odločanje v podjetju danes potrebujemo podatke, ki so pravi, to pa pomeni tudi ustrezno količino kakovostnih podatkov. Ponavadi je bilo mišljenje tako, da bodo odločitve veliko boljše, če bomo imeli na razpolago več podatkov in bomo med njimi izbirali. Prevelika količina podatkov v podatkovnem skladišču lahko tistega, ki se odloča, zmede, zato se ne mora odloči pravilno ali pa najboljše. Ravno tako pa je tudi v primeru, da imamo v podatkovnem skladišču preveliko zbirko podatkov, saj se iz njih lahko ne najdemo. Zagotoviti ustrezno količino kakovostnih podatkov za vse ravni odločanja je torej pomembna naloga vsakega podjetja.

Posledice odločitev, ki jih sprejemajo v podjetju, so lahko pozitivne ali negativne. Pozitivne so skladne s pričakovanji, pri čemer smo zagrabili in uporabili ponujeno priložnost, povečali vrednost in ugled podjetja, postavili temelje za osebno in družinsko kariero, povečali samozavest in zadovoljstvo udeležencev. Negativne odločitve pa imajo za svojo posledico razočaranje in izgubo samozavesti, izgubo časa in denarja, ugleda in samospoštovanja, neizkoriščene priložnosti drugod (oportunitetni stroški), propadli projekti in podjetja (Stemberger, 2003, str.14).

Sklep

V diplomski nalogi sem predstavila pomen in zagotavljanje kakovosti informacije pri podatkovnem skladiščenju. Pri tem sem ugotovila, da je področje kakovosti informacije v

današnjem času zelo napredovalo glede na preteklost. Vse več podjetij po svetu in tudi doma se zaveda, kako pomembno je, da imamo za prave odločitve tudi prave informacije. Od pravih odločitev pa je zelo odvisen uspeh in nadaljnje razvijanje podjetij. Saj le takrat, ko ima najvišje vodstvo kakovostne informacije na voljo v trenutku, ko jih potrebuje, lahko pravočasno zazna vse nepravilnosti in sproti rešuje težave.

Šele takrat, ko podatke prečistimo, integriramo ter jih uskladimo z obstoječimi poslovnimi pravili podjetja, jih lahko napolnimo v podatkovno skladišče, ki je jedro sistemov za podporo odločanju. Podatkovno skladišče mora vsebovati kakovostne podatke, ki jih potem uporabniki uporabljajo pri odločanju. Narobe je, če podatkovno skladišče ne vsebuje kakovostnih podatkov in se jih potem uporablja pri sprejemanju strateških odločitev. Odločitve, sprejete na podlagi takih podatkov, lahko podjetje pripeljejo tudi do propada.

Pri branju literature za diplomsko delo sem ugotovila, da veliko analitskih hiš po svetu napoveduje podatkovnim skladiščem še veliko več razvoja in pomembnosti kot ga imajo danes.

Ko hočemo narediti analizo kakovosti podatkov v našem podjetju, moramo najprej ugotoviti kakšno je stanje kakovosti podatkov ter določiti ustrezno raven kakovosti, ki jo hočemo doseči. Takrat ko kakovost podatkov zastavljenega cilja ne dosega, je potrebno podatke očistiti. Vendar pa kot sem že napisala, samo čiščenje ni dovolj, saj bodo podatki v tem primeru čez nekaj časa spet slabše kakovosti. Zato je nujno hkratno odpravljanje vzrokov nekovosti podatkov. Pregledati je potrebno postopke zajemanja, shranjevanja, obdelave podatkov ter prenosa v podatkovno skladišče. Določiti je potrebno skrbnike za kakovost podatkov, vzpostaviti ustrezne organizacijske predpise in izšolati uporabnike, kar je odvisno od najdenih vzrokov nekovostnih podatkov.

Kar se moramo zavedati pri vsej stvari je, da absolutne kakovosti ne moremo doseči, saj je ta predraga in lahko tudi nepotrebna. Ni nekega splošnega pravila o tem, koliko kakovosti je še smiselno zagotavljati v podjetju. To je odvisno je od tega kakšen je namen, tip podatkovnega skladišča in povratek investicije. V nekaterih primerih, glede na namen uporabe podatkov iz podatkovnega skladišča, ne potrebujemo zelo natančnih podatkov. V nekaterih primerih pa potrebujemo izredno natančno definirane podatke. Od vsega naštetega pa je ključnega pomena, da se organizacija zaveda pomena kakovosti podatkov in z njimi začne tudi ustrezno ravnati.

Literatura in viri

1. Adelman, S. & Moss, L. (2000). *Data warehouse project management*. Boston: Addison Wesley.
2. Berry, M. J. A. & Linoff, G. (2000). *Mastering data mining: The art and science of customer relationship management*. New York: J. Wiley.
3. Burns, L. (2005). *The ugly truth about data quality*. Najdeno 1. aprila 2009 na spletnem naslovu <http://www.information-management.com/specialreports/20050531/1028545-1.html>
4. Dvornik, U. (2007). *Sprotna analizična obdelava podatkov – OLAP*. Maribor: Ekonomsko – poslovna fakulteta.
5. English, L. (1999). *Improving data warehouse and business information quality: methods for reducing costs and increasing profits*. New York: Wiley.
6. Eppler, M. J. (2003). *Managing information quality : increasing the value of information in knowledge-intensive products and processes*. Berlin, New York: Springer.
7. *ETL, Ekstrakcija, transformacija in prenos podatkov [podjetja Nps – poslovne rešitve]*. Najdeno 2. oktobra 2008 na spletnem naslovu <http://www.nps.si/etl.html>
8. Ferle, M. (2005, september-oktober). INTERVJU: dr. Ralph Kimball. *Infosrc.si*. str. 5-7.
9. Gerkeš, M. *Kakovost informacij na WWW in eklektični dokumenti*. Najdeno 24. aprila 2009 na spletnem naslovu http://home.izum.si/cobiss/cobiss_obvestila/1999_1/html/clanek_02.html
10. Godnov, U. (2008, jul.-avg.). Čas je za kakovost podatkov. *Sistem*, str. 10-11
11. Golob I. & Walzer T. & Brumen B. (2003). Analiza in primerjava pristopov pri gradnji podatkovnih skladišč. Uporabna informatika.
12. Golob, I. & Welzer, T. *Arhitektura podatkovnih skladišč*. Najdeno 7. junija 2008 na spletnem naslovu http://www.drustvo-informatika.si/dogodki/arhiv/dsi2001/sekcija_a/golob_welzer.doc
13. Golob, I. (2001). *Arhitecture podatkovnih skladišč* (magistrsko delo). Maribor: Ekonomska fakulteta.
14. Gradišar, M. & Resinovič, G. (1998). *Informatika v organizaciji*. Kranj: Moderna organizacija.
15. Greenfield, L. *What Data Errors You May Find When Building a Data Warehouse*. Najdeno 24. marca 2009 na spletnem naslovu <http://www.dwinfocenter.org/errors.html>
16. Horjak M. *Vpliv varne informacijske tehnologije na ekonomsko uspešnost podjetja*. Najdeno 2. maja 2009 na spletnem naslovu http://www.mfc2.si/pdf/horjak_vpliv_varne_informacijske_tehnologije_na_ekonomsko_uspesnost_podjetja.pdf
17. Informacijska podpora odločanje, drugi del. Najdeno 4. marec 2009 na spletnem naslovu http://www.ujemi-znanje.si/vsebina/knjiga_odlocanje1.pdf
18. Imhoff, C. (2001). *Building the Customer Centric Enterprise*. New York: John Wiley & Sons Inc.

19. Inmon, W.H. (1998). *Corporate Information Factory*. New York: John Wiley & Sons Inc.
20. Integracija podatkov- ključni proces BI projekta [BILAB d.o.o.]. Najdeno 13. decembra 2008 na spletnem naslovu <http://www.bilab.si/?show=content&id=12&men=16&oce=13>
21. Islovar. Najdeno 14. maja 2009 na spletnem naslovu http://www.islovar.org/iskanje_enostavno.asp
22. Jagarinec, D. (2005, Oktober). OLAP in podatkovna skladišča. *Moj Mikro*. Najdeno 8. junija 2008 na spletnem naslovu http://www.bilab.si/uploads/clanki/arhivirana_datoteka_5.pdf
23. Jaklič, J. (2002). *Upravljanje in uporaba podatkov*. Ljubljana: Ekonomska fakulteta.
24. Jaklič, J. (2003). *Upravljanje podatkov. Gradiva pri predmetu Baze podatkov in njihova uporaba*. Ljubljana: Ekonomska fakulteta.
25. Jaklič, T. (2001). Podatkovna skladišča v finančnih institucijah. *Infosrc.si*, str. 3-5.
26. Kimball, R. (1996a). *Building The Data Warehouse second edition*. New York: John Wiley & Sons.
27. Kimball, R. (1996b). *The Data Warehouse Toolkit*. New York: John Wiley & Sons.
28. Kimball, R. & Ross, M. (2002) *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. 2nd Edition*. New York: Wiley
29. Krsnik, J. (2008). Sprotno analitično procesiranje – OLAP. Najdeno 4. oktobra 2008 na spletnem naslovu http://www.bfro.uni-lj.si/Kat_center/objave/objave.htm
30. Lalič R. (2005). Vpeljava sistema OLAP v podporo obveščevalnemu in analitičnemu delu policije na področju prepovedanih drog. Ljubljana: Ekonomska fakulteta.
31. Larry, B (2005, maj). The Ugly Truth about Data Qality. *Information Management Speciak Reports*. Najdeno 23. aprila 2009 na spletnem naslovu <http://www.information-management.com/specialreports/20050531/1028545-1.html>
32. Marolt, J. (1994). *Menedžment in tehnologija zagotavljanja kvalitete*. Kranj: Moderna organizacija.
33. Marolt, J. (2006, nov.-dec.). ABC skladiščenja podatkov iz poslovnega obveščanja. *Infosrc.si*, str. 20-23.
34. Milavec, H. T. (2008, št.54). Izgradnja podatkovnega skladišča. *Infosrc.si*, str. 28-34.
35. Milavec, H.T. (2006, 14.marec). Zakaj banke potrebujejo podatkovno skladišče. *Info.SRC.si*, str. 9-11.
36. Mohorič, T. (1999, okt.-nov.). O podatku in informaciji. *Organizacija, Fakulteta za računalništvo in informatiko*, str. 445-448.
37. Podatkovna skladišča [BILAB d.o.o.]. Najdeno 13. decembra 2008 na spletnem naslovu <http://www.bilab.si/?show=content&id=5&men=10&oce=6>
38. Podatkovna skladišča: Želite prehiteti konkurenco in povečati dobiček [Ixtan forum d.o.o.]. Najdeno 10. septembra 2008 na spletnem naslovu <http://www.ixtlan.si/Razvoj.asp?start=2>
39. Pregled ponudbe poslovnih informacij [Gospodarska Zbornica Slovenije]. Najdeno 4. marec 2009 na spletnem naslovu <http://www.gzs.si/slo/panoge/21371>

40. Slovar informatike. Najdeno 3. maja 2009 na spletnem naslovu http://www.islovar.org/iskanje_enostavno.asp?reset=true
41. Stemberger M. (2003). Prilagodljivost organizacije, procesov in delovnih tokov kot odziv na spremembe iz okolja. Ljubljana: Fakulteta za računalništvo in informatiko.
42. TIQM, Quality System. Najdeno 7. Junija 2009 na spletnem naslovu <http://www.infoimpact.com/tiqmmethodology.cfm>
43. Tegeltija, B. (2008). Uporabite moč informacije. *Infosrc.si*, str. 21-23.
44. Welzer, T. (1998). Zbornik posvetovanja Dnevi slovenske informatike Portorož 1998. *Kakovost podatkov in meta podatkovni model* (stran 72). Ljubljana: Slovensko društvo informatika.
45. White, C. (2000, Marec). The federated Data Warehouse. *DM Review Magazine*. Najdeno 8. junija 2008 na spletnem naslovu <http://www.dmreview.com/issues/20000301/1953-1.html>
46. *Zakaj je kakovost podatkov tako pomembna?* [Inštitut za kakovost podatkov]. Najdeno 12. decembra 2008 na spletnem naslovu <http://www.ikp-dqi.com>
47. Žalik Rizman, K. (2004). Podatkovna skladišča in kakovost podatkov. *Uporabna informatika*, str. 19-29.