

**UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO**

**VALTER ŠORLI**

**UPRAVLJANJE MATIČNIH PODATKOV – INTEGRACIJA  
PODATKOV O STRANKAH**

**MAGISTRSKO DELO**

Mentor: prof. dr. Viljan Mahnič

Ljubljana, 2014



Št.: 125-MAG-ISO/2012

Datum: 04. 09. 2012



Valter ŠORLI, univ. dipl. inž. rač. in inf.

## L j u b l j a n a

Fakulteta za računalništvo in informatiko Univerze v Ljubljani izdaja naslednjo magistrsko nalogo

Naslov naloge: **Upravljanje matičnih podatkov – integracija podatkov o strankah**

**Master data management – customer data integration**

Tematika naloge:

Za uspešno poslovanje vsake organizacije sta kakovost in urejenost podatkov bistvenega pomena. Z upravljanjem matičnih podatkov (angl. *master data management* – MDM) želimo zagotoviti, da organizacija ne uporablja podvojenih in pogosto nekonsistentnih verzij istih matičnih podatkov, ki se lahko pojavijo v različnih delih njenega informacijskega sistema, tako da skušamo povezati zapise, ki pripadajo isti entiteti. Ker je upravljanje matičnih podatkov zahteven proces in vsi podatki niso enako pomembni, se pogosto osredotočimo na tiste, ki so za organizacijo najpomembnejši. Največkrat so to podatki o strankah.


V magistrski nalogi opredelite, kaj so matični podatki in predstavite koncept njihovega upravljanja s posebnim poudarkom na integraciji podatkov o strankah. Preučite tri možne načine implementacije upravljanja z matičnimi podatki: registrski, hibridni in transakcijski, izvedite vrednotenje in naredite medsebojno primerjavo. Umestite MDM v arhitekturo informacijskega sistema, preučite mehanizme zagotavljanja kakovosti podatkov in načine uparjanja na primeru uporabniške domene. Za MDM domeno uporabnika pripravite pregled obstoječih komercialnih rešitev in njihov položaj na trgu.

Mentor:

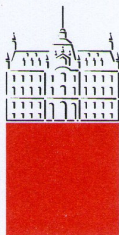
  
prof. dr. Viljan Mahnič



Dekan:

  
prof. dr. Nikolaj Zimic

Št.: 125-MAG-ISO/2012  
Datum: 04. 09. 2012



**Valter ŠORLI**, univ. dipl. inž. rač. in inf.

### Ljubljana

Fakulteta za računalništvo in informatiko Univerze v Ljubljani izdaja naslednjo magistrsko nalogo

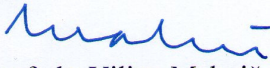
Naslov naloge: **Upravljanje matičnih podatkov – integracija podatkov o strankah**  
**Master data management – customer data integration**

Tematika naloge:

Za uspešno poslovanje vsake organizacije sta kakovost in urejenost podatkov bistvenega pomena. Z upravljanjem matičnih podatkov (angl. *master data management* – MDM) želimo zagotoviti, da organizacija ne uporablja podvojenih in pogosto nekonsistentnih verzij istih matičnih podatkov, ki se lahko pojavijo v različnih delih njenega informacijskega sistema, tako da skušamo povezati zapise, ki pripadajo isti entiteti. Ker je upravljanje matičnih podatkov zahteven proces in vsi podatki niso enako pomembni, se pogosto osredotočimo na tiste, ki so za organizacijo najpomembnejši. Največkrat so to podatki o strankah.


V magistrski nalogi opredelite, kaj so matični podatki in predstavite koncept njihovega upravljanja s posebnim poudarkom na integraciji podatkov o strankah. Preučite tri možne načine implementacije upravljanja z matičnimi podatki: registrski, hibridni in transakcijski, izvedite vrednotenje in naredite medsebojno primerjavo. Umestite MDM v arhitekturo informacijskega sistema, preučite mehanizme zagotavljanja kakovosti podatkov in načine uparjanja na primeru uporabniške domene. Za MDM domeno uporabnika pripravite pregled obstoječih komercialnih rešitev in njihov položaj na trgu.

Mentor:

  
prof. dr. Viljan Mahnič



Dekan:

  
prof. dr. Nikolaj Zimic



# IZJAVA O AVTORSTVU

## magistrskega dela

Spodaj podpisani Valter Šorli z vpisno številko 63000289, sem avtor magistrskega dela z naslovom »Upravljanje matičnih podatkov – integracija podatkov o strankah«.

S svojim podpisom zagotavljam, da:

- sem magistrsko delo izdelal samostojno pod vodstvom mentorja prof. dr. Viljana Mahnič
- so elektronska oblika magistrskega dela, naslova (slov., angl.), povzetka (slov., angl.) ter ključne besede (slov., angl.) identični s tiskano obliko magistrskega dela
- in soglašam z javno objavo elektronske oblike magistrskega dela v zbirki »Dela FRI«.

V Ljubljani, dne 01.09.2014

Podpis avtorja \_\_\_\_\_



# Zahvala

Zahvaljujem se mentorju dr. Viljanu Mahničju, ki me je z dragocenimi nasveti vodil skozi obdobje izdelave tega magistrskega dela.

Hvala Andreji za potrpežljivost in podporo v času mojega ustvarjanja.



# Kazalo

Povzetek .....	1
Abstract.....	3
Uvod .....	5
1   Matični podatki.....	9
1.1   Kaj so matični podatki .....	9
1.2   Kako pridemo do matičnih podatkov?.....	12
1.2.1   Od zgoraj navzdol.....	12
1.2.2   Od spodaj navzgor .....	13
1.2.3   Razpoznavanje metapodatkov .....	13
1.3   Kaj je upravljanje matičnih podatkov .....	14
1.4   Upravljanje matičnih podatkov o strankah .....	15
2   Sistem upravljanja matičnih podatkov .....	17
2.1   Prednosti upravljanja matičnih podatkov .....	17
2.2   Umestitev in arhitektura.....	20
2.3   Razsežnosti upravljanja matičnih podatkov .....	22
2.3.1   Domena.....	22
2.3.2   Metode uporabe .....	23
2.3.3   Načini implementacije.....	25
2.3.4   Izbira načina implementacije.....	34
2.4   Postopek združevanja zapisov strank .....	40
2.4.1   Skupine atributov.....	41
2.4.2   Primerjanje zapisov in njihovih atributov .....	41
2.4.3   Uparjanje na primeru .....	44
2.5   Zagotavljanje kakovosti podatkov .....	48
2.6   Obvladovanje podatkov .....	50
2.6.1   Skupina za obvladovanje podatkov .....	51
2.6.2   Stopnje zrelosti obvladovanja upravljanja matičnih podatkov.....	52
3   Vpeljava projekta.....	59
3.1   Vzpostavitev delovne skupine .....	60
3.2   Izbira metodologije .....	63
4   Pregled ponudbe na trgu .....	69
4.1   Poročili tržnih raziskav .....	69
4.2   Lastnosti vodilnih ponudnikov .....	71

5	Sklep.....	75
6	Priloge .....	79
7	Literatura.....	83

## Kazalo tabel

Tabela 1: Primeri matičnih podatkov. ....	11
Tabela 2: Primerjava med načini implementacije. ....	35
Tabela 3: Ocenjene vrednosti pomembnosti. ....	39
Tabela 4: Minimalni pogoji avtomatskega združevanja. ....	40
Tabela 5: Primeri atributnih uteži. ....	46
Tabela 6: Rezultati izračuna podobnosti zapisov. ....	47
Tabela 7: Predloga plana vpeljave upravljanja matičnih podatkov. ....	61

## Kazalo slik

Slika 1: Postopek iskanja matičnih podatkov. ....	12
Slika 2: Prekrivanje matičnih podatkov o strankah. ....	16
Slika 3: Deleži rešitev za upravljanje matičnih podatkov. ....	17
Slika 4: Umestitev v informacijski sistem. ....	20
Slika 5: Arhitektura sistema za upravljanje matičnih podatkov. ....	21
Slika 6: Dimenzije sistema za upravljanje matičnih podatkov. ....	22
Slika 7: Registrski način implementacije. ....	26
Slika 8: Registrski način implementacije z oznako vira. ....	27
Slika 9: Transakcijski način implementacije. ....	30
Slika 10: Spekter načinov implementacije. ....	36
Slika 11: Zvezdni diagram aplikacije poročanja. ....	39
Slika 13: Zvezdni diagram aplikacije CRM. ....	39
Slika 12: Zvezdni diagram aplikacije HRM. ....	39
Slika 14: Razdelitev ocen primerjanja zapisov z dvojnimi pragom. ....	44
Slika 15: Primer iskanja ujemanj zapisov. Množici zapisov K in M. ....	45
Slika 16: Izračunane ocene glede na mejne vrednosti. ....	47
Slika 17: Postopek izboljševanja kakovosti podatkov. ....	50
Slika 18: Hierarhija upravljanja s podatki. ....	51
Slika 19: Štirje pristopi vpeljave projekta upravljanja matičnih podatkov. ....	64
Slika 20: Metoda za integrirano okolje znanja. ....	66
Slika 21: Forrester diagram za upravljanje matičnih podatkov. ....	69
Slika 22: Gartnerjev diagram za upravljanje matičnih podatkov. ....	70



# Seznam uporabljenih kratic

BI – poslovno obveščanje (ang. *business intelligence* )

BPM – modeliranje procesov (ang. *business process modeling*)

CDI – integracija podatkov o strankah (ang. *customer data integration*)

CRM – upravljanje odnosov s strankami (ang. *customer relationship management*)

DQ – kakovost podatkov (ang. *data quality* )

DWH – podatkovno skladiščenje (ang. *data warehousing*)

ERP – celovita programska rešitev (ang. *enterprise resource planning*)

HRM – upravljanje človeških virov (ang. *human resource management*)

IS – informacijski sistem (ang. *information system*)

IT – informacijska tehnologija (ang. *information technology*)

MDM – upravljanje z matičnimi podatki (ang. *master data management*)

MIKE – metoda za integrirano znanje okolja (ang. *method for an integrated knowledge environment*)

PIM – upravljanje informacij o izdelkih (ang. *product information management*)

SOA – storitveno usmerjena arhitektura (ang. *service oriented architecture*)



# Povzetek

V magistrskem delu se ukvarjamo s problematiko upravljanja matičnih podatkov o strankah. Organizacije se pri svojem delovanju pogosto srečujejo z nekonsistentnimi podatki, ki so razpršeni po različnih silosnih aplikacijah. Ker silosi živijo vsak svoje življenje, so matični podatki v njih medsebojno neuskklajeni, poslovni uporabniki pa pogosto ne vedo, katera kopija odraža zadnje veljavno stanje. Z željo po dvigu kakovosti matičnih podatkov o strankah, je potrebno vzpostaviti sistem za njihovo upravljanje.

Namen dela je opisati in predstaviti lastnosti takšnih sistemov. Uveljavile so se tri metode uporabe sistema za upravljanje: sodelovalna, operativna in analitična, ki poleg štirih načinov implementacije, opredeljujejo sistem za upravljanje matičnih podatkov o strankah. V delu obravnavamo vse štiri načine implementacije, ki bistveno določajo lastnosti vozlišča MDM. Opisano je delovanje registrskega, usklajevalnega, kombiniranega in transakcijskega načina, izvedena pa je tudi njihova medsebojna primerjava. Predstavljena je metoda za ugotavljanje stopnje zrelosti upravljanja matičnih podatkov, ki organizaciji služi kot merilo kakovosti trenutne rešitve in predlaga nadaljnje korake za izboljšave.

Projekt vzpostavitve sistema za upravljanje je izjemno obsežen in drag proces, ki zaradi poseganja v poslovne procese organizacije s sabo prinaša tudi določena tveganja. V delu predlagamo metodologijo razvoja in vodenja projekta, podane pa so tudi usmeritve za vodstveno skupino, ki bo projekt vodila.

Na trgu se pojavlja veliko ponudnikov rešitev za upravljanje matičnih podatkov o strankah. Da organizacija lažje izbere potencialne kandidate in zoži seznam ponudnikov, predstavljamo izsledke dveh analitskih hiš, ki periodično spremljata ponudbo na segmentu upravljanja matičnih podatkov o strankah. Za vodilnih pet ponudnikov navajamo glavne prednosti in slabosti njihovih rešitev.

V delu so na enem mestu zbrane informacije, ki organizaciji koristijo pri vpeljavi sistema za upravljanje matičnih podatkov o strankah.

Ključne besede: upravljanje matičnih podatkov, integracija podatkov o strankah, vodenje podatkov, kvaliteta podatkov.



# Abstract

In this Master's thesis we deal with the problem of managing customer master data. Organizations are often faced with inconsistent data, which are scattered throughout the various silos applications. Since silos are living their own lives, the master data in them remains uncoordinated and business users often do not know which copy reflects the latest valid state. With a desire to raise the quality of customer master data, it is necessary to establish a system to manage them.

The purpose of this work is to present and describe the properties of such systems. We introduce three usage methods of systems for management: collaborative, operational and analytical, which in addition to the four modes of implementation, define a system for managing customer master data. Considered are all four methods of implementation that significantly determine the properties of MDM hub. We describe the operation of the registry, consolidation, transactional and coexistence hub, and perform their mutual comparison. A method for determining the degree of maturity of master data management is presented. An organization can use it as a measure of the quality of its current solution and for seeking of further steps for improvement.

The project of establishing a system of governance is extremely extensive and expensive process which, due to interference with the business processes of the organization, involves certain risks. In this work, we propose an appropriate methodology for development and project management, additionally there are given some guidelines for management team that will lead the project.

On the market there are many providers of customer master data management solutions. To narrow the list of bidders and facilitate the organization selection, we present the results of two analytical houses, which periodically monitor the supply of the customer master data management segment. For five of the leading providers we sets out the main advantages and disadvantages of their solutions.

Information that can be beneficial for an organization which is establishing a system for customer master data management, are in this work gathered on one place.

Key words: master data management, customer data integration, data governance, data quality.

# Uvod

Organizacije se v svojem razvoju, bodisi zaradi svoje rasti, bodisi zaradi pripojitev drugih organizacij, pogosto znajdejo v situaciji, ko se isti podatki o strankah podvajajo v različnih sistemih [14]. To predstavlja velike težave, saj ti podatki živijo ločeno življenje in se medsebojno ne usklajujejo. Slednje privede do situacije, ko poslovni uporabniki ne vedo več, kateri podatki prikazujejo dejansko stanje.

Vse to velja tudi za podatke o strankah nekega podjetja. Ker so kritični podatki o strankah raztreseni po različnih sistemih in so medsebojno različni, se pojavljajo problemi. S takimi podatki je težko upravljati, hkrati pa jim ne moremo nikoli popolnoma zaupati. Samo zaupanja vredni podatki so lahko podlaga za sprejemanje boljših odločitev.

Za reševanje težav s slabimi podatki obstaja več pristopov. Gre za najrazličnejše postopke profiliranja, čiščenja in vodenja podatkov, ki jih skupaj združimo v ogrodje za upravljanje matičnih podatkov (angl. *master data management* – MDM). MDM torej sestoji iz procesov in tehnologij, cilj pa je vzpostaviti in zagotavljati avtorizirano, zanesljivo, trajnostno, natančno in varno podatkovno okolje, ki predstavlja »enotno različico resnice«. V to okolje sprejeti sistem zapisov se uporablja v celotni organizaciji, v vseh množicah aplikacijskih sistemov, poslovnih linijah in uporabniških skupnostih [3, 15].

MDM se tako nahaja v središču informacijskega sistema organizacije. Je torej ključna točka, ki skrbi za konsistentnost in poenotenje izbranih ključnih matičnih podatkov.

Če hočemo matične podatke upravljati, jih je potrebno v množici podatkov najprej razpoznati. Prvo poglavje je zato namenjeno opredelitvi matičnih podatkov. Opisali bomo njihove lastnosti in navedli po čem se razlikujejo od drugih skupin podatkov. Podali bomo dva pristopa za njihovo razpoznavanje v informacijskem sistemu (ang. *information system* - IS) organizacije.

Ker so prav matični podatki za organizacijo najvrednejši [15], jih je smiselno upravljati. Opredelili bomo, kaj je upravljanje matičnih podatkov (ang. *master data management* – MDM) in kdaj v okviru MDM govorimo o integraciji podatkov o strankah (angl. *customer data integration* – CDI). Prav MDM-CDI nas v tem magistrskem delu najbolj zanima, zato se vsi primeri navezujejo ravno na domeno stranke, ki jo bomo tudi podrobneje opisali.

Drugo poglavje je namenjeno podrobnejšemu opisu sistema MDM. Najprej bomo predstavili nekatere koristi, ki jih je deležna organizacija v primeru vzpostavitve sistema MDM. Nato bomo sistem MDM umestili v IS organizacije in opisali njegovo arhitekturo. Kot bomo videli v nadaljevanju, se vsaka implementacija MDM giblje v treh razsežnostih [6, 15]: razsežnost domene, metode uporabe in načina implementacije. Ker so to bistveni dejavniki, ki imajo močan vpliv na nadaljnje poslovanje organizacije, bomo vsako izmed razsežnosti podrobneje opisali [3]. Poseben poudarek bomo dali na načine implementacije, ki jih literatura pojmuje kot arhitekturne sloge [3] in izvedli njihovo medsebojno primerjavo. Kot pomoč za izbiro ustreznega načina implementacije bomo predlagali vizualizacijo nad že znanimi kriteriji [15], ki imajo vpliv na izbiro.

Pomemben del sistema MDM je način združevanja zapisov, ki v MDM vstopajo iz zunanjih aplikacij. Na konkretnem primeru bomo prikazali postopek združevanja zapisov strank, opisali možne realizacije primerjalne funkcije in avtomatizacije njenega delovanja z uporabo pragov (ang. *threshold*).

V sistemu MDM imata mesto tudi kakovost podatkov (ang. *data quality* – DQ) in obvladovanje podatkov (angl. *data governance*). Čeprav sta del arhitekture MDM lahko rečemo, da zaobjemata politike MDM [5]. Podali bomo nekaj metrik za spremljanje kakovosti podatkov in jih zapeljali v proces, ki organizaciji omogoča iterativno izboljševanje kakovosti le-teh. Organizaciji je, pri določevanju nivoja na katerem se nahaja njeno obvladovanje podatkov, v pomoč model štirih stopenj zrelosti vodenja MDM [23]. Vsako stopnjo opredeljuje štiri komponente: ljudje, politike, tehnologije ter tveganja in koristi. Dodatno model določa še posebne kriterije, ki jih je potrebno doseči za prehod v višjo stopnjo zrelosti. Ko organizacija pozna svoje trenutno stanje in si zastavi cilj, ki ga želi doseči, ima pred sabo jasno zarisano pot, katero mora prehoditi za dosego le-tega.

Ker je uvedba MDM v IS organizacije velik podvig, bomo v tretjem poglavju govorili o vpeljavi projekta. Glede na negativne posledice, ki jih za organizacijo prinaša neuspeh projekta [3], bomo okoli nekaj točk najprej strnili potrebne razmisleke vodstva. Ne gre zanemariti, da so predvsem oni nosilci odgovornosti in zato glede uvedbe MDM v poslovanje ne morejo in ne smejo ostati neopredeljeni. Podpora vodstva je namreč eden ključnih faktorjev za premik organizacije in dosedanjih načinov dela na višjo stopnjo.

Za namen vpeljave bomo predlagali vzpostavitev delovne skupine, katere naloga je vodenje projekta in usklajevanje predvsem poslovnih uporabnikov z osebjem oddelka IT. Določili

bomo poglobitne naloge, ki naj bi jih skupina opravljala in podali nekaj začetnih korakov, ki naj se izvejo na začetku projekta. Glede na lastnosti projektov MDM, bomo predstavili štiri pristope vpeljave projekta MDM. Pogledali bomo njihove glavne značilnosti, v nadaljevanju pa predlagali konkretno metodologijo katere se organizacija lahko posluži za vodenje in implementacijo projekta.

Zaradi pomanjkanja usposobljenega kadra se organizacije redko odločijo za izdelavo lastne rešitve MDM-CDI, zato bomo predstavili nekaj glavnih komercialnih ponudnikov. Glede na pričakovano visoko rast segmenta MDM-CDI, ta naj bi do leta 2015 dosegel vrednost 3,2 milijarde dolarjev [25], se na trgu pojavlja veliko manjših ponudnikov. Velja opozoriti, da uvedba MDM v poslovanje ni prehodne narave, ampak bistveno poseže v poslovne procese in spremeni način delovanja organizacije. To nam daje slutiti, da ima izbira pravega ponudnika rešitve MDM velik pomen na bodoče poslovanje.

Ker se želimo izogniti situaciji, da bi nas morebiti trenutna zadovoljiva rešitev ovirala v prihodnosti, je organizacija primorana izbrati dolgoročnega partnerja z vizijo. Tukaj so nam lahko v pomoč poročila analitskih hiš, ki periodično spremljajo dogajanje na trgu segmenta MDM-CDI. V tem delu bomo predstavili izsledke dveh, in sicer Forrester in Gartner. Njuni diagrami podajajo informacijo o položaju rešitve posameznih ponudnikov in organizaciji omogočajo izbiro tistih, ki s svojo ponudbo predstavljajo verodostojne partnerje.

V magistrski nalogi so na enem mestu zbrane poglobitne informacije, ki organizaciji nudijo vpogled v področje upravljanja matičnih podatkov. Opredeljeni so osnovni pojmi, prikazani načini delovanja različnih konceptov MDM, podani napotki za vpeljavo projekta in pripravljen pregled glavnih ponudnikov programskih paketov MDM-CDI. Delo lahko služi kot pomoč organizaciji, ki se v svojem poslovanju sooča s težavami z naslova slabega upravljanja matičnih podatkov.



# 1 Matični podatki

Da bi razumeli upravljanje matičnih podatkov, je potrebno najprej opredeliti, kaj so matični podatki, in povedati, kako jih razlikujemo od preostalih podatkov v informacijskem sistemu (ang. *information system* - IS) organizacije.

## 1.1 Kaj so matični podatki

Kadar spremljamo poslovanje organizacij, ne moremo mimo entitet, kot so kupci, posojila, artikli, dobavitelji, pacienti, kupoprodajne pogodbe itd. Uporaba tovrstnih entitet je široko razširjena po različnih sistemih v organizaciji in predstavlja matične podatke te organizacije. Kot bomo videli v nadaljevanju, lahko ključne entitete, glede na njihove lastnosti, grupiramo v nekaj področij ali domen kot na primer: stranka, produkt, račun in lokacija [3].

V literaturi se pogosto pojavlja naslednja definicija matičnih podatkov: »Objekti matičnih podatkov so tisti jedrni poslovni objekti, ki se skupaj z njihovimi metapodatki, atributi, definicijami, vlogami, povezavami in hierarhijami uporabljajo v različnih aplikacijah v organizaciji. So tisti objekti, ki so za organizacijo najbolj pomembni« [15, str. 6].

Ko govorimo o kategorizaciji podatkov, se navadno srečujemo z mnogo različnimi pogledi in delitvami. Če za trenutek ob strani pustimo metapodatke, lahko podatke v grobem ločimo v tri skupine [4]:

- **Transakcijski podatki:** so posledica izvajanja poslovnih transakcij. Tipični predstavniki te skupine so naročila, plačila itd. Količina transakcijskih podatkov močno narašča s povečevanjem obsega poslovanja organizacije. Če na primer banka proda N plačilnih kartic, se bo dnevni obseg transakcij povečal za približno N krat povprečno število dnevni transakcij na komitenta.
- **Referenčni podatki:** predstavljajo usklajen nabor vrednosti, kot so oznake: stanj, valut, spola itd., njihova vloga pa je zagotavljanje konsistentnosti vrednosti atributov. Smiselno je, da so referenčni podatki poenoteni na nivoju celotne organizacije, saj je njihova naloga okarakterizirati druge podatke. Podatki so shranjeni v šifrantih in se redko spreminjajo. Če se navežemo na primer iz prejšnje točke, se za N prodanih kartic, ki predstavljajo obstoječe produkte banke, količina referenčnih podatkov ne poveča.

- Matični podatki: se nanašajo na entitete, ki so jedro poslovanja in jih organizacija uporablja preko različnih poslovnih procesov in sistemov. V našem primeru bi se količina matičnih podatkov povečala za število novih komitentov.

Matični podatki nastopajo tako v transakcijskih sistemih, ki so posledica izvajanja poslovnih procesov, kot tudi v analitičnih sistemih, ki vodstvu organizacij predstavljajo podlago za sprejemanje odločitev. Matični podatki torej ne predstavljajo zgolj poenotenega pogleda, ki ga je potrebno upravljati, ampak morajo odražati tudi pravilen pogled na informacije v določenem trenutku. Slednje je mogoče doseči samo na način, da oba omenjena sistema delujeta nad istim naborom matičnih podatkov. Pogosto se srečujemo z aplikacijami, ki vsebujejo vsaka svojo kopijo podatkov, ki taki aplikaciji omogočajo delovanje. Za take aplikacije se je uveljavil izraz silosno organizirane aplikacije (ang. *silos application*). Čeprav silosno organizirane aplikacije omogočajo »lokalno« upravljanje z matičnimi podatki, imamo še vedno problem z usklajevanjem matičnih podatkov med aplikacijami. Želja je, da imamo en sam repozitorij, ki je konsolidiran, avtomatiziran in upravljan. To dosežemo tako, da omenjene ključne podatke izvzamemo iz okrilja silosnih aplikacij, ter jih prenesemo v ločen sistem. Da bi ta cilj dosegli, moramo matične podatke najprej prepoznati in jih ločiti od drugih podatkov [3].

»Matične podatke lahko določimo kot podatke, ki so bili očiščeni, racionalizirani in integrirani v sistem zapisa (ang. *system of record*). Tak zapis je veljaven za aktivnosti jedrnega poslovanja na nivoju celotne organizacije« [3, str. 8].

Ker se večina najpomembnejših poslovnih funkcij v podjetju vrti prav okoli matičnih podatkov, lahko rečemo, da so matični podatki med najbolj dragocenimi informacijami, ki jih ima organizacija v lasti [6]. Predstavljajo jedrne informacije o poslovanju kot so: stranke, dobavitelji, produkti, računi, ter razmerja med njimi, brez katerih si poslovanja praktično ne moremo predstavljati.

Sedaj, ko smo določili, kaj so matični podatki, si pogledjmo katere karakteristike jih opredeljujejo:

- Konstantna količina podatkov  
Kot smo že navedli ob kategorizaciji podatkov, se obseg matičnih podatkov skozi čas ne spreminja bistveno, medtem ko se obseg transakcijskih podatkov s poslovno aktivnostjo močno povečuje [15]. Če kot primer vzamemo klice mobilnega operaterja, nam postane hitro jasno, da pri rahli rasti števila naročnikov, število opravljenih klicev navadno močno narašča.

- **Obstajajo neodvisno**  
Za razliko od transakcijskih podatkov, so objekti matičnih podatkov neodvisni od ostalih objektov [6]. Na primeru iz prejšnje alineje je jasno, da klicatelj in klicani lahko obstajata samostojno, medtem ko klic, kot tipični transakcijski objekt, sam brez njiju ne more obstajati.
- **Nizka frekvenca spreminjanja**  
Objekti matičnih podatkov so, v primerjavi s transakcijskimi podatki, v manjši meri podvrženi spreminjanju. Večina atributov objektov matičnih podatkov ostaja skozi življenjski cikel objekta nespremenjenih oz. se spreminjajo zelo poredko [15]. V kolikor pogledamo elektronsko sporočilo, se njegov status in vsebina močno spreminjata med prehajanjem od novega sporočila, preko predloge in sporočila v pošiljanju, do poslanega in prebranega sporočila. Na drugi strani imamo pošiljateljev račun, ki v svojem življenjskem ciklu ostaja bolj kot ne nespremenjen.

Iz zgoraj navedenega lahko povzamemo, da so matični podatki tisti podatki, ki opisujejo glavne entitete organizacije, se uporabljajo v različnih poslovnih procesih v podjetju in se ob prehodu skozi transakcije redko spreminjajo. Gre za množico podatkov, ki se uporabljajo s strani več aplikacij.

Kot primer si pogledajmo matične podatke naslednje transakcije: Janez Novak, z naročniškim paketom Povezani, je dne 1.4.2013 ob 10:00 z bazne postaje MB264, v 6 minutnem pogovoru navezal stik z Manco Novak na bazni postaji LJ341. Tabela 1 predstavlja razčlenitev transakcije na posamezne matične podatke, ki so razvrščeni v pripadajoče domene.

**Tabela 1: Primeri matičnih podatkov.**

<b>Domena matičnega podatka</b>	<b>Matični podatek</b>
Stranka	Janez Novak
Stranka	Manca Novak
Produkt	naročniški paket Povezani
Lokacija	bazna postaja LJ341
Lokacija	bazna postaja MB264

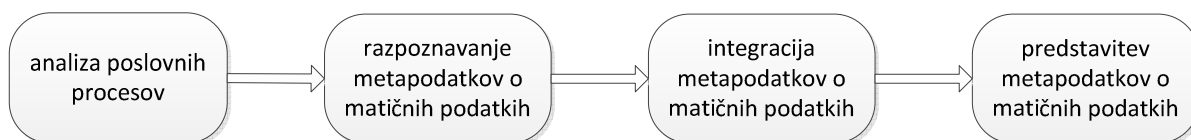
Iz navedenega primera lahko hitro sklepamo, da izluščeni matični podatki izpolnjujejo vse njihove značilnosti. Obe stranki, produkt in lokaciji obstajajo neodvisno eden od drugega in od same transakcije. Prav tako se matični podatki skozi omenjeno transakcijo ne spreminjajo, niti se ne porajajo novi.

## 1.2 Kako pridemo do matičnih podatkov?

Spoznali smo lastnosti matičnih podatkov, na tem mestu pa si pogledjmo, kakšni pristopi so nam na voljo pri razpoznavanju matičnih podatkov. Kadar se določen podatek pojavi v 80 odstotkih poslovnih aplikacij, takrat z veliko verjetnostjo lahko trdimo, da gre za matični podatek [2]. V nasprotnem primeru za podatek, ki se pojavlja zgolj v 20 odstotkih poslovnih aplikacij lahko trdimo, da ne gre za matični podatek. Za razpoznavanje bomo seveda potrebovali nekoliko bolj natančne postopke, zato si pogledjmo dva koncepta, in sicer od spodaj navzgor ter od zgoraj navzdol.

### 1.2.1 Od zgoraj navzdol

Poglejmo si najprej postopek iskanja matičnih podatkov od zgoraj navzdol. Slika 1 prikazuje postopek iskanja, ki ga lahko razčlenimo na štiri stopnje: analiza poslovnih procesov, razpoznavanje metapodatkov o matičnih podatkih, integracija metapodatkov o matičnih podatkih in predstavitev metapodatkov o matičnih podatkih [19].



**Slika 1: Postopek iskanja matičnih podatkov.**

Avtorji navajajo, da si postopek iskanja matičnih podatkov lahko predstavljamo kot izgradnjo novega informacijskega sistema (IS), ki je namenjen izključno upravljanju z matičnimi podatki drugih operativnih IS. V stopnji analize poslovnih procesov je potrebno določiti podatke, ki podpirajo poslovni proces in jih zmodelirati v enem od standardiziranih jezikov za modeliranje. Zanimajo nas podatki, ki nastopajo v več kot enem poslovnem procesu in so osnovani nad ključnimi entitetami. Ti so dobri kandidati, da jih poslovni analitiki določijo kot matične podatke. V drugi stopnji je izmed podatkov potrebno pridobiti metapodatke o matičnih podatkih. Metapodatki o matičnih podatkih so podatki, ki opisujejo matične podatke. Gre za opis njihovega poslovnega pomena, format, tip, dovoljen nabor vrednosti itd. Za ta namen uporabimo pravilo za prepoznavanje, ki bo opisano v nadaljevanju. V naslednji stopnji poskrbimo za integracijo metapodatkov iz vseh poslovnih procesov. Tako zagotovimo, da bo končni MDM sistem en silos, ki ga umestimo v IS organizacije. Vsak matični podatek mora biti jasno opredeljen z metapodatki, ki ga določajo na nivoju celotne organizacije. Sedaj nam ostane samo še predstavitev metapodatkov o matičnih podatkih. Potrebno je zagotoviti način, s katerim bo omogočeno vzdrževanje in avtomatsko posodabljanje vrednosti matičnih podatkov. Sistem mora biti sposoben prejetja sprememb in nato obveščanja vseh

podsystemov, ki jih sprememba zadeva. Zgoraj opisani postopek se v literaturi pojavlja kot pristop od zgoraj navzdol, saj izhaja iz modela poslovnega procesa [15].

### 1.2.2 Od spodaj navzgor

Alternativa zgoraj opisanemu je pristop od spodaj navzgor [15]. Bistvo tega pristopa je, da za izhodiščno točko vzamemo že obstoječe aplikacije in v njih identificiramo matične podatke. Namesto da iščemo matične podatke v poslovnih pravilih, jih poskušamo prepoznati direktno v že obstoječih podatkih aplikacij. V prvem koraku iz aplikacij izluščimo metapodatke o matičnih podatkih in jih zberemo na enem mestu. Pravila za razpoznavanje so vsebinsko enaka kot v drugem koraku koncepta od zgoraj navzdol in so opisana v podpoglavju 1.2.3. Naslednji korak predstavlja identifikacijo matičnih podatkov, ki ustrezajo metapodatkom. V zadnjem koraku je potrebno še standardizirati način predstavitve in način posredovanja matičnih podatkov vsem aplikacijam, ki bodo uporabljale storitev MDM.

Odločitev o tem, katerega izmed omenjenih pristopov bomo izbrali, je v veliki meri povezana s tem, ali MDM uvajamo v že obstoječi IS, ali pa dodajamo novo komponento k obstoječemu MDM sistemu. Koncept od zgoraj navzdol je primeren, kadar postavljamo nov IS ali na obstoječ MDM sistem pridružujemo novo aplikacijo. Koncepta od spodaj navzgor se navadno poslužujemo takrat, kadar MDM uvajamo v obstoječo množico že delujočih aplikacij.

### 1.2.3 Razpoznavanje metapodatkov

Kot smo že omenili, je pri pridobivanju matičnih podatkov bistven korak razpoznavanje metapodatkov o matičnih podatkih. Kaj torej določa metapodatke o matičnih podatkih in kako jih medsebojno poenotiti med različnimi aplikacijami, ima velik pomen za dokončno uspešno implementacijo MDM projekta.

Metapodatke o matičnih podatkih določimo v naslednjih treh korakih [2, 19]:

- Pridobivanje metapodatkov o atributih  
Iz izvornih aplikacij poslovnih procesov ali druge obstoječe dokumentacije pridobimo čim večji nabor podatkov o atributih matičnih podatkov. Zanimajo nas imena, podatkovni tipi, predpisane dolžine, razne omejitve nad podatki, porazdelitve in zaloge vrednosti atributov itd. Vse tako pridobljene informacije shranimo v centralizirano zbirko.
- Iskanje in razreševanje podobnosti  
Iz množice zbranih podatkov je potrebno izoblikovati nabor metapodatkov, ki opredeljujejo attribute matičnih podatkov. Pogosto se pojavijo delna prekrivanja atributov, za katera iz metapodatkov ne moremo biti popolnoma prepričani, da gre za

vsebinsko en in isti atribut. V takih primerih morajo poslovni analitiki sprejeti odločitev, ki se glede na njihovo vsebinsko znanje, zdi najbolj pravilna.

- Poenotenje pomena

Ko so metapodatki izluščeni in urejeni, jim je potrebno pripisati tudi pomen. Pogosto se dogaja, da imamo v dveh ali več aplikacijah atribut z enakim imenom in tipom vendar drugačnim pomenom. Tak primer je recimo atribut naslov z dolžino 400 znakov. V eni aplikaciji je njegov pomen naslov stalnega prebivališča, medtem ko druga aplikacija v atributu hrani naslov za dostavo. V tem koraku je torej potrebno metapodatkom, ki smo jih združili, pripisati tudi enak pomen.

Brez jasnih definicij metapodatkov o matičnih podatkov se organizacije, ki želijo uvesti MDM, hitro znajdejo v težavah, saj siva področja navadno hitro nastopijo.

### 1.3 Kaj je upravljanje matičnih podatkov

Znotraj podjetja imamo navadno več aplikacij, ki pokrivajo različna področja, kot so npr. upravljanje odnosov s strankami (ang. *customer relationship management* – CRM), podatkovno skladiščenje (ang. *data warehousing* – DWH), upravljanje človeških virov (ang. *human resource management* – HRM), finance, marketing ipd., in imajo vsaka svojo podatkovno bazo. Veliko matičnih podatkov je zato podvojenih v več bazah in kmalu se pojavijo vprašanja, kateri zapis je veljaven, kje se nahaja zadnja ažurirana informacija itd.

»Upravljanje z matičnimi podatki je ogrodje, sestavljeno iz procesov in tehnologij. Cilj je vzpostaviti in zagotavljati avtorizirano, zanesljivo, trajnostno, natančno in varno podatkovno okolje, ki predstavlja »enotno različico resnice«. V to okolje sprejeti sistem zapisov se uporablja v celotni organizaciji, v vseh množicah aplikacijskih sistemov, poslovnih linijah in uporabniških skupnostih« [3, str.11].

Kot si iz zgornjega opisa lahko predstavljamo, vpeljava MDM-a v organizacijo zahteva veliko organizacijsko, časovno in finančno zavezanost celotne organizacije. Gre namreč za več kot zgolj umestitev nove aplikacije v sistem. V veliko primerih so potrebne korenite spremembe v poslovnih pravilih organizacije, po vzpostavitvi sistema pa je potrebno matične podatke obdržati čiste, konsistentne in v zadnjem stanju.

## 1.4 Upravljanje matičnih podatkov o strankah

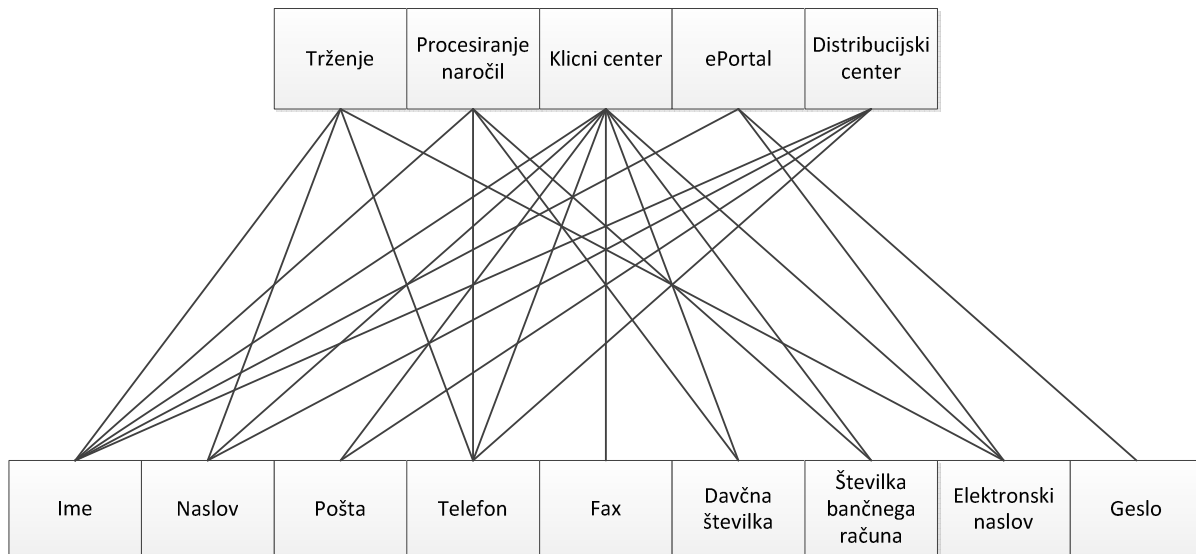
Kadar za namene obdelovanja različnih področij podatkov v povezavi s strankami uporabimo strategijo in arhitekturo MDM, to pojmuje kot integracijo podatkov o strankah [18].

»CDI je obširna množica tehnoloških gradnikov, servisov in poslovnih procesov, ki ustvarjajo, vzdržujejo in zagotavljajo razpoložljivost pogleda na stranko. Pogled je točen, pravočasen, integriran in celovit preko vseh linij poslovanja, kanalov in poslovnih partnerjev« [3, str. 14].

Gre torej za sistem MDM, ki je specializiran zgolj za upravljanje podatkov o strankah. Na tem mestu gre omeniti, da ima termin stranka v našem primeru nekoliko širši pomen. V nabor strank tako spadajo termini kot so: fizična oseba, pravna oseba, klient, kontakt, pacient, naročnik itd.

V povezavi z rešitvami MDM-CDI se v literaturi pogosto uporablja termin vozlišče CDI (ang. *hub*), ki predstavlja uskladitev podatkovnih virov o stranki in jih navzven prikazuje kot poenoten vir [3]. Vozlišče CDI ima nalogo, da očisti in uredi podatke, sistemom, ki se preko servisov nanj povezujejo, pa predstavlja transparentne podatke o strankah.

CDI močno pripomore organizaciji pri doseganju nižjih operativnih stroškov, skladnosti poslovanja s predpisi in doseganju multidisciplinarnih poslovnih zahtev za okrepitev dobičkonosnosti [18]. Slika 2 nazorno prikazuje kako so matični podatki o strankah porazdeljeni po aplikacijah, kjer so aplikacije navedene v zgornjem nizu, matični podatki pa so poenoteni v spodnjem nizu.



**Slika 2: Prekrivanje matičnih podatkov o strankah.**

Iz prepletenih povezav med aplikacijami in atributi na sliki je jasno, da to vodi do konfliktnih informacij, saj je isti podatek navadno ročno zaveden v več različnih sistemih. Razpolagati s točnimi in popolnimi podatki o strankah in njihovih stikih s podjetjem je bistvenega pomena pri zagotavljanju uspešnega poslovanja.

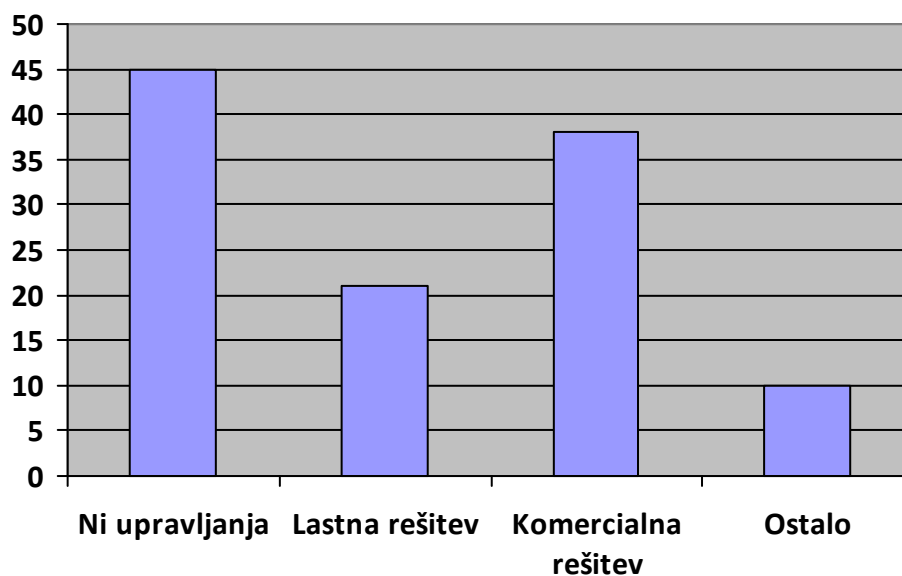
CDI platforma vključuje celoten 360 stopinjski pogled na podatke o stranki, vključujoč vse relacije, ki jih stranka ima do organizacije. Kot primer lahko navedemo podatke o vseh možnih kanalih, preko katerih lahko s stranko komuniciramo [11]. Danes stranka noče biti več dosegljiva samo preko telefona, ampak želi npr. prejemati kupljeno blago na svoj poštni naslov, reklamni material na elektronsko pošto in podajati svoje mnenje o izdelkih preko socialnih omrežij.

Za izpolnjevanje želja uporabnikov mora organizacija razumeti potrebe, cilje, zmožnosti svojih strank. Hkrati pa mora imeti možnost ponujanja novih produktov, novih storitev in iskanje priložnosti za navzkrižno prodajo (ang. *cross-sell*). Vse naštetu lahko organizacija doseže zgolj z natančnimi in celovitimi informacijami o svojih strankah.

## 2 Sistem upravljanja matičnih podatkov

### 2.1 Prednosti upravljanja matičnih podatkov

Globalno MDM pridobiva močan zagon. Po ocenah analitskega podjetja Gartner so svetovni prihodki z naslova programske opreme MDM v letu 2012 znašali 1,9 milijarde ameriških dolarjev, kar pomeni 21 odstotno zvišanje v primerjavi z letom 2011. Po napovedih naj bi trg



**Slika 3: Deleži rešitev za upravljanje matičnih podatkov.**

do leta 2015 dosegel vrednost 3,2 milijarde dolarjev [25]. Po analizi, izvedeni v letu 2012, je bil delež organizacij, ki v svojem poslovanju niso imele formaliziranega upravljanja z matičnimi podatki, kar 45 odstotkov [17]. Kot prikazuje graf na sliki 3, je delež organizacij z lastno rešitvijo upravljanja malenkost nad 20 odstotki, medtem ko je od vseh rešitev najbolj pogosta uporaba enega od komercialnih paketov MDM. Samo 10 odstotkov organizacij se odloči za uporabo odprtokodnih rešitev ali drugih oblik upravljanja kot na primer storitve v oblaku. Zadnja aktualna analiza, ki jo je opravil Gartner, pa ugotavlja, da se 40 odstotkov vseh vprašanih podjetij ravno pripravlja na vzpostavitev sistema MDM [16],

Uvedba MDM je kompleksen, dolgotrajen in seveda tudi drag proces [15]. Ker uspeh projekta ni vnaprej zagotovljen, se organizacije za ta korak odločijo zgolj v primeru, ko imajo jasno predstav o tem, kakšne koristi bodo s tem pridobile. Cilj investicije ne sme biti zgolj uvedba

MDM v poslovanje, ampak zavedanje, da zagotovitev poenotenega pogleda nad ključnimi poslovnimi entitetami omogoča občutne izboljšave v produktivnosti, ravnanju s tveganji in zmanjševanju stroškov.

Poglejmo torej natančneje, kaj so doprinosi uvedbe MDM-CDI v poslovanje organizacije. Če glavne prednosti, ki jih navajajo različni avtorji [3, 15], strnemo v nekaj ključnih točk, lahko povzamemo sledeče doprinose MDM-CDI:

- Izboljšana izkušnja stranke

Vsaka odgovorna organizacija se zaveda, da se bo samo zadovoljna stranka vračala in koristila njene storitve. Poleg tega, da so storitve kakovostne, odzivne, varne in vedno na voljo, je za bogato izkušnjo stranke potrebno zagotoviti personalizirano obravnavo. Slednje ni mogoče brez urejenih podatkov o strankah, ki so osnova za prilagajanje storitve potrebam posamezne stranke.

- Povečevanje prihodkov

Ko ima organizacija urejene podatke o strankah, jim lahko ponudi dodatne storitve ali blago na podlagi poznavanja na primer družinskih razmerij in tako poveča prodajo. Gre torej za širjenje posla na podlagi dodatno ponujenih storitev po različnih grupacijah obstoječih strank. Kot primer lahko navedemo paketne ponudbe zavarovalnic, kjer ob sklenitvi zavarovanja, zavarovalnica družini ponudi popust v primeru, da ostali družinski člani prenesejo svoja zavarovanja k tej zavarovalnici. Z ozirom na to, da so stroški zadrževanja strank nižji kot stroški pridobivanja novih, je pomembno najti način kako zadržati obstoječe stranke. Urejeni podatki o strankah in posledično spremljanje njihovih interakcij z organizacijo bistveno olajšajo iskanje vzroka za odhajanje strank h konkurenci.

- Nižanje stroškov

Ker MDM-CDI predstavlja nabor v celoti informatiziranih poslovnih procesov, to seveda pomeni manj administrativnih stroškov. Potrebe po kadrih, ki pred poenotenjem poslovanja v različnih oddelkih opravljajo podobne naloge, se zmanjšajo, poslovni procesi postanejo bolj poenoteni in transparentni. Ker je sistem centraliziran, se stroški vzdrževanja informacijske tehnologije bistveno skrajšajo. Odpadejo namreč vsakršna podvajanja infrastrukture, kar pomeni manj strežnikov, manj licenc za programsko opremo, manj porabljene energije in prostorov, manj podizvajalcev in veliko cenejša nadaljnja integracija novih sistemov. Seveda so za implementacijo MDM potrebna dodatna vlaganja v infrastrukturo in izobraževanje kadrov, vendar se po navedbah nekaterih svetovalcev [30] investicija povrne že v dobrem letu, z naslova prihrankov ukinjanja starih aplikacij.

- Izboljšano upravljanje s tveganji

Kadar pride do spremembe v zakonodaji, imamo v razvejanih in samostojnih silosnih aplikacijah ogromno dela s prilagoditvami novim predpisom. Navadno pride do veliko ponavljajočih nadgradenj, ki jih opravljajo zunanji izvajalci in tako celoten proces precej stane. V primeru poenotnega sistema MDM se nadgradnja izvede zgolj nad enim sistemom, kar je veliko ceneje, celoten proces pa je veliko bolj obvladljiv. Pri nadgradnjah posameznih aplikacij se pogosto dogaja, da hkratna posodobitev ne uspe oziroma le-ta iz najrazličnejših vzrokov ni mogoča. V takem primeru imamo lahko velike težave pri poslovanju in medsebojnem povezovanju aplikacij in nekonsistentnosti pri spremljanju poslovanja in izdelavi poročil.

- Centralizacija matičnih podatkov

Centralizacija matičnih podatkov se, ob predpostavki, da je poskrbljeno za varnostne kopije podatkov, z vidika možnosti izgube podatkov izkaže za bolj varno alternativo v primerjavi s silosnimi aplikacijami. Veliko bolj pregledno in učinkovito je reševati kritične situacije na enem mestu, kot zagotavljati nemoteno delovanje verige aplikacij in njihovih parcialnih zbirk podatkov. Z drobljenjem zbirke podatkov na več enot, se veča tudi verjetnost izpada zbirke kot celote.

- Nižanje tveganja zaradi slabih in nepopolnih informacij

Kadar so podatki porazdeljeni po več sistemih, je za kakovostno in učinkovito upravljanje z njimi potrebnega kar nekaj kadra, ki za preverjanje usklajenosti in pravilnosti porabi tudi veliko časa. Kljub velikim naporom se podatki v posamezne sisteme vnašajo z zamikom, napake in podvojeni zapisi pa se ne odkrivajo pravočasno. Bolj kot so podatki zanesljivi, manjše je tveganje za napačno sprejete odločitve vodstva organizacij.

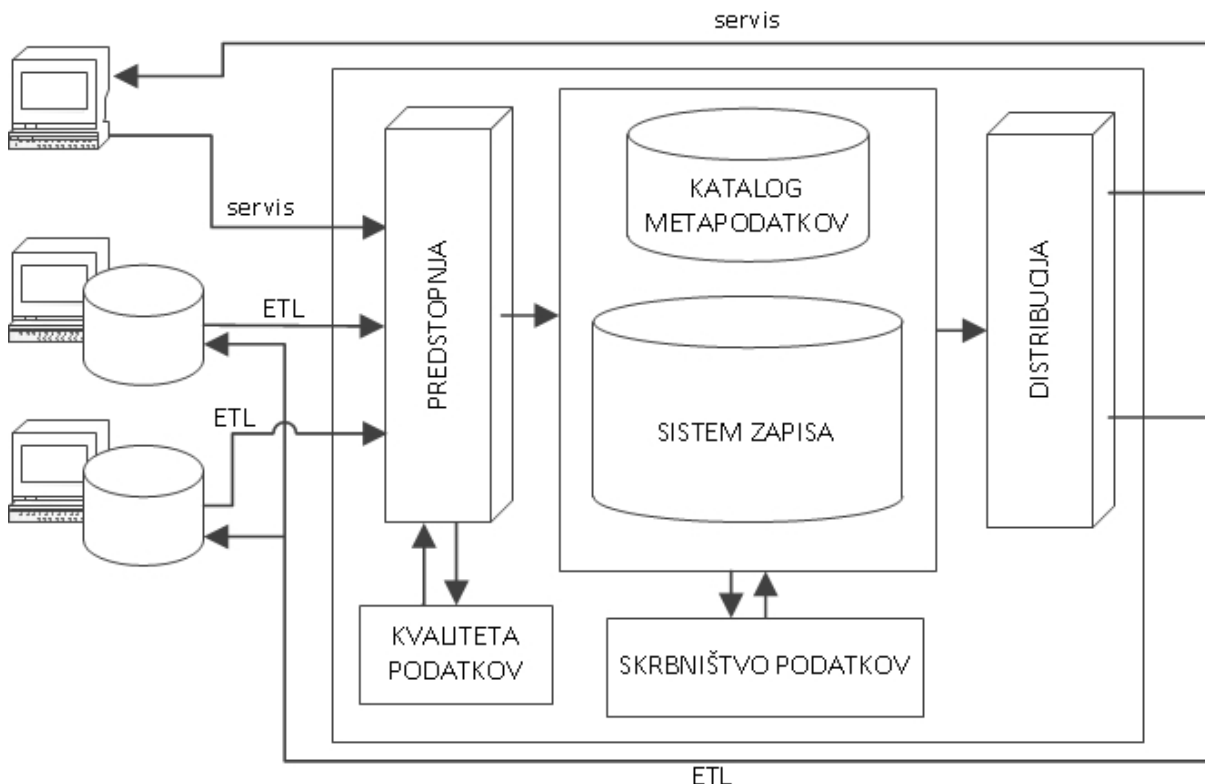
- Izboljšano planiranje poslovanja

Popolni podatki o strankah in njihova jasna segmentacija pomenijo uspešnejše ciljno oglaševanje in ponujanje storitev, ki jih stranka potrebuje. Sledenje omogoča planiranje oglaševanja in namenjanje sredstev v segmente, kjer si organizacija želi krepiti svojo pozicijo. Zopet gre poudariti, da so urejeni podatki temelj za sprejemanje odločitev vodstev. Zaradi še vedno prepogoste silosne ureditve aplikacij v IS se pogosto dogaja, da so poročila, čeprav se nanašajo na iste stranke, medsebojno neprimerljiva. Razlog za to so velikokrat nekonsistentnosti v eni izmed lokalnih kopij registra strank. CDI s poenotenjem zagotavlja boljše primerljivost končnih poročil in tako boljše podlago vodstvu za sprejemanje odločitev.

Imeti celotno sliko o stranki, organizaciji omogoča bogat nabor posamezni stranki prilagojenih servisov in storitev, ter njihovo primerno obravnavanje. Slednje ima za posledico



Pogosto se srečamo s potrebo prenosa večje količine podatkov, kjer nam vodilo SOA ne zadošča več. Taki primeri so recimo prenosi presekov stanj v podatkovno skladišče ali postopki za usklajevanje matičnih podatkov silosnih aplikacij s stanjem v vozlišču MDM. V teh primerih smo primorani uporabiti postopke paketnega prenosa podatkov, ki niso transakcijsko naravnani in se navadno izvajajo v treh stopnjah: pridobivanje (ang. *Extract*), preoblikovanje (ang. *Transform*) ter nalaganje (ang. *Load*), ki jih skupaj označujemo kot postopki ETL.



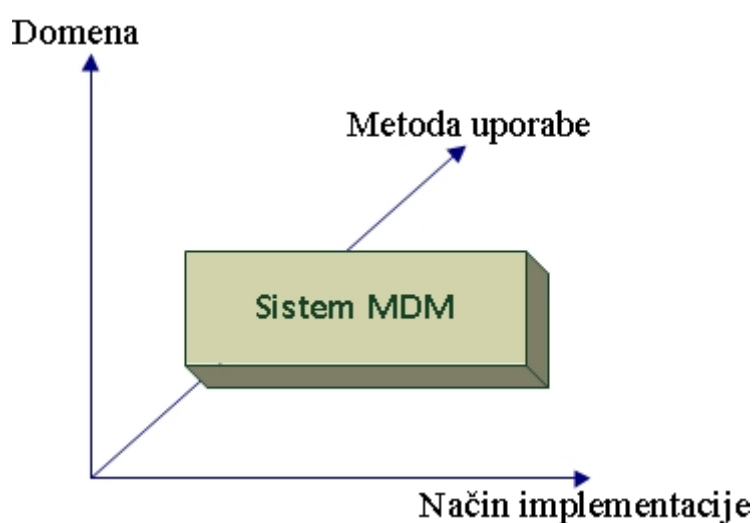
**Slika 5: Arhitektura sistema za upravljanje matičnih podatkov.**

Če si sistem MDM na sliki 5 pogledamo nekoliko podrobneje ugotovimo, da jedro sestoji iz dveh zbirk podatkov, katerima sta pridruženi stopnji za sprejem in distribucijo ter modula za kakovost in skrbništvo nad podatki [21]. Prva zbirka predstavlja sistem zapisa, ki je edina veljavna različica matičnih podatkov na nivoju celotne organizacije, druga pa predstavlja centraliziran opis matičnih podatkov v obliki enotnega repozitorija metapodatkov. Predstopnja predstavlja vmesnik, na katerega aplikacije odlagajo spremembe nad metapodatki. Prav v predstopnji se vrši kontrola kakovosti prejetih podatkov, kjer gre predvsem za procese standardizacije, odprave duplikatov (ang. *de-duplication*), združevanja in obogatitve podatkov. V modulu za skrbništvo podatkov se odkriva konflikte med zapisi matičnih podatkov, naloga skrbnikov pa je, da si označene zapise ogledajo ter se odločijo kateri zapis je pravi. To se navadno zgodi, kadar so avtomatska pravila za določanje

prevladujočega zapisa pomanjkljiva in je zato potrebna pozornost skrbnika, ki napako odpravi.

## 2.3 Razsežnosti upravljanja matičnih podatkov

Prav tako, kot se razlikujejo potrebe, zaradi katerih se organizacija odloči za vpeljavo sistema MDM v svoje poslovanje, se razlikujejo tudi lastnosti posameznih rešitev. Literatura navaja tri razsežnosti, v katerih se posamezna implementacija lahko giblje [6, 15]. Vse tri razsežnosti so nazorno prikazane na sliki 6, kjer se MDM giblje v tro-dimenzijskem prostoru [6, str. 12].



Slika 6: Dimenzije sistema za upravljanje matičnih podatkov.

Dimenzija domene vsebinsko opredeljuje podatke, ki bodo v MDM shranjeni. Pove nam katere matične podatke želimo upravljati. V večini primerov gre za podatke o strankah, produktih ali računih. Naslednja dimenzija je metoda uporabe, ki predstavlja enega od treh načinov, na katerega bo organizacija uporabljala sistem MDM. Gre za sodelovalni, operativni ali analitični način, ki so podrobneje opisani v nadaljevanju. Končno nam dimenzija načina implementacije podaja informacijo o tem, na kakšen način bomo načrtovali rešitev. Vsako izmed navedenih dimenzij bomo podrobneje opisali v naslednjih razdelkih.

### 2.3.1 Domena

Glede na značilnosti matičnih podatkov te lahko členimo v domeno strank, produktov in računov [6]. Kot stranke smatramo vse osebe in organizacije, ki na nek način stopajo v kontakt z organizacijo. Po večini gre tukaj za raznovrstne matične podatke o kupcih, dobaviteljih, zaposlenih itd. Kot že omenjeno, domeno stranke označujemo s kratico CDI in je

izmed vseh domen najbolj popularna. V domeno produktov spadajo vse bistvene informacije o proizvodih in storitvah, ki se že ponujajo na trgu ali so še v fazi razvoja. Označujemo jo s kratico PIM (ang. *Product Information Management*), njena posebnost pa je, da se njen podatkovni model stalno spreminja in dopolnjuje [31]. Gre za dodajanje novih atributov, saj so proizvajalci pod stalnim pritiskom konkurence in želja svojih strank prisiljeni v prilagoditev svojih proizvodov. Klasičen primer so avtomobili, kjer stranka lahko izbira med celo paleto dodatne opreme, barve, tipom, močjo motorja itd. Proizvajalec mora imeti jasno razdelane vse kombinacije omenjenih produktov in njihovih variacij in to mu MDM mora omogočati. Domena računov je mišljena za shranjevanje podatkov o finančnih računih in pogodbah s strankami. V tej domeni najdemo tudi konsolidirano informacijo o skupni vrednosti izdanih in plačanih računih, črne liste in ostale informacije, ki organizaciji omogočajo boljše pogajalsko izhodišče. Ker v tem delu posvečamo večjo pozornost podatkom o strankah, se bomo v nadaljevanju omejili zgolj na domeno strank.

Na tem mestu velja omeniti, da poleg omenjenih osnovnih domen obstajata še dve vrsti razmerji med domenami [6]:

- Razmerja znotraj domene

Tu gre za razmerja med matičnimi podatki znotraj ene domene. Do tega primera pride pri izdelavi poročil, ko znotraj domene obstaja hierarhija matičnih podatkov. Zanima nas recimo iz katerih polizdelkov je sestavljen določen izdelek.

- Razmerja med različnimi domenami

V nasprotju s predhodnim razmerjem gre tukaj za povezovanje matičnih podatkov med dvema ali več rizičnimi domenami. V praksi nas recimo zanimajo vse stranke, ki so kupile določen model avtomobila pri enem izmed naših posrednikov. V tem primeru smo vzpostavili relacijo med domenami stranka, račun in produkt.

Vsak matični podatek lahko tvori razmerja znotraj svoje domene ali pa se navezuje na matične podatke drugih domen.

## 2.3.2 Metode uporabe

Obstajajo tri metode uporabe sistema MDM [6, 15], ki se razlikujejo glede na način uporabe podatkov in načina dostopa do njih.

### 2.3.2.1 Sodelovalna (ang. collaborative)

Predstavljamo si, da isti nabor zapisov v MDM sistemu uporablja in posledično spreminja več različnih poslovnih procesov. Da ti procesi ne bi spreminjali podatkov, kadar do njih še niso upravičeni, v MDM uvedemo poseben status zapisa, ki določa, kateri poslovni procesi lahko spreminjajo podatke v določenem trenutku. Oglejmo si dogajanje na primeru postavljanja

novega paketa mobilne telefonije na trg. V prvi fazi lahko novi produkt kreira in dopolnjuje zgolj skrbnik produkta. Predlaga model telefona, določi zakupljene količine, omogočene opcije in ostale podrobnosti, ki ustrezajo segmentu trga, za katerega se nov paket pripravlja. V drugi fazi skrbnik financ določi finančne okvirje, preveri ceno paketa in pričakovan donos novega produkta. V tretji fazi pride na vrsto oddelek trženja in glede na segment ter razpoložljive finance predlaga način promocije. Omenjene faze se seveda lahko večkrat ponavljajo, dokler se vsi deležniki ne strinjajo in pošljejo tako nastali produkt v prodajo. Kot opisano gre torej za zagotavljanje doseganja konsenza vseh deležnikov, ki sodelujejo pri določanju neke podmnožice matičnih podatkov. To seveda ni mogoče brez detajlno določenega poslovnega procesa, kar je bistvenega pomena pri sodelovalnem načinu metode uporabe. Kot že omenjeno v primeru, se ta metoda uporablja predvsem v navezavi z domeno produkta [6].

### **2.3.2.2 Operativna (ang. *operational*)**

To metodo najdemo v rešitvah, kjer prihaja do sprotne obdelave transakcij (ang. *online transaction processing* – OLTP). MDM je primoran odgovarjati zahtevam iz različnih sistemov in aplikacijskih uporabnikov v realnem času. V nasprotju s sodelovalno metodo ne hrani informacije o statusu, a zagotavlja zadnje in ažurno stanje ne glede na veliko količino sprememb nad posameznim zapisom. V primeru dodajanja nove stranke v MDM-CDI je sistem odgovoren za izvajanje vseh kontrol, kot na primer preprečevanje podvojenosti in za osveževanje vseh morebitnih preostalih silosnih sistemov, ki se nanj navezujejo.

### **2.3.2.3 Analitična (ang. *analytical*)**

Uporablja se predvsem v orodjih za poslovno obveščanje (ang. *business intelligence* – BI), kjer MDM predstavlja zaupanja vreden vir podatkov. Če se za trenutek ozremo nazaj na domene, o katerih smo govorili v prejšnjem poglavju, opazimo da se le-te vsebinsko ujemajo z dimenzijami v zvezdnih diagramih, ki se uporabljajo v podatkovnih skladiščih [12]. To nam jasno nakazuje da je MDM zelo primeren vir za analitično obdelavo podatkov. Hkrati velja opozoriti, da MDM rešitve že vsebujejo nekatere relevantne KPI (ang. *key performance indicators* - KPI) indikatorje, kot na primer število novih strank na teden, tako da se orodja za poročanje lahko priklapljajo direktno na MDM. Taki sistemi že omogočajo uporabo tako enostavnih, kot tudi kompleksnejših analitičnih funkcionalnosti. Kot enostavno funkcijo lahko navedemo iskanje gospodinjstev z uporabo pravil okoli informacij o imenu, naslovu in starosti. Kompleksnejše funkcionalnosti pa omogočajo določevanje vplivnosti posameznega komitenta na podlagi socialne mreže, ki jo pridobimo z analizo poslovnih in zasebnih telefonskih števil, elektronskih naslovov in drugih podatkov o osebi.

### 2.3.3 Načini implementacije

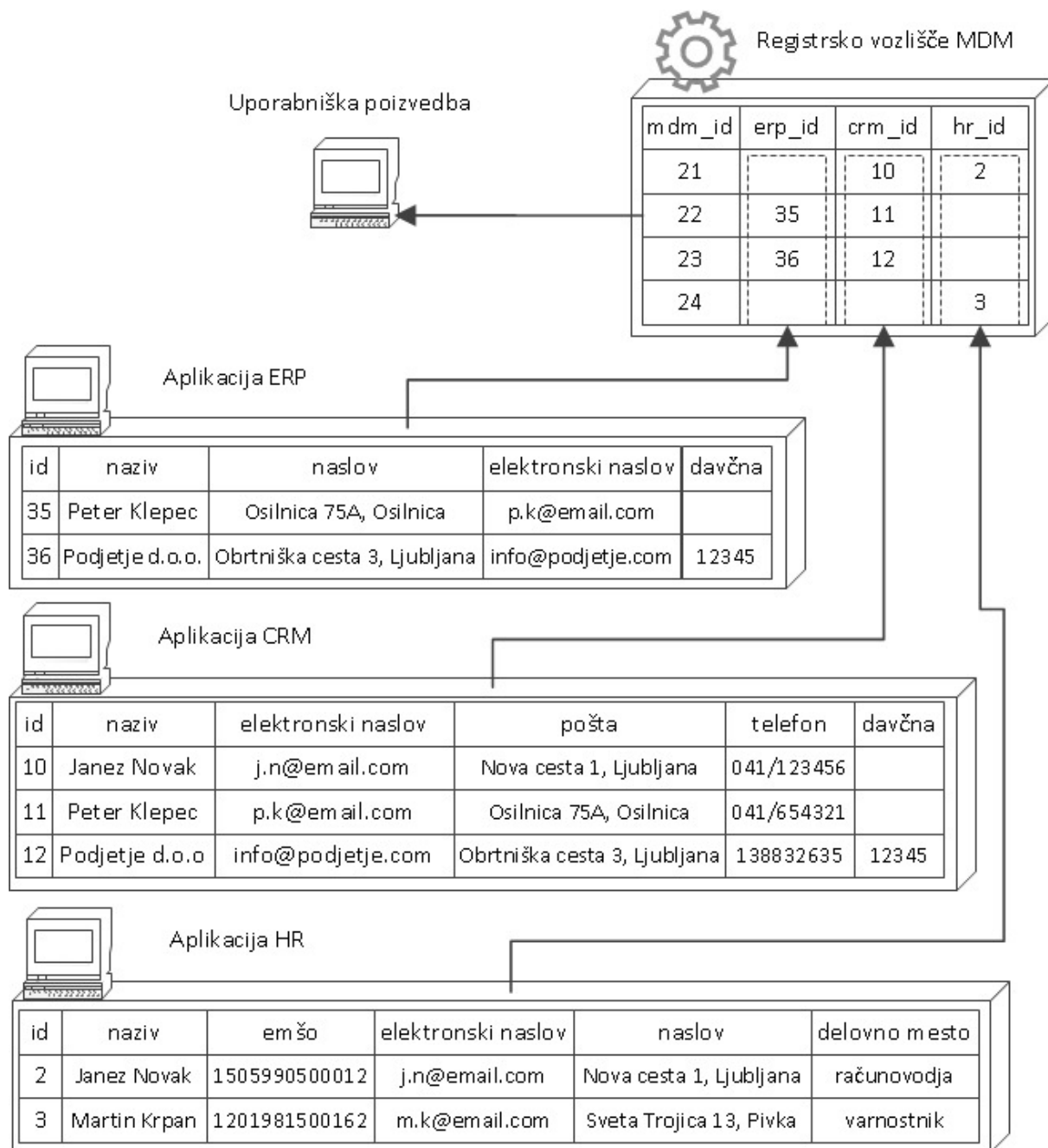
Avtorji načine implementacije poimenujejo različno. Medtem ko eni govorijo o arhitekturi sistema MDM [3], drugi navajajo načine implementacije [6]. V tem delu bomo privzeli slednjo terminologijo in govorili o načinih implementacije, saj smo kot arhitekturo sistema predstavili osnovno modularno zgradbo MDM.

V implementacijskem smislu avtorji navajajo štiri v nadaljevanju opisane pristope [3, 6, 15]. Kot primer predpostavimo, da IS organizacije sestavljajo silosne aplikacije CRM, ERP in HRM, med katerimi ima vsaka svoj register strank ali uporabnikov. Z željo, da bi navedene registre združili in jih navzven predstaviti kot eno urejeno zbirko si pogledjmo, kako bi na tem primeru izgledal vsak izmed sledečih načinov implementacije.

#### 2.3.3.1 Registrski način

Registrsko vozlišče (ang. *registry hub*) je najenostavnejša implementacija MDM. Glavna poanta te ideje je, da čez obstoječe silosne aplikacije poveznemo enostavno strukturo v obliki kazala. To kazalo nato predstavlja vmesnik, ki podrejene silosne aplikacije poenoti in jih navzven predstavlja kot en urejen vir podatkov. Kot si lahko predstavljamo, takšno kazalo ni nič drugega kot nabor metapodatkov, ki povedo od kje posamezni atributi matičnih podatkov izhajajo in pravil, s pomočjo katerih določimo katera različica zapisa in na katerem viru najbolj ustreza uporabnikovim željam. Takšno kazalo lahko sestavimo na dva načina.

Prva in enostavnejša od obeh možnosti je, da vsaka vrstica v registru hrani zgolj ključne identifikatorje in povezave do izbranega dela iskane vrednosti matičnega podatka v vseh virih [15]. To pomeni, da en zapis v kazalu povezuje vse pripadajoče zapise v vseh virih. Prav zaradi shranjevanja zgolj ključnih identifikatorjev, se ta različica registrskega načina v literaturi pojavlja tudi pod imenom identitetno vozlišče (ang. *identity hub*) [3]. Kot prikazuje slika 7 to dosežemo tako, da kazalo sestavimo iz enoličnega identifikatorja vrstice registrskega vozlišča in nato nanizamo vse notranje enolične identifikatorje posameznih virov. V našem primeru imamo tri silosne aplikacije (ERP, HRM in CRM) zato bo, ob pridruženem enoličnim identifikatorju MDM, preslikovalna tabela sestavljena iz štirih stolpcev. Ker ena vrstica vozlišča predstavlja povezavo do vseh pripadajočih zapisov, je potemtakem število vrstic enako številu različnih strank po vseh virih.

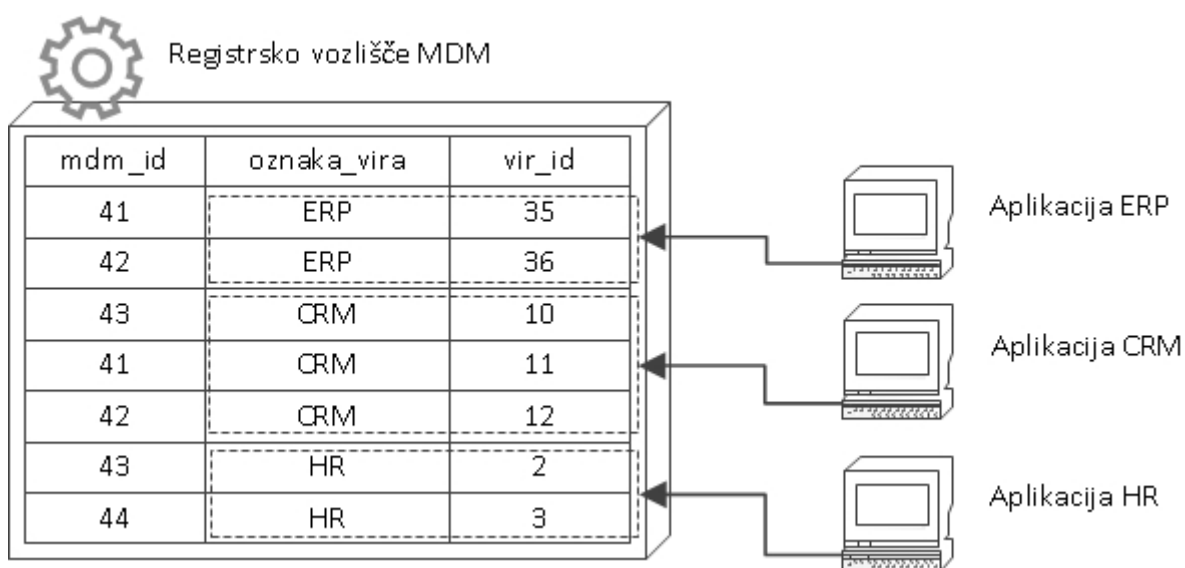


**Slika 7: Registrski način implementacije.**

Dodajanje novega vira v MDM pomeni dodajanje novega stolpca v preslikovalno tabelo, kar za sabo prinese dopolnitev manjkajočih vrednosti. Obstoječe zapise v tabeli je potrebno opremiti s pravilnimi identifikatorji pripadajočih zapisov novega vira in korigirati vse poizvedbe, s katerimi želimo dostopati do dodanega vira. To samo po sebi ni tako problematično, bolj neugodno je dejstvo, da opisana rešitev predpostavlja en zapis za eno stranko. Če si pogledamo primer za MDM\_ID=21, lahko vidimo, da imamo natanko po enega predstavnika v aplikacijah CRM in HRM, medtem ko v ERP zapis ne nastopa. Da bi to

zagotovili, je potrebno v vseh virih sistema MDM in še pred prvo implementacijo, torej v vseh silosnih aplikacijah, izvesti čiščenje podatkov. Vsako tako opravilo v zalednih sistemih se velikokrat izkaže za zelo drago in zamudno opravilo, ki hkrati zelo podaljša čas integracije [8].

Tako pridemo do druge možnosti pri kateri vozlišče sestavimo tako, da v njega prenesemo identifikatorje vseh zapisov iz vseh virov ter oznako vira, iz katerega identifikator prihaja [3]. Imamo torej oznako vira in identifikator zapisa, ki skupaj tvorita enolični identifikator zapisa v vozlišču, katerima dodamo identifikator MDM, ki medsebojno združi istovrstne zapise.



**Slika 8: Registrski način implementacije z oznako vira.**

Slika 8 prikazuje enostavno registrsko vozlišče MDM s pripadajočimi podrejenimi viri, iz katerih se v vozlišče prenesejo tri stranke CRM, dve stranki ERP in dve stranki iz aplikacije HRM. Kot je prikazano na sliki s črtkanim okvirjem, se za omenjene zapise preneseta enolični identifikator zapisa in oznaka vira. Na podlagi predhodno določenih pravil se zapisi opremijo z vrednostmi MDM\_ID, ki »podvojene« zapise iz istih ali različnih virov združi v poenoten zapis. V našem primeru MDM\_ID=41 združuje zapis ERP\_ID=35 iz aplikacije ERP in zapis CRM\_ID=11 iz aplikacije CRM. Omenjena zapisa MDM smatra kot eno in isto stranko, zato se na izhodu, ki ga lahko predstavlja sistem za poročanje, pojavi zgolj en zapis. Sistemi MDM poleg pravil za poenotenje zapisov o strankah navadno omogočajo tudi določanje pravil za izbor vrednosti atributa, v primeru da je le-ta zastopan na več virih hkrati. Kot primer takega pravila pri izboru atributa naslova bi lahko navedli dva kriterija. Prvi kriterij naj poišče naslov, ki je najbolj popoln, kar pomeni, da ima tako ulico kot hišno številko. Drugi kriterij pa bi, v primeru več enakovrednih kandidatov prvega kriterija, narekoval izbiro naslova iz

aplikacije HRM pred aplikacijo CRM, aplikacijo ERP pa bi postavil na zadnje mesto. Na ta način proces čiščenja podatkov prestavimo v vozlišče MDM, kar odpravi potrebo po poseganju v silosne aplikacije, sam proces čiščenja pa porazdelimo na daljše časovno obdobje in tako ta ne predstavlja več časovno kritične operacije uvedbe MDM.

Ob pojavu novega zapisa v enem izmed virov se ta prenese v vozlišče. Tu se nato preveri ali zapis že ustreza pravilom katere od obstoječih strank. Če je temu tako, se ga uvrsti v množico zapisov, ki predstavljajo to stranko, novo pridobljene atributi pa se ustrezno ovrednoti. V nasprotnem primeru se ustvari novo stranko, z novo vrednostjo MDM\_ID, katere edini predstavnik je ta zapis. Ob kreiranju zapisa z novo vrednostjo MDM\_ID obstaja velika verjetnost, da se bo v kratkem iz katerega izmed preostalih virov pojavil podoben zapis, ki bo ustrezal tej isti stranki.

Oglejmo si nekaj glavnih prednosti in slabosti, ki jih prinaša registrski način implementacije [6, 15].

Prednosti:

- Ažurnost podatkov: V primeru registrskega načina implementacije poteka združevanje podatkov iz različnih virov v realnem času. To pomeni, da ima organizacija v MDM vedno na voljo zadnje veljavno stanje.
- Hitra in poceni rešitev z nizkim tveganjem za organizacijo: Kot smo spoznali, se podatki prenašajo zgolj v eno smer, in sicer od virov skozi vozlišče MDM proti sistemu za poročanje. To pomeni, da med implementacijo ni nikakršnega poseganja v obstoječe aplikacije in je zato implementacija hitra in dokaj enostavna. Vse aplikacije med uvajanjem MDM delujejo nemoteno in zato sam proces uvedbe nima nikakršnih neposrednih vplivov na tekoče poslovanje organizacije.
- Ni sprememb poslovnih procesov: Ker se poslovanje ne spreminja, lahko ostajajo vsi poslovni procesi nespremenjeni. Kar pridobimo z uvedbo MDM je, da so vhodni podatki v poslovnem procesu popolnejši in odražajo stanje, ki je bližje realnemu.

Slabosti:

- Slabe performance poizvedb: Dejstvo, da se vrednosti atributov pridobivajo neposredno iz aplikacij in se ne shranjujejo v vozlišču MDM, prinaša s seboj nekatere zmogljivostne težave. Ker je ob vsaki poizvedbi potrebno najprej pridobiti podatke iz različnih virov in jih nato z različnimi algoritmi poenotiti v en zapis, postanejo poizvedbe hitro prezahtevne za izvedbo v realnem času. To predstavlja veliko težavo, saj gre razvoj poslovnega obveščanja ravno v smeri zajema podatkov v realnem času [20].

- Namenjen samo branju: Registrski način implementacije ne omogoča nikakršnega spreminjanja ali vnašanja podatkov preko sistema MDM. Vse spremembe je tako potrebno vnašati direktno v podrejene aplikacije.
- Zgolj operativna raba: Ker sodelovalna raba in analitična raba nista mogoči, smo omejeni zgolj na operativno rabo matičnih podatkov. Sodelovalno rabo je nemogoče izvesti, saj nimamo možnosti popravljanja podatkov preko MDM. Zaradi omejitve na enostavne poizvedbe, je analitična raba praktično neuporabna za resno delo.
- Podatki ostajajo neurejeni: Ker register predstavlja zgolj pogled nad dejanskimi podatki v silosnih aplikacijah, ostajajo ti neurejeni.
- Delovanje pogojeno z delovanjem podrejenih sistemov: Ker je MDM zgolj pogled in kot tak ne hrani matičnih podatkov, je za delovanje preslikave potrebno imeti razpoložljive vse podrejene sisteme. Kakršna koli nedostopnost enega izmed virov pomeni nedelovanje celotnega sistema MDM.

Kot lahko vidimo sta glavni slabosti opisanega načina prav slaba zmogljivost pri kompleksnejših nalogah in pogojevanje delovanja z delovanjem vseh podrejenih aplikacij. Z željo po odpravi teh osnovnih napak se je pojavil usklajevalni način implementacije.

### **2.3.3.2 Usklajevalni način implementacije**

Usklajevalni (ang. *consolidation*) način implementacije je zelo podoben registrskemu, s to razliko, da v vozlišču ne hranimo zgolj identifikatorjev, ampak kar dejanske vrednosti. V MDM ne preslikamo samo identifikatorje ampak vrednosti atributov iz vseh virov in jih s podobnimi pravili, kot pri registrskemu načinu implementacije, uredimo v konsolidiran nabor podatkov. Kot rezultat dobimo nabor urejenih podatkov, ki je seveda redundanten, a prinaša kar nekaj pozitivnih lastnosti.

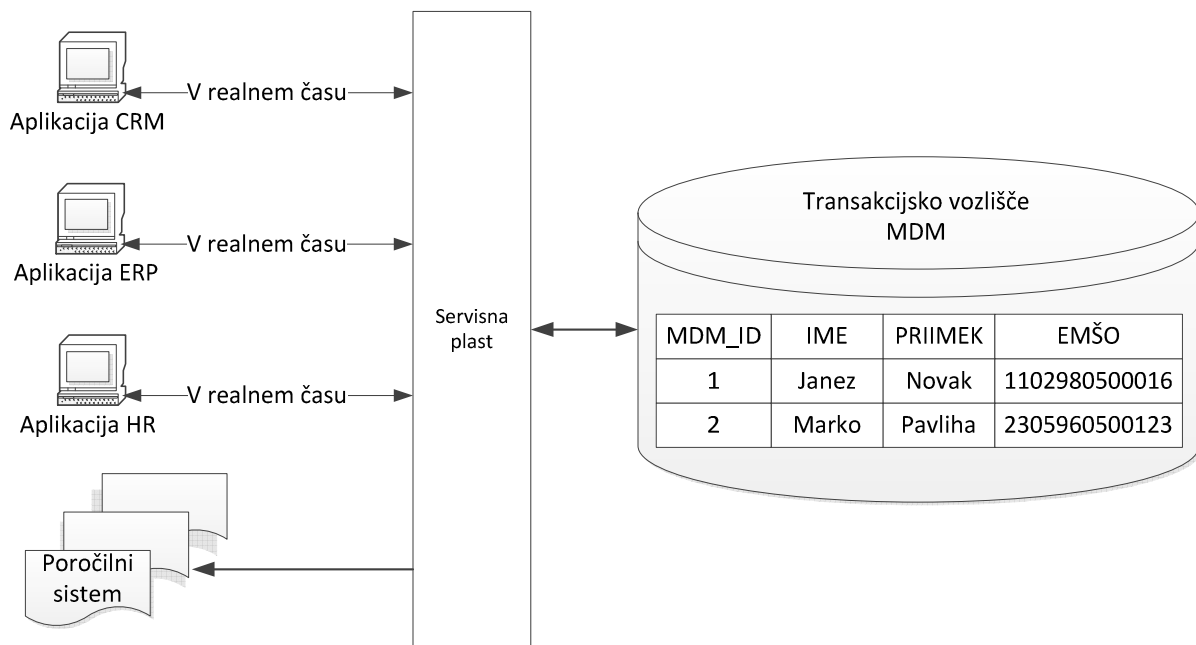
Ker so vsi atributi sedaj na voljo v MDM, nam to omogoča izvajanje kompleksnejših poizvedb brez nevarnosti preobremenjevanja silosnih aplikacij. MDM sedaj za svoje delovanje ne potrebuje več stalnega dostopa do virov, kar zelo izboljša razpoložljivost, saj izpad enega od virov ne pomeni več izpada celotnega vozlišča MDM. Povezavo potrebujemo zgolj tedaj, ko se stanje v MDM usklajuje s stanjem v aplikacijah iz katerih črpamo podatke.

MDM periodično osvežuje svoje podatke, po navadi v nočnem času, kadar aplikacije niso obremenjene. To pomeni, da v vozlišču nimamo zadnjega stanja, ampak z vsakim trenutkom starejše podatke vse do naslednje osvežitve. Za analitične potrebe je to v večini primerov sprejemljivo, ni pa to primerno za spremljanje tekočega poslovanja. Prav tako kot pri

registrskem načinu tudi pri usklajevalnem spreminjanje podatkov skozi MDM ni mogoče, saj je podatkovni tok od silosnih aplikacij proti MDM še vedno enosmeren.

### 2.3.3.3 Transakcijski način implementacije

V nasprotju z registrskim pristopom imamo v tem primeru, ki ga prikazuje slika 9 celoten nabor matičnih podatkov shranjen v shrambi MDM. Tu se nahaja edina in veljavna različica matičnih podatkov z vsemi pripadajočimi atributi in metapodatki. Kar vodi v dejstvo, da morajo vse ostale aplikacije uporabljati podatke iz vozlišča MDM, njihove morebitne lokalne kopije pa je potrebno ukiniti. Spreminjanje in branje matičnih podatkov je mogoče samo v MDM, prestali sistemi pa za pridobitev oziroma spreminjanje podatkov uporabljajo mehanizme, ki se za ta namen pripravljene na strani MDM sistema. Vse težave zaradi neuskklajenih različic podatkov odpadejo v trenutku, ko matične podatke poenotimo in centraliziramo.



**Slika 9: Transakcijski način implementacije.**

Način, s katerim omogočimo uporabo podatkov vozlišča na nivoju celotnega IS, je vzpostavitev servisne plasti s pripadajočimi servisi, s pomočjo katerih bodo aplikacije dostopale do podatkov in nad njimi izvajale operacije [6]. Servisi morajo omogočati branje, kreiranje, odstranjevanje in spreminjanje zapisov matičnih podatkov in predstavljajo edino pot, po kateri je možen dostop do matičnih podatkov. Zaželeno je torej, da je MDM-CDI rešitev načrtovana na platformi storitveno usmerjene arhitekture, kjer je SOA »programska arhitektura pri kateri programske komponente izpostavimo kot omrežne servise in so zato

lahko ponovno uporabljene s strani drugih aplikacij« [3, str. 93]. Če povzamemo definicijo konzorcija W3C [33], je arhitektura SOA okarakterizirana s sledečimi lastnostmi:

- Servis je abstraktni pogled na dejanski program, podatkovno bazo ali poslovni proces, ki tipično izvede operacijo na poslovnem nivoju.
- Servis je formalno definiran v smislu izmenjave sporočil med agentom, ki je podal zahtevek ter agentom, ki je odgovor posredoval in ne z lastnostmi agentov samih.
- Servis je opisan z metapodatki, ki jih je mogoče obdelati strojno. Gre za zgolj tiste podrobnosti, ki so vidne navzven in pomembne za uporabo servisa.
- Servisi se nagibajo k uporabi majhnega števila operacij z relativno velikimi in kompleksnimi sporočili.
- Zaželeno je, da se servisi uporabljajo preko mreže.
- Sporočila se navadno pošiljajo z uporabo formata XML, ki je neodvisen od okolja in v standardizirani obliki.

Iz opisanega lahko izluščimo kar nekaj pozitivnih lastnosti, ki jih arhitektura SOA prinese v MDM. Servisna plast doprinese poenoten način dela z matičnimi podatki, ki skupaj z možnostjo ponovne uporabe močno zmanjša število napak in čas potreben za razvoj novih aplikacij. Servisi omogočajo izgradnjo nadgradljivih aplikacij (ang. *scalable applications*), kar se izkaže kot zelo pomembno pri širitvi poslovanja organizacije.

Lahko sklenemo, da MDM in SOA živita v nekakšni simbiozi, saj se servisno orientirana poslovna rešitev zanaša na obstoj shrambe matičnih podatkov, uspešnost implementacije MDM rešitve pa se zanaša na zmožnost upravljanja matičnih podatkov z uporabo spletnih servisov [15]. Dobre MDM-CDI rešitve naj bi bile načrtovane kot k metapodatkom usmerjeno SOA okolje, ki bi omogočalo organizaciji hiter prehod od osredotočenosti poslovanja na račune k osredotočenosti poslovanja na stranke [3].

Predstavljamo si lahko, da izvedba takega tipa vozlišča pomeni velike posege v obstoječe silosne aplikacije in predstavlja precejšnje tveganje za nemoteno poslovanje organizacije med samo implementacijo. Ne gre zanemariti dejstva, da gre v tem primeru za veliko spremembo toka podatkov, ki ima lahko velike posledice v izvajanju poslovnih pravil in, v smislu količine prenašanja podatkov, tudi večje obremenitve na informacijsko strukturo organizacije. Vsi matični podatki, kateri so se do sedaj hranili v silosnih aplikacijah, se bojo sedaj prenašali po računalniškem omrežju, število poizvedb nad vozliščem MDM pa bo na nivoju vsote poizvedb na vseh zalednih aplikacijah. Treba je torej poskrbeti, da bodo aplikacije poizvedovale zgolj po podatkih, ki jih res potrebujejo, izvedba posameznega servisa pa mora biti hitra in na splošno zmogljivostno nezahtevna.

Podobno kot pri registrskem načinu, lahko tudi na tem mestu prednosti in slabosti strnemo okoli nekaj glavnih točk [6, 15].

#### Prednosti:

- Celostna rešitev: Transakcijski način implementacije predstavlja celostno rešitev upravljanja z matičnimi podatki, saj so podatki centralizirani, enotno upravljani in predstavljajo edino veljavno različico. MDM se tako nahaja v središču IS organizacije. Je torej ključna točka, ki skrbi za konsistentnost in poenotenje izbranih ključnih matičnih podatkov.
- Poenostavljena izdelava novih aplikacij: Ker MDM temelji na arhitekturi SOA, je vključevanje novih aplikacij v IS precej poenostavljeno, kajti celotna logika okoli upravljanja matičnih podatkov se skrči na enostavne klice predhodno že pripravljenih servisov [6].
- Poenotenje metapodatkov o matičnih podatkih: Uvedba transakcijskega vozlišča zahteva poenotenje metapodatkov o matičnih podatkih na nivoju celotne organizacije. Ker so podatki shranjeni na enem mestu, to onemogoča pojavljanje različnih interpretacij po posameznih sistemih, ki matične podatke uporabljajo, kar pripomore k skladnosti poslovanja in bolj kakovostnemu poročanju.
- Avtonomnost: MDM sedaj deluje neodvisno v primeru, ko kateri od podrejenih sistemov odpove. Vse transformacije se dogajajo neposredno nad podatki v vozlišču, katero je samozadostno.
- Poenostavljeno spreminjanje poslovnih procesov: Zaradi usklajenosti matičnih podatkov na nivoju organizacije, zaupamo v njihovo kakovost in imamo servise za enostavno uporabo. Spreminjanje poslovnih procesov nam tako ne predstavlja večjih težav.
- Omogoča vse tipe uporabe: Tako analitična kot tudi operativna in sodelovalna metoda uporabe sedaj niso več pod vprašajem.

#### Slabosti:

- Kompleksnost vzpostavitve: Proces ni več enostaven kot pri registrskemu načinu. Potrebno je zagotoviti pravila, po katerih se bodo izvajali procesi standardizacije, normalizacije in čiščenja podatkov ter zagotoviti inicialni prenos podatkov iz zalednih sistemov [3].
- Poseg v vse aplikacije: Upoštevajoč dejstvo, da so sedaj matični podatki shranjeni samo v vozlišču MDM, je potrebna predelava vseh obstoječih aplikacij, ki namesto svojih lokalnih podatkov sedaj potrebne podatke pridobivajo iz vozlišča MDM.

- Uvedba predstavlja določeno tveganje za poslovanje: Med spreminjanjem obstoječih aplikacij lahko prihaja do izpada delovanja, kajti navadno gre za spreminjanje kompleksnih in za poslovanje pomembnih informacijskih rešitev.

#### **2.3.3.4 Kombinirani način implementacije**

Pri kombiniranem načinu implementacije, v nekateri literaturi znanem tudi pod imenom koeksistenčni (ang. *coexistence*) [3], gre za kombinacijo zgoraj omenjenih paradigem registrskega in transakcijskega načina [15]. Njegova uporaba se nakazuje predvsem v primerih, ko želimo, zaradi zahtevne ali onemogočene predelave, nekatere izmed podedovanih sistemov ohraniti, nekaj pa posodobiti oziroma jih nadomestiti z novimi. Na tem mestu omenimo, da se zaradi izgubljene dokumentacije in izvorne kode pogosto znajdemo v situaciji, ko predelava nekaterih podedovanih sistemov enostavno ni mogoča ali pa predstavlja drago in časovno potratno nalogo [8].

Takšna delna integracija je mogoča, ker je vozlišče sposobno hraniti in vzdrževati nekaj matičnih podatkov in nekaj referenc na obstoječe zaledne sisteme. Transakcije, ter s tem tudi novi zapisi in spremembe nad njimi, se v nekaterih aplikacijah še vedno dogajajo v zalednih sistemih in ne direktno nad vozliščem, MDM pa poskrbi, da se spremembe na enem viru prenesejo in uskladijo tako v vozlišču MDM kot tudi v lokalnih kopijah preostalih aplikacij. Ob spremembi v eni izmed aplikacij se transakcija, bodisi v realnem času, ali pa kot del paketne obdelave, prenese v vozlišče MDM. Tu se določi novo množico veljavnih matičnih podatkov, ki se v nadaljevanju uporablja za poročanje in je hkrati osnova za posodobitev podrejenih aplikacije. Vozlišče MDM je v tem primeru tako mesto, kjer je mogoče spreminjanje kopije matičnih podatkov, kot tudi referenčna točka, ki z možnosti hrambe nekaterih skupin matičnih podatkov zagotavlja boljše zmogljivosti poizvedb, medtem ko zaledni sistemi še vedno igrajo vlogo sistema zapisa.

Prednosti:

- Nižja cena: Namesto popolne predelave silosnih aplikacij potrebujemo zgolj vzpostavitev mehanizma paketnega osveževanja podatkov, zato je cena nižja kot pri transakcijskem načinu.
- Manj tvegana vpeljava: Med samo vpeljavo aplikacije ostajajo silosne aplikacije skoraj nespremenjene, kar pomeni, da je tveganje izpada poslovanja bistveno nižje kot pri vpeljavi transakcijskega MDM.
- Zmogljivost poročanja: Hramba najpomembnejših matičnih podatkov v vozlišču omogoča bistveno hitrejše izvajanje poizvedb kot pri registrskem načinu implementacije.

- Mogoče so vse tri metode uporabe: Glede na to, da vozlišče MDM hrani kopijo matičnih podatkov, ki jo je mogoče spreminjati in dopolnjevati, so mogoče vse tri metode uporabe.
- Delovanje ni pogojeno z delovanjem podrejenih sistemov: Ker hranimo kopijo podatkov v samem vozlišču, občasno nedelovanje podrejenih aplikacij ne vpliva na delovanje vozlišča.

#### Slabosti:

- Kompleksna pravila: Pravila s pomočjo katerih ohranjamo pravilno stanje vozlišča so lahko zelo kompleksna in nepregledna ter so zelo podobna pravilom pri registrskem načinu implementacije.
- Neuskklajenost podatkov: V nekaterih primerih usklajevanje podatkov ne poteka transakcijsko, zato se lahko zgodi, da vrednosti v vozlišču niso usklajene s tistimi na virih. To lahko privede do nekonsistentnega poročanja ali neuskklajenega poslovanja, kajti atributi v različnih aplikacijah nimajo enakih vrednosti.
- Redundanca v podatkih: Podatki iz silosnih aplikacij so v informacijskem sistemi praktično podvojeni. Medtem ko so najbolj ažurni matični podatki razpršeni po aplikacijah, se njihova kopija hrani v vozlišču. Glede na dejstvo, da matični podatki v sodobnem poslovanju obsegajo do 15 odstotkov vseh ne-transakcijskih podatkov organizacije, potrebne kapacitete za hranjenje kopij podatkov niso zanemarljive [28].

### 2.3.4 Izbira načina implementacije

Če predhodno omenjene prednosti in slabosti posameznih implementacijskih načinov na enem mestu zberemo okoli nekaj skupnih točk, pridemo do rezultata ki se nahaja v tabeli 2. V stolpcih so navedeni načini implementacije, vrstice pa predstavljajo primerjalne kriterije. Kriteriji so nam že poznani iz opisov posameznih načinov, obrazložiti pa velja pomena sistema zapisa in sistema reference. Tukaj gre za to predstavitev matičnih podatkov v vozlišču MDM. Pri sistemu reference se v vozlišču MDM hrani zgolj referenca na matični podatek, ki je shranjen v eni izmed silosnih aplikacij. Pri sistemu zapisa je v MDM shranjen dejanski podatek in ne zgolj referenca.

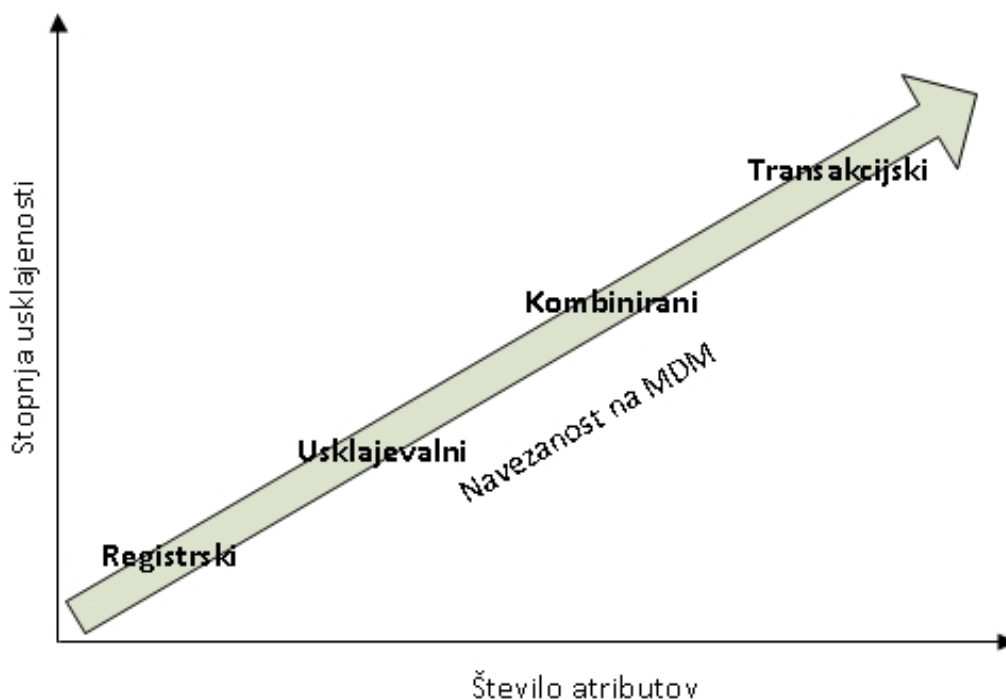
**Tabela 2: Primerjava med načini implementacije.**

	Registrski	Usklajevalni	Kombinirani	Transakcijski
Možne metode uporabe	Operativna	Analitična	Operativna, analitična in sodelovalna	Operativna, analitična in sodelovalna
Sistem zapisa ali reference	Sistem reference	Sistem zapisa	Delni sistem zapisa	Sistem zapisa
Ažurnost matičnih podatkov	Ažurno	Ni ažurno	Delno ažurno	Ažurno
Pogojenost delovanja podrejenih sistemov	Pogojeno	Ni pogojeno	Ni pogojeno	Ni pogojeno
Urejenost metapodatkov organizacije	Ni urejenosti	Ni urejenosti	Delna urejenost	Popolna urejenost
Poseganje v obstoječe silosne aplikacije	Ni poseganja	Ni poseganja	Nizko poseganje	Globoko poseganje
Kompleksnost implementacije	Nizka	Nizka	Srednja	Visoka
Način dostopa do MDM podatkov	Samo branje	Samo branje	Branje in pisanje	Branje in pisanje
Lokacija matičnih podatkov	V izvornih aplikacijah	V MDM in izvornih aplikacijah	V MDM in izvornih aplikacijah	Samo v MDM
Podvojenost podatkov	Ni podvojenosti	Je podvojenost	Delna podvojenost	Ni podvojenosti

Pri natančnem pogledu lastnosti v tabeli opazimo, da se ob prehodu iz registrskega, prek kombiniranega v transakcijski način implementacije, lastnosti vedno bolj nagibajo proti višji stopnji usklajenosti matičnih podatkov (ang. *degree of consolidation*), tesnejšemu navezovanju aplikacij na sistem MDM (ang. *tightly coupled application*) in večanju števila atributov matičnih podatkov [15].

Večanje stopnje usklajenosti matičnih podatkov lahko razumemo tudi skozi prizmo usklajenosti metapodatkov o matičnih podatkih. Kot primer si oglejmo hipotetični atribut *naslov1*. Pri registrskem slogu imajo lahko posamezne aplikacije naslov zapisan v različnih formatih, posamezen naslov pa ima lahko več pomenov. Lahko govorimo o začasnih, stalnih, dostavnih in še kakšnih naslovih, medtem ko je pri transakcijskem slogu pomen atributa *naslov1* točno določen na nivoju celotne organizacije. Prav tako je dotični naslov zapisan v točno določenem formatu in tako dilema, ali je na prvem mestu ulica ali kraj, ne obstaja več.

Bolj kot se pomikamo proti transakcijskemu vozlišču, tesneje so aplikacije navezane na MDM. Pri registrskem načinu smo omenili, da je za delovanje registra nujno potrebno delovanje vseh aplikacij iz katerih MDM črpa podatke. Pri pogledu transakcijskega sloga pa ugotovimo, da je situacija ravno obratna. Ne samo, da delovanje transakcijskega vozlišča ni več odvisno od podrejenih aplikacij, ampak je situacija ravno nasprotna. Aplikacije so tako



**Slika 10: Spekter načinov implementacije.**

tesno vezane na vozlišče MDM, da brez njega praktično ne morejo delovati. Na tem mestu se seveda vprašamo, kaj smo potem sploh pridobili, če so sedaj aplikacije odvisne od sistema MDM? S stališča IS organizacije veliko, saj za konsistentno stanje ni več potrebno delovanje vseh aplikacij, temveč zgolj enega sistema MDM.

Večje kot je število atributov matičnih podatkov, slabšo zmogljivost dosegamo z uporabo registrskega vozlišča. Kot že omenjeno, je to ena glavnih težav tega načina implementacije; zato ni primeren za kompleksnejše rešitve. Pravo nasprotje je transakcijsko vozlišče, ki hrani vse podatke na enem mestu; kompleksnejše poizvedbe ne predstavljajo več ovire pri poslovanju organizacije.

Vsi načini implementacije torej ležijo vzdolž spektra, sestavljenega iz treh dimenzij: števila atributov, stopnje usklajenosti in stopnje navezanosti aplikacij na MDM, kar prikazuje slika 10 [15]. Iz tega lahko sklepamo, da si rešitve sledijo v smiselnem zaporedju in si organizacija z izbiro določenega načina implementacije ne zapre vrat za postopen prehod k alternativni rešitvi. Nekateri ponudniki MDM rešitev navajajo, da je v primeru prehoda iz registrskega na transakcijski način mogoče, z naslova ponovne uporabe že implementirane rešitve, število potrebnih korakov zmanjšati za 50 odstotkov in tako veliko privarčevati na času, potrebnem za vpeljavo transakcijskega vozlišča [26]. Organizacija si torej lahko zastavi iterativno vpeljavo MDM-a v svoje poslovanje. V prvi stopnji se brez poseganja v obstoječe aplikacije postavi registrsko vozlišče, kasneje pa se z »mehkim« prehodom preko kombiniranega preide do transakcijskega MDM. Registrski način implementacije torej ni nujno slaba izbira, izbira pa je potemtakem odvisna od potreb in zmožnosti organizacije.

#### **2.3.4.1 Izbira na podlagi vplivov na obstoječe aplikacije**

Kljub temu, da je prehajanje med alternativnimi rešitvami mogoče, vseeno ni nepomembno, da se organizacija že na začetku odloči za pravo rešitev. Torej tisto, ki ji bo v sprejemljivem času in stroškovnem okviru prinesla zadovoljivo rešitev. To je predvsem pomembno pri vpeljavi MDM v že delujoč IS organizacije, kjer si ne moremo privoščiti načrtovanja od začetka, ampak moramo upoštevati obstoječe stanje. Kot pomoč pri odločanju na tem mestu lahko predlagamo preprosto vizualno analizo, kjer na podlagi trenutnih zahtev aplikacij, ki se bojo povezovale na novo vzpostavljeni MDM, ocenimo pričakovanja in določimo kateri implementacijski način nam najbolj ustreza. Kriterije oziroma lastnosti organizacija seveda lahko določa sama, za osnovo pa si lahko vzame v strokovni literaturi opredeljenih šest [15]:

- Število atributov matičnih podatkov (Š): V kolikor imamo opravka zgolj z nekaj atributi, je navaden register primerna rešitev. Kadar so matični podatki kompleksnejše entitete ali v času pričakujemo razširitev množice, takrat se je bolje premikati proti transakcijskemu načinu. Višje kot je število atributov, bolj je kriterij pomemben.
- Usklajevanje (U): Sorazmerno z večanjem števila virov, iz katerih se »napaja« register, raste tudi kompleksnost pravil s katerimi določamo veljavno vrednost posameznega atributa. Bolj kot so pravila kompleksna, manj so primerna za

implementacijo registra in bolj se približujemo transakcijskemu MDM. Višanje potrebe po usklajevanju viša tudi pomembnost tega kriterija.

- **Sinhronizacija (S):** Kadar govorimo o transakcijskih aplikacijah, je bistvenega pomena, da so matični podatki enaki, ne glede na to s katerega zornega kota gledamo na njih. Pri teh aplikacijah je sinhronizacija vseh morebitnih kopij nujna in stopnja sinhronizacije visoka. Na nasprotni strani so analitične aplikacije, ki se navadno navezujejo na podatke podatkovnih skladišč in pri katerih se proces polnjenja izvaja na dnevnem nivoju. Tukaj zadošča usklajevalni način implementacije, saj ni potrebe po zadnjem stanju matičnih podatkov. Bolj kot je kriterij sinhronizacije pomemben, bližje smo transakcijskem načinu.
- **Dostop (D):** Omenili smo, da v organizaciji prav matični podatki predstavljajo podatke z najvišjo vrednostjo. Zato je pomembno, da imajo dostop do njih samo pooblaščen osebe. Bolj kot so podatki centralizirani, lažje je preverjati, kdo dostopa do podatkov, in določati, do katerih podatkov ima dostop. Ker so podatki najbolj centralizirani prav v transakcijskem MDM, je sklep jasen: večjo kontrolo kot rabimo za dostop do podatkov, bolj se gibljemo proti transakcijskemu vozlišču.
- **Kompleksnost servisov (K):** Bolj kot so servisi, s pomočjo katerih manipuliramo s podatki, kompleksni in več poslovne logike kot je vsebovane v njih, bolj se nagibamo proti transakcijskemu načinu.
- **Zmogljivost (Z):** Ker imajo silosne aplikacije vsaka svojo podatkovno bazo, do zmogljivostnih težav navadno ne prihaja. Drugače je pri transakcijskem MDM, kjer so podatki centralizirani na enem mestu. Najrazličnejše aplikacije poskušajo dostopati do istega nabora podatkov, zato se lahko pojavi problem ozkega grla. Ker v MDM govorimo o matičnih podatkih, katerih lastnost je počasno spreminjanje skozi čas, težava ozkega grla ni tako kritična, da je s preišljenim načrtovanjem ne bi mogli razrešiti [15]. Bolj kot je zmogljivost procesiranja transakcij pomembna, bolj se nagibamo proti registrskemu implementacijskem načinu.

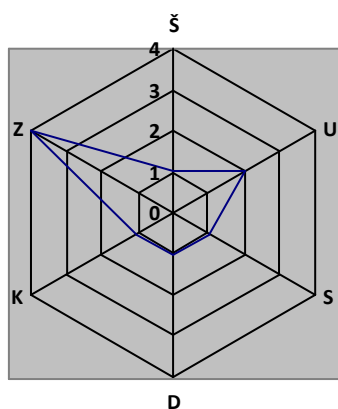
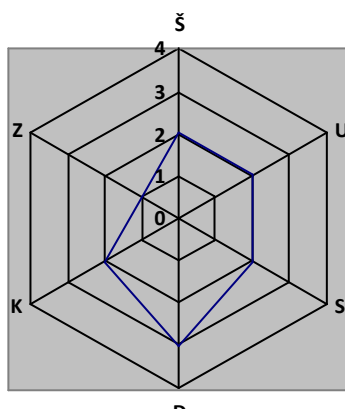
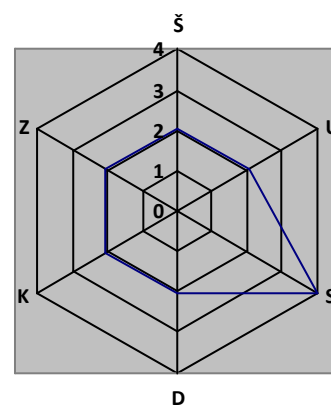
Za vsako aplikacijo torej pridobimo nabor ocenjenih vrednosti za vse zgoraj opisane kriterije. Vsak kriterij ocenimo z oceno od 1 do 5 na podlagi ocenjene pomembnosti kriterija, ki jih prikazuje tabela 3. Kriterij ocenjen z 1 pomeni najnižjo stopnjo pomembnosti, kriterij ocenjen s 5 pa najvišjo stopnjo pomembnosti.

**Tabela 3: Ocenjene vrednosti pomembnosti.**

Vrednost kriterija	Opisna vrednost kriterija
0	Nepomembno
1	Manj pomembno
2	Pomembno
3	Bolj pomembno
4	Zelo pomembno

Pri prvih petih opisanih kriterijih višja stopnja pomembnosti pomeni višjo usmerjenost kriterija proti transakcijskemu načinu implementacije. Izjema je le kriterij zmogljivosti, kjer višja stopnja pomembnosti usmerja proti registrskemu načinu.

Po pridobljenih ocenah lahko za vsako aplikacijo izrišemo zvezdni diagram, kjer na posameznem kriteriju označimo oceno in pridobljene točke povežemo s krivuljo. Bolj kot krivulja poteka po notranjosti diagrama, bolj se kot primernejši kaže registrski način

**Slika 11: Zvezdni diagram aplikacije poročanja.****Slika 13: Zvezdni diagram aplikacije HRM.****Slika 12: Zvezdni diagram aplikacije CRM.**

implementacije. Nasprotno je krivulja, ki poteka po zunanosti diagrama, znak za transakcijsko vozlišče. Na slikah 11, 12 in 13 so s svojimi značilnostmi prikazani trije zvezdni diagrami za primere aplikacij. Kot je razvidno nobena izmed njih v popolnosti ne ustreza nobenemu izmed skrajnih implementacijskih načinov, kar se v praksi tudi najpogosteje dogaja. Razberemo lahko, da ima aplikacija CRM večji poudarek na sinhronizaciji, saj mora vedno odražati zadnje veljavno stanje, medtem ko lahko aplikacija za poročanje na podatke zadnjega stanja počaka do naslednje osvežitve podatkovnega skladišča. Po drugi strani je zmogljivost na preizkušnji ravno pri izdelavi kompleksnih poročil, medtem ko, zaradi nizkega

števila transakcij, v aplikaciji HRM zmogljivost ni bistvenega pomena. Pri izbiri moramo upoštevati vse lastnosti aplikacij in zadostiti njihovim potrebam. Glede na opisani primer bi se organizacija lahko odločila za kombinirani implementacijski slog, kjer bi se stanje v aplikaciji CRM osveževalo v realnem času, medtem ko bi aplikacija HRM in DW, podatke prevzemala s pomočjo periodičnega osveževanja. DWH in lokalno podatkovno bazo aplikacije HRM bi tako lahko osveževali s pomočjo postopkov ETL, kar pomeni, da v aplikaciji ne bi bili podvrženi bistvenim posegom v delovanje.

## 2.4 Postopek združevanja zapisov strank

V predhodnih poglavjih smo veliko pisali o združevanju zapisov, ki predstavljajo isto stranko. Ta razdelek pa je namenjen obravnavi tematike samega postopka. Zاپise, ki v MDM prihajajo z različnih virov, je potrebno primerjati z obstoječimi zapisi v MDM in ugotoviti ali zapisi ustrezajo že obstoječi stranki ali predstavljajo nove.

Za vsak zapis, ki se pojavi na vhodu postopka združevanja, najprej preverimo ali ustreza minimalnim pogojem za avtomatsko obdelavo. Pogoji, s pomočjo katerih se določa kateri zapisi so primerni za avtomatsko obdelavo, so navadno sestavljeni iz množice kombinacij atributov matičnih podatkov. Da zapis izpolnjuje zahtevani pogoj pomeni, da ima vrednosti v vseh zahtevanih atributih. Primer takega nabora je prikazan v tabeli 4, kjer mora zapis, ki kandidira, zadostiti vsaj enemu od postavljenih pogojev.

**Tabela 4: Minimalni pogoji avtomatskega združevanja.**

Številka	Ime	Priimek	EMŠO	Datum rojstva	Kraj rojstva	Naslov	Telefonska številka	Elektronski naslov
1			+					
2	+	+		+	+			
3	+			+		+		
4	+	+		+			+	
...	+	+						+

V primeru, da zapis ne zadosti nobenemu od navedenih pogojev, pomeni, da se množica atributov z izpolnjenimi vrednostmi ne prekriva z nobeno kombinacijo v pogojih. Tak zapis se zavrne z zahtevo po dopolnitvi, ali pa se ga umakne v čakalno vrsto in obvesti skrbnika o novem zapisu, ki potrebuje njegovo pozornost.

### 2.4.1 Skupine atributov

Iz primera v tabeli 4 lahko vidimo, da vsi atributi niso medsebojno enakovredni. EMŠO je tako selektiven atribut, da že sam po sebi enolično določi stranko, medtem ko datum rojstva za enolično identifikacijo ne zadostuje. Na splošno delimo attribute v tri glavne skupine [3]:

- **Identitetni atributi:** So atributi s pomočjo katerih lahko identificiramo posamezno stranko. To so že omenjeni EMŠO, ime, priimek, davčna številka, številka certifikata z izdajateljem, številka računa, telefonska številka in podobni atributi.
- **Diskriminacijski atributi:** To je skupina atributov s pomočjo katerih ne moremo direktno povezovati dveh zapisov, lahko pa ju razlikujemo. Tak primer je datum rojstva, kjer imata oče in sin lahko enaka ime in priimek, se pa razlikujeta po letnici rojstva. Tipično imajo diskriminacijski atributi nižjo kardinalnost kot identitetni, zato jih ne moremo uporabljati za iskanje ujemanj med dvema zapisoma, lahko pa znižajo verjetnost za lažno pozitivna ujemanja. V to skupino poleg datuma rojstva lahko uvrstimo še spol, datum smrti itd.
- **Atributi tipa zapisa:** S pomočjo teh atributov ugotavljamo za kakšen tip zapisa gre in katera identifikacijska pravila naj za ta zapis uporabimo. Primeri teh atributov so lahko državljanstvo, tip stranke kateri pove ali gre za fizično ali pravno osebo itd. Kadar naletimo na pravno osebo, ne moremo računati na podatke o priimku in spolu, zato nas ti atributi usmerijo v izbiro prave podmnožice množice pravil.

### 2.4.2 Primerjanje zapisov in njihovih atributov

Sedaj, ko smo attribute matičnih podatkov grupirali v različne tipe nam ostane, da določimo na kakšen način bomo te attribute medsebojno primerjali. Za izvedbo primerjave imamo na voljo veliko pristopov, omenimo pa zgolj nekaj najpogostejših:

- **Natančno ujemanje [3]:** Tukaj govorimo o popolnem ujemanju vrednosti dveh atributov. Pri numeričnih in datumskih atributih je preverjanje ujemanja enostavno, pri znakovnih nizih pa se lahko sprejme dogovor, da se pred primerjavo naredi konverzijo v velike črke, odvečne presledke in posebne znake pa se eliminira.
- **Ujemanje na podlagi »zdravega razuma« (ang. *common sense*) [3]:** Na podlagi vsebinskega znanja analitikov se določi neka standardna pravila, s pomočjo katerih želimo izboljšati iskanje ujemanj med atributi. Pravila so uporabna zgolj za določen tip zapisov, ki ga določimo s pomočjo atributov tipa zapisa. Če na primer ugotovimo, da je stranka iz Slovenije, lahko ujemanje med naslovoma nekoliko obogatimo s pravili, da je »ul.« in »u.« enako kot »ulica«, pri imenih fizičnih oseb pa, da ima »ml.« enak pomen kot »mlajši«. Opisani primer se v literaturi pojavlja tudi kot iskanje ujemanj s pomočjo odstranjevanja sinonimov [29].

- Ujemanje na podlagi baze znanja [3]: V primerih, kadar že imamo pripravljeno bazo znanja, ki jo lahko uporabimo pri obogatitvi podatkov, to lahko s pridom izkoristimo tudi za iskanje ujemanj med atributi. Če ima stranka znano hišno številko in ulico, ne pa tudi naselja, lahko poiščemo število vseh možnih naselij za to ulico in številko. V primeru, da odkrijemo natančno en zadetek, lahko podatke o naselju in občini oplemenitimo in tako omogočimo primerjavo med več atributi.
- Statistični pristop [22]: Gre za poizkus primerjanja atributov na podlagi predhodno pridobljenih statističnih porazdelitev atributov, ki nastopajo v primerjavi. Znano je, da je ime Janez v Sloveniji precej pogosto in zato zelo neselektivno. Če to isto ime iščemo v Turčiji je velika verjetnost, da smo že samo s primerjanjem po imenu našli pravo osebo. S statističnega vidika so uspešne primerjave na manj pogostih vrednostih atributa bolj signifikantne od ujemanj na bolj pogostih vrednostih. Po ocenah nam upoštevanje takšnih verjetnostnih pristopov doprinese 4 do 10 odstotno izboljšanje natančnosti [3].
- Mehko ujemanje (ang. *fuzzy matching*) [29]: Bistvo metode je ocenjevanje oddaljenosti med vrednostima atributov v skupnem prostoru. Gre za to, da se na nek način določi matematično funkcijo kot cenilko, ki vrednosti atributov preslika v skupni prostor ter pove koliko sta vrednosti medsebojno oddaljeni. Pri izbiri cenilke imamo seveda veliko možnosti, saj lahko temelji na najrazličnejših konceptih. Kot primer lahko navedemo fonetično analizo dveh atributov pri kateri so besede, ki se slišijo podobno bliže, ali pa podobnost izračunamo na podlagi najmanjšega števila zamenjav znakov za pretvorbo enega niza v drugi niz.

V množici različnih pristopov se nam postavi vprašanje, katerega izbrati. Na splošno je to odvisno od posameznega atributa ter podatkov, ki jih hrani, zato je splošno navodilo težko podati, lahko pa podamo sledečo usmeritev. Kadar imamo na voljo veliko identitetnih atributov, ki so zastopani v vseh virih, so deterministične metode navadno boljša izbira. Kadar temu ni tako, se implementacija postopkov za iskanje ujemanja oprime statističnih pristopov, mehke metode pa se nekako najboljše obnesejo pri procesiranju velike količine zapisov z veliko atributi [22].

Rezultat primerjave dveh atributov je lahko binarna vrednost, ki pove ali se atributa ujemata ali ne, ali pa ocena (ang. *score*), ki podaja neko vrednost ali stopnjo ujemanja med atributoma. Kadar za primerjavo uporabljamo ocene, moramo določiti tudi neko mejno vrednosprag (ang. *threshold*), ki bo nabor možnih vrednosti razdelila na dva dela. Če je ocena nižja od praga, bomo sklepali, da se atributa ne ujemata v dovoljšni meri, da bi ju proglasili za enaka.

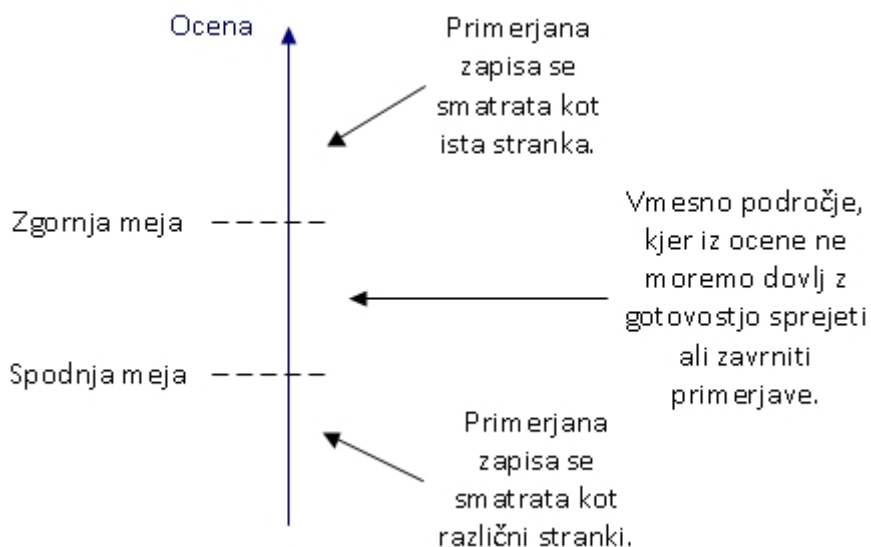
Nasprotno velja, da atributa smatramo kot dovolj podobna, če je ocena njunega ujemanja nad pragom.

Naslednji korak pri združevanju zapisov strank je določitev pravil za primerjanje zapisov na podlagi zgoraj opisanih pravil za primerjanje posameznih atributov. Podobno kot pri primerjavi atributov je tudi pri primerjavi zapisov rezultat možno izraziti kot binarno vrednost ali oceno. Funkcijo primerjave lahko torej sestavimo na naslednje tri načine [3]:

- Binarno pravilo ujemanja za attribute in binarno pravilo za ujemanje zapisov: Ker so vsi parametri primerjalne funkcije zapisa binarni in funkcija kot rezultat prav tako vrača binarno vrednost, se za dosego želenega rezultata funkcije določa kateri parametri naj se primerjajo in delež izpolnjenih glede na število. Gre torej za to, da se na podlagi razmerja med številom izpolnjenih primerjav med atributi in številom vseh atributov, ki nastopajo v funkciji, določi ali smatramo zapisa kot enaka ali ne. Funkcijo lahko dopolnimo do te mere, da posamezne attribute medsebojno grupiramo podobno kot pri določanju minimalnih kriterijev za avtomatsko primerjavo v tabeli 4. Težava takih pravil je, da hitro postanejo obsežna in težko razumljiva.
- Binarno pravilo ujemanja za attribute in ocenjevalno pravilo za ujemanje zapisov: Ta princip deluje in je, kar se tiče primerjanja atributov, enak predhodnemu. Razlika se pojavi pri prenosu vrednosti v funkcijo za primerjanje zapisov. Vsakemu atributu lahko določimo utež, ki poveča oziroma zmanjša vpliv ujemanja posameznega atributa na končno primerjalno funkcijo. Atributa kot sta EMŠO in davčna številka bosta zagotovo deležna višje uteži kot na primer datum in naslov stalnega prebivališča. Takšna ocena ocenjevalne funkcije je torej vsota binarnih vrednosti primerjav atributov pomnoženih s pripadajočimi utežmi. Ocena kot taka ni več binarna vrednost, zato je potrebno, za interpretacijo ali gre za ujemanje ali neujemanje zapisov, določiti še mejno vrednost.
- Ocenjevalno pravilo ujemanja za attribute in ocenjevalno pravilo za ujemanje zapisov: Tukaj imamo opravka z ocenami na dveh nivojih. Najprej imamo ocene primerjav posameznih atributov in nato še skupno oceno za primerjavo celotnega zapisa. Prav tako kot ocene se tudi uteži lahko postavijo na obeh nivojih. Smisel uteži na nivoju posameznega atributa smo že pojasnili, posameznemu pravilu pa lahko dodamo utež, kadar nad določenim zapisom izvedemo več primerjav in ocene združimo v eno.

Poleg določanja uteži za posamezne primerjave ima velik vpliv na proces ujemanja tudi pomikanje praga. Kadar govorimo o višanju praga se moramo zavedati, da postaja pravilo za določanje, ali se zapisa ujemata, vedno bolj konzervativno. Ko mejno vrednost zvišujemo, je zaradi strožjega kriterija med zapisi, ki so označeni kot ujemanja, vedno manj napačnih.

Znižuje se torej število lažno pozitivnih ujemanj. Hkrati s tem pa se nam na drugi strani povečuje število lažno negativnih ujemanj, torej takih, ki jih nismo razpoznali kot duplikate, pa bi jih morali. V tem primeru si lahko pomagamo z določitvijo dveh mej, ki jih postavimo



**Slika 14: Razdelitev ocen primerjanja zapisov z dvojnimi pragom.**

kot prikazuje slika 14. Zgornjo mejo postavimo tako, da kot duplikate okarakteriziramo zgolj tiste zapise, za katere resnično verjamemo, da so si medsebojno enaki. Vse primerjave, ki se uvrstijo nad to mejo torej proglasimo za ujemanja in tako za te zapise v MDM že obstajajo stranke. Na drugi strani primerjave, ki se znajdejo pod spodnjo mejo, proglasimo kot neuspela ujemanja in zapisi se v MDM prenesejo kot nove stranke. Ker sta obe meji nekoliko razmaknjeni, se med njima vzpostavi vmesno polje. Ideja je, da se v to vmesno področje polovi vse zapise, za katere ne moremo z gotovostjo trditi, da za njih v MDM že obstajajo stranke. Hkrati pa ne moremo biti prepričani, da bi naredili pravilno, če bi jih v sistem vnesli kot nove stranke. Medtem ko vsi zapisi nad zgornjo in zapisi pod spodnjo mejo v MDM sistem zapisa preidejo avtomatsko, se vse primerjave iz srednjega pasu posreduje skrbnikom sistema v ročno obravnavo.

### 2.4.3 Uparjanje na primeru

Za boljšo predstavo si oglejmo nekaj primerov preverjanj ujemanj na sliki 15. Prikazani sta dve množici podatkov. Zgornjo množico podatkov sestavljajo zapisi od 'a' do 'd'; to je skupek zapisov, ki vstopajo v MDM in jih označimo kot množico K. Spodnja množica zapisov, ki jo poimenujmo množica M, pa sestoji iz zapisov 1 do 3 in se že nahaja v MDM kar pomeni, da so že prestali proces preverjanja ujemanja in združevanja ter predstavljajo posamezne stranke organizacije.

K	Ime in priimek	EMŠO	Datum rojstva	Naslov	Tel. številka	Elektronski naslov
a	M. Novak		11.7.1995			
b	Krpan		12.1.1981	Sveta Trojica 13	041/654321	
c	Peter Klepec			Osilnica 54A, Osilnica		p.k@email.com
d		1505990500012			031/987654	j.n@email.com



M	Ime in priimek	EMŠO	Datum rojstva	Naslov	Telefonska številka	Elektronski naslov
1	Janez Novak	150599050001 2	15.5.1990	Nova cesta 1, Ljubljana	041/123456	j.n@email.com
2	Manca Novak	210799505012 3	21.7.1995	Slovenska ulica 2		
3	Martin Krpan	120198150016 2	12.1.1981	Sveta Trojica 13, Pivka	041/654321	m.k@email.co m

**Slika 15: Primer iskanja ujemanj zapisov. Množici zapisov K in M.**

Potrebno je torej izvesti primerjavo izbranega zapisa iz množice kandidatov za vstop v MDM z zapisi, ki so že vključeni v MDM. V našem primeru bomo uporabili način z ocenjevalnim pravilom ujemanja za attribute in ocenjevalnim pravilom za ujemanje zapisov. Najprej določimo formulo, po kateri bomo določili oceno primerjave zapisov  $k \in K$  in  $m \in M$ . Kot že rečeno, je ocena ujemanja zapisov določena tako, da medsebojno primerjamo attribute na enakih položajih, pridobljene ocene primerno utežimo in jih seštejemo v skupno oceno. Formula za oceno posameznega zapisa se torej glasi:

$$Z(k, m) = \sum_{i=1}^n A(k, m)_{(i)} * W_{(i)},$$

kjer je  $n$  število atributov matičnega podatka,  $A(k, m)_{(i)}$  predstavlja oceno primerjave posameznega atributa,  $W(i)$  pa so uteži, s katerimi korigiramo izračunane ocene primerjave atributa in so prikazane v tabeli 5. Podane uteži so zgolj simbolne in se spreminjajo glede na način poslovanja organizacije in kakovost podatkov. Z njihovo pomočjo lahko skrbniki MDM natančno določijo pomembnost posameznih atributov in tako prilagajajo ocenjevalno funkcijo dokler ne dosežejo pričakovanega delovanja mehanizma uparjanja zapisov matičnih podatkov.

**Tabela 5: Primeri atributnih uteži.**

Atribut	Ime in priimek	EMŠO	Datum rojstva	Naslov	Tel. številka	elektronski naslov
<b>Utež atributa (W)</b>	0,5	1	0,4	0,5	0,3	0,7

Poleg formule za ocenjevanje zapisov je potrebno določiti še način primerjanja atributov. V našem primeru bomo attribute primerjali na tri načine. Atribut EMŠO bo zaradi kontrolne vsote, ki se že preverja na vnosnih maskah aplikacij, primerjan s pomočjo natančnega ujemanja. Tu imamo torej primer binarnega rezultata, ki bo rezultiral samo v vrednosti 0 ali  $1 \cdot W$ . Naslov bomo v prvem koraku s pomočjo ujemanja na podlagi baze znanja najprej poskušali dopolniti z morebitnimi dodatnimi podatki, nato pa ga bomo s pomočjo mehke metode primerjali z zapisom MDM. Rezultat te primerjave bo ocena, ki je realno število in ne več binarna vrednost. Za preostale primerjave iz primera bomo uporabili zgolj mehko primerjanje s pripadajočo utežjo.

Kot mehke mere za primerjanje vrednosti atributov se bomo na tem mestu poslužili primerjave med znakovno predstavljenima vrednostma atributov. Kot začetno točko vzamemo znakovni predstavitev primerjanega atributa obeh kandidatnih zapisov. V naslednjem koraku nad prvim zapisom izvajamo operacije zamenjave, izbrisa in dodajanja posameznega znaka tako dolgo, dokler prvi niz ne postane enak drugemu nizu in zraven štejemo potrebno število zamenjav. Dobljeno število predstavlja oddaljenost med vrednostma atributov. Sedaj lahko določimo oceno med primerjanima atributoma zapisov kot:

$$A(k, m)_{(i)} = \frac{D(m)_{(i)} - S(k, m)_{(i)}}{D(m)_{(i)}}, \text{ kjer je } D(m)_{(i)} \text{ dolžina znakovne predstavitve obstoječega}$$

atributa v MDM,  $S(k, m)_{(i)}$  pa predstavlja število potrebnih sprememb za pretvorbo vrednosti  $i$ -tega atributa zapisa  $k$  v vrednost  $i$ -tega atributa zapisa  $m$ . Na tem mestu ponovimo, da gre v zgornjem primeru zgolj za enega od načinov, s pomočjo katerega določimo, kako podobna sta

si atributa. Na konkretnem primeru zapisa b si pogledjmo, kako po navedeni formuli izračunajmo podobnosti  $Z(b, 1)$ .

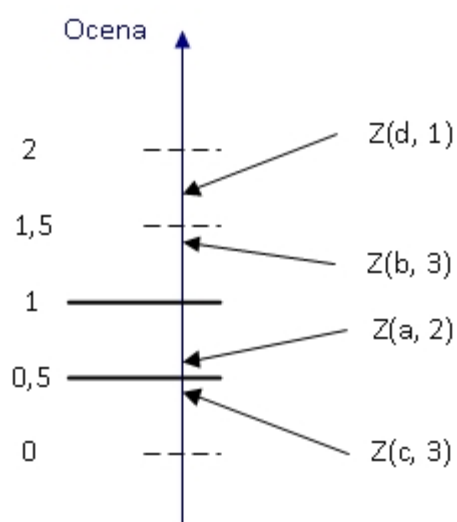
$$Z(b,1) = \frac{11-10}{11} * 0,5 + \frac{8-4}{8} * 0,4 + \frac{23-25}{23} * 0,5 + \frac{9-6}{9} * 0,3 = 0,3$$

Če ta isti izračun ponovimo še za ostale primerjave, pridemo do rezultata primerjav med zapisi, ki ga prikazuje tabela 6.

**Tabela 6: Rezultati izračuna podobnosti zapisov.**

Z(k,m)		Zapisi množice M		
		1	2	3
Zapisi množice K	a	0,43	<b>0,61</b>	0,28
	b	0,3	0,17	<b>1,41</b>
	c	0,08	-0,6	<b>0,43</b>
	d	<b>1,77</b>	0	0,38

Celoten izračun za preostale primerjave se nahaja v prilogi. Vsaka vrednost podaja rezultat primerjave  $Z(k, m)$ , ki je izvedena za vse primerjave med množicama K in M. Za vsako primerjavo zapisa k z vsemi zapisi množice M, kot relevantno vzamemo tisto, ki ima najvišjo vrednost podobnosti. Za kandidatni zapis a je torej merodajna primerjava  $Z(a, 2)$ , saj je njena ocena najvišja med vsemi  $Z(a, *)$ .



**Slika 16: Izračunane ocene glede na mejne vrednosti.**

Edino težavo imamo le pri zapisu a, čigar ocena ga je uvrstila v vmesno

Za vsak zapis  $k \in K$  smo torej pridobili par  $(k, m)$  za katerega je pripadajoča vrednost  $Z$  najvišja in tako predstavlja najboljšega kandidata za primerjanje novega in obstoječega MDM zapisa. Z zavedanjem, da so to zgolj ocene, je potrebno preveriti ali pridobljene ocene ustrezajo postavljenim mejnim vrednostim za potrditev ujemanja posameznega para. Tako kot je prikazano na sliki 16, bomo za naš primer privzeli, da sta meji postavljeni pri vrednostih 1 za zgornjo in 0,5 za spodnjo mejo. Zapis d in b glede na doseženo oceno ustrezata že obstoječim zapisom strank v MDM. Ocena  $Z(c, 3)$  zapisa c je dovolj nizka, da omenjeni zapis proglašimo kot novo in še neobstoječo stranko v MDM, saj je padel pod

področje in zato zanj avtomatična uvrstitev ni mogoča. V tem primeru bo, kot smo že omenili, potrebno posredovanje skrbnika podatkov, ki bo s svojim vsebinskim znanjem zapis uvrstil v pravo množico in, če se mu zdelo potrebno, korigiral uteži ali mejne vrednosti.

## 2.5 Zagotavljanje kakovosti podatkov

Kot smo že opisali, je za dobre odločitve in planiranje potrebno imeti dobre podatke. Tudi ko organizacija vpelje strategijo upravljanja metapodatkov in vzpostavi register metapodatkov, je vsebina tega registra navadno še vedno nepopolna ali vsebuje napake [3]. Prav zato kot komponento MDM smatramo tudi modul za upravljanje s kakovostjo podatkov (ang. *data quality* – DQ). Zelo težko je namreč iz slabih informacij, ki jih pridobimo iz virov, skonstruirati kakovosen zapis MDM, ki bi odražal dejansko stanje. Navadno se organizacije niti ne zavedajo, da je kakovost njihovih podatkov slabša od pričakovane. Po analizi, ki jo je opravil TDWI [24], skoraj polovica vprašanih organizacij meni, da so njihovi podatki bodisi točni ali vsaj zadovoljive kakovosti glede na potrebe. V realnosti se je izkazalo, da ima kar 44 odstotkov organizacij slabše podatke kot so pričakovali [7].

Zaradi slabih podatkov lahko sledijo napačne odločitve, ki pa organizacijo stanejo veliko sredstev. Staro razvojno pravilo pravi, da je vsaka denarna enota, vložena v načrtovanje, ekvivalentna 10 enotam vložnim v razvoj in 100 enotam vložnim v podporo [9]. Na to pravilo lahko gledamo tudi z vidika kakovosti podatkov. Vsak dodaten denar in čas, ki ga organizacija v začetku posveti zagotavljanju višje kakovosti podatkov, se hitro povrne v primerjavi s stroški povezanimi z odpravljanjem posledic slabih podatkov in posledično ceno sprejemanja napačnih odločitev. TDWI ocenjuje, da stroški, povezani s slabimi podatki, v ZDA na letni ravni znašajo 611 milijard ameriških dolarjev [24].

Po definiciji bi lahko rekli, da je: »kakovost podatkov ena ključnih komponent katerekoli uspešne podatkovne strategije in pobude za upravljanje podatkov. Prav tako je ena temeljnih zahtev, ki omogočajo upravljanje z matičnimi podatki in integracijo podatkov o strankah« [3, str. 112].

Glede na raziskave je pri vzpostavitvi podatkovnega skladišča kar 75 odstotkov dela vložnega v pripravo podatkov, preverjanje pravilnosti in postopke ETL in kar polovico tega dela je namenjenega čiščenju in standardizaciji podatkov [3]. Delež torej ni zanemarljiv in vredno se je problematike DQ lotiti nekoliko bolj sistematično. Čeprav sta z vsebinskega vidika postopka za zagotavljanje kakovosti podatkov v DWH in MDM medsebojno zelo podobna, obstajata dve bistveni razliki. Prva je, da se pri DWH vsi mehanizmi za spremljanje

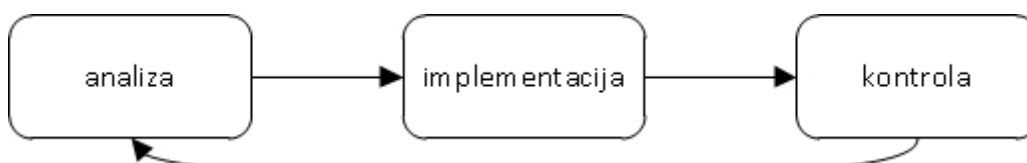
kakovosti izvajajo znotraj postopkov polnjenja skladišča (ETL) in so omejeni zgolj na skladišče samo, medtem ko se pri konceptu MDM kakovost podatkov preverja ne samo ob času polnjenja, ampak tudi med rednim poslovanjem organizacije. Druga razlika je ta, da v nasprotju z DWH, pri MDM tok podatkov poteka v obe smeri. Medtem ko se v DWH izvaja samo polnjenje, mora MDM poskrbeti tudi za osveževanje lokalnih kopij svojih virov.

Če želi organizacija po vzpostavitvi MDM spremljati kakovost svojih podatkov, si mora najprej postaviti metrike in tako določiti prostor v katerem bo sledila premikom metrik v času. To je za poslovanje bistveno, saj se v primeru, da se stanje ne spremlja, stopnja kakovosti podatkov v času poslabšuje. Stopnja upadanja kakovosti je seveda različna za različne organizacije in njihove načine poslovanja, vendar vseeno lahko govorimo o neki splošni stopnji, ki znaša okoli 2 odstotka na mesec [3]. Navedimo torej nekaj metrik, s pomočjo katerih lahko spremljamo kakovost podatkov [7]:

- Popolnost (ang. *completeness*): Kriterij veleva, da podatki nimajo manjkajočih vrednosti in da v MDM ne vnašamo praznih atributov. Poleg popolnosti podatkov trenutnega zapisa, organizacija skozi čas spremlja odstotek manjkajočih vrednosti za posamezne attribute ali entitete matičnih podatkov in na podlagi odstopanja od statistike odkriva problematične zapise.
- Veljavnost (ang. *validity*): Tukaj preverjamo, ali so podatki ustreznega tipa in dolžine, ali je podatek obvezen, ali mora biti vrednost določenega atributa unikatna in podobno. V kolikor so podatki vsebinsko strukturirani, kot na primer EMŠO ali številka bančnega računa, imamo na voljo posebne mehanizme kot so kontrolne vsote, da zagotovimo pravilnost vnesenih podatkov.
- Točnost (ang. *accuracy*): Zagotavljanje točnosti podatkov se mnogokrat izkaže kot težavna naloga. Ena od opciji je, da organizacija vodi statistiko podobnih vnosov in išče prevelika odstopanja od povprečja, ki bi pomenila opozorilo za skrbnike MDM.
- Konsistentnost (ang. *consistency*): Pri nekaterih implementacijah MDM se pogosto srečujemo z redundantnostjo podatkov. Na tem mestu lahko spremljamo kateri, v kolikšni meri in koliko časa so matični podatki v medsebojno redundantnih kopijah neusklajeni.
- Enoličnost (ang. *uniqueness*): To načelo pomeni, da vsak zapis predstavlja natanko en objekt v realnem svetu. Če to prenesemo na MDM-CDI to pomeni, da vsak zapis v MDM predstavlja natanko eno stranko organizacije. Organizacija v poslovanje lahko vpelje avtomatiziran postopek, ki na podlagi predhodno določenih pravil išče kandidate za podvojene zapise.

- Pravočasnost (ang. *timeliness*): Mero lahko določimo kot pretečeni čas med trenutkom, ko se je uporabniku pojavila potreba po določeni informaciji do trenutka, ko mu je informacija na voljo.

Ko ima organizacija na voljo kriterije, skozi katere spremlja kakovost matičnih podatkov, se lahko sistematično loti preverjanja DQ novih matičnih podatkov in odpravljanja napak v metapodatkih ter obstoječih matičnih podatkih, pravilih za združevanje matičnih podatkov iz različnih virov itd. Kot prikazano na sliki 17, postopek poteka iterativno skozi tri faze: analiza, implementacija in kontrola [4].



**Slika 17: Postopek izboljševanja kakovosti podatkov.**

V fazi analize se s pomočjo zgoraj opisanih kriterijev in informacij skrbnikov podatkov pregleda sumljive in napačne podatke. Faza implementacije je namenjena prilagoditvi postopkov uparjanja, sprememb metapodatkov in pravil ali zgolj sprememb podatkov na virih. V zadnji stopnji, ki je faza kontrole, je potrebno preveriti ali nove izboljšave slučajno niso v nasprotju z obstoječimi pravili. Po vsaki uspešni iteraciji so podatki in metapodatki pravilnejši, poslovanje bolj konsistentno, poročila pa odražajo stanje bolj podobno dejanskemu.

## 2.6 Obvladovanje podatkov

V sklop MDM sodi tudi obvladovanje podatkov (angl. *data governance*). »To je proces, osredotočen na zagotavljanja kakovosti, konsistence, uporabnosti, varnosti, dostopnosti in lastništva nad informacijami« [3, str. 109]. Za uspešno poslovanje je upravljanje z matičnimi podatki in integracijo podatkov o stranki potrebno dopolniti s celovito strategijo in procesom obvladovanja. Glede na hierarhijo upravljanja s podatki, ki jo prikazuje slika 18, je DQ vsebovan v sistemu MDM, obvladovanje podatkov pa z določanjem procesov in politik upravljanja s podatki zaobjema MDM [5].

Upravljanje z matičnimi podatki o strankah je torej širše področje kot zgolj zagotavljanje kakovosti matičnih podatkov in manj kot obvladovanje matičnih podatkov. Obvladovanje podatkov se dotika tako informacijske tehnologije, kot tudi poslovanja organizacije. Na strani informacijske tehnologije je usmerjeno k zagotavljanju čim učinkovitejše rabe podatkov in učinkovitosti delovanja razpoložljivih resursov, medtem ko na poslovni strani stremi k izkoriščanju podatkov s ciljem po izboljšavi poslovnih procesov in poslovanja organizacije [5].



**Slika 18: Hierarhija upravljanja s podatki.**

Pod termin obvladovanje podatkov štejemo upravljanje z metapodatki in poslovnimi pravili, določanje politik in procesov, vezanih na kakovost podatkov. Obvladovanje podatkov torej predstavlja ogrodje, ki nam omogoča periodično spremljanje in zviševanje kakovosti podatkov. Za izvajanje nalog organizacije navadno vzpostavijo posebno delovno ekipo, saj je količina in odgovornost delovnih nalog precej visoka.

### **2.6.1 Skupina za obvladovanje podatkov**

Gre za manjšo skupino ljudi, v kateri je poleg zagotavljanja multidisciplinarnega znanja njihovih članov, potrebno zagotoviti tudi kadre z določeno mero avtoritete in sposobnostjo vodenja. Odločitve in usmeritve skupine morajo biti utemeljene in na visokem kakovostnem nivoju, hkrati pa je potrebno zagotoviti njihovo dosledno izvajanje.

Ekipa mora neprestano stremeti k iskanju konsenza med osebjem odgovornim za informacijsko tehnologijo in željami poslovnih uporabnikov, hkrati pa zagotavljati pravilnost podatkov v MDM. Kadri, ki sestavljajo ekipo, morajo poznati poslovne procese in šifrate organizacije ter skupaj s poslovnimi uporabniki iskati ideje za njihove izboljšave. Hkrati je potrebno sodelovati z oddelkom IT in določiti poenotene načine na katere bodo uporabniki dostopali do podatkov, termine za vzdrževanje infrastrukture in posege v programsko opremo, preverjati zmožnosti IS itd.

Spodaj povzemamo nekaj nalog, ki jih mora biti sposobna opravljati skupina za obvladovanje [3]:

- Določitev in poenotenje standardov poimenovanj tabel, atributov, podatkovnih baz, datotek itd.
- Poenotena navodila za kreiranje servisov SOA in njihovih sporočil.
- Standardizacija razvoja in izbira orodij za modeliranje procesov, podatkovnih modelov, entitet, razredov itd.
- Določitev nabora dovoljenih načinov dostopanja do podatkov. V tem primeru imamo v mislih predvsem nabor tehnologij (SQL, JDBC, ...) s pomočjo katerih uporabniki pridejo do podatkov v bazi.
- Postavitev standardov za procesiranje porazdeljenih transakcij.
- Postavitev in spremljanje, v poglavju o zagotavljanju kakovosti podatkov, že omenjenih metrik kakovosti podatkov.
- Kreiranje predlog vnosnih mask in določitev standardov za validacijo vnesenih podatkov. S pomočjo poslovnih uporabnikov, ki imajo domensko znanje in poznajo vsebino, je potrebno postaviti omejitve, ki preprečujejo vnos nesmiselnih podatkov v sistem.
- Skrb za MDM-DCI. Tukaj gre predvsem za postavitev in upravljanje s podatkovnim modelom MDM ter pripravo in razpošiljanje ustrezne dokumentacije. Velja poudariti, da je na tem mestu treba zagotoviti zadostno vključenost vseh deležnikov, da lahko aktivno sodelujejo pri vzpostavljanju procesov obvladovanja podatkov MDM.
- Sodelovanje pri načrtovanju, upravljanju in dokumentiranju sistemov, ki predstavljajo vire MDM.
- Soudedežba pri postavljanju pravil, po katerih se izvajajo primerjave med zapisi, ki vstopajo v MDM, in zapisi, ki so že del tega sistema. Veliko o teh pristopih smo povedali v poglavju, kjer so opisani postopki združevanja zapisov o strankah.

Iz navedenega lahko razberemo, da je delo skupine za obvladovanje MDM zelo razgibano in polno sodelovanja z vsemi deležniki sistema MDM. Izkušnje kažejo, da imajo projekti, ki niso vodeni s strani dobre skupine za obvladovanje MDM, zelo zmanjšane možnosti za dosego zastavljenih ciljev [3].

## **2.6.2 Stopnje zrelosti obvladovanja upravljanja matičnih podatkov**

Da organizacija lahko spremlja in uspešno napreduje v svojem razvoju, se jo razvrsti v eno od stopenj zrelosti obvladovanja MDM. DataFlux predlaga model po katerem se organizacija razvrsti v eno od štirih stopenj zrelosti [23], ki jih bomo v nadaljevanju tudi na kratko opisali. Ko organizacija prehaja v naslednje stopnje, mora za vsak prehod vložiti določen napor in

sredstva, hkrati pa se ob prehodu v višjo stopnjo zrelosti znižajo tveganja in zvišajo koristi, ki jih je deležna. Vsako stopnjo opredeljujejo štiri komponente:

- **Ljudje:** Kdo so deležniki, ki so vključeni v tej stopnji in kaj se od njega pričakuje?
- **Politike:** Katera poslovna pravila morajo biti implementirana za ustrezno obvladovanje podatkov v tej stopnji?
- **Tehnologije:** Katere investicije v tehnologijo morajo biti izpeljane in katere tehnologije morajo biti na voljo?
- **Tveganje in koristi:** Opredeljena so tveganja, katerim je organizacija izpostavljena v tej stopnji, ter koristi oz. doprinosi, ki jih stopnja doprinese.

Poleg navedenih komponent je potrebno podati še kriterij za prehod v naslednjo stopnjo. Vsak prehod nosi določene zahteve, ki morajo biti izpolnjene za uspešen prehod. Značilnosti posamezne stopnje in pogoji za prehod v naslednjo stopnjo so predstavljeni v naslednjih razdelkih.

### **2.6.2.1 Prva stopnja – Nediscipliniranost**

To je stopnja, v kateri so podatki organizacije raztreseni po aplikacijah in v veliki meri medsebojno redundantni a neuskklajeni. Organizacija navadno niti ne zna oceniti kakovosti podatkov ali stroškov, ki so posledica njihove slabe kakovosti. Presenetljivo je, da je organizacij, ki sodijo v to skupino kar 30 odstotkov [23]. V takih organizacijah se pojavlja zgolj nekaj posameznih poslovođij, ki so motivirani za parcialno konsolidacijo podatkov predvsem za analitične namene [15].

Opredeljenost stopnje [23]:

- **Ljudje:** Kakovost je odvisna od posameznikov, ki najpogosteje niso poslovni analitiki in zato nimajo potrebnega vsebinskega znanja. Na voljo ni nobene skupine, katere primarna naloga bi bila skrb za kakovost podatkov, zato ti zanesenjaki delujejo vsak na svojem področju in med njimi ni nikakršnega medsebojnega usklajevanja. Poslovođstvo ne posveča pretirane pozornosti kakovosti podatkov oz. se problemov niti ne zaveda in celotno krivdo za napake v podatkih pripisujejo IT oddelkom [23]. Obstaja zgolj zavedanje, da ima organizacija redundantne podatke, ki so razpršeni po več aplikacij [15].
- **Politike:** Proces spremljanja kakovosti podatkov ne obstaja, podatki pa so organizirani silosno. Zaradi tega se s slabimi podatki organizacije ukvarjajo samo kadar se za to pojavi potreba, reševanje težav pa običajno poteka z direktnimi intervencijami in brez vnaprej izdelanega načrta [15].

- Tehnologija: Na voljo ni nobenega mehanizma za profiliranje, analiziranje in revizijo podatkov. Podatki so silosno urejeni in čiščenje podatkov se izvaja samo na določenih ozkih podmnožicah.
- Tveganje in koristi: Tveganja so v takih razmerah seveda visoka. Organizacija lahko celo izgublja stranke ali sprejema napačne strateške odločitve, krivdo pa je težko komurkoli pripisati, saj naloge zagotavljanja kakovosti niso nikomur dodeljene. Koristi za tak način vodenja kakovosti podatkov so minimalne in kratkotrajne, saj se po manjših popravkih v podatke hitro pričejo vračati nove napačne vrednosti.

Ko začne organizacija seštevati stroške, ki so posledica nekakovostnih podatkov, se lahko prične proces približevanja drugi stopnji. Organizacija mora postaviti cilje obvladovanja podatkov in identificirati ključne entitete, ki bodo podvržene obvladovanju. Potrebno je uvesti profiliranje, standardizacijo in verifikacijo podatkov na nivoju celotnega IS organizacije. Iz zbranih ugotovitev mora organizacija pripraviti nabor pravil za kakovost podatkov in jih centralizirano shraniti na enem mestu. Ta pravila niso sama sebi namen, ampak morajo biti uporabljena za preverjanje kakovosti podatkov v vseh razpoložljivih aplikacijah.

### **2.6.2.2 Druga stopnja – Reaktivnost**

Kot že samo ime pove se organizacija ukvarja z napakami šele, ko do njih pride. Takrat se prične iskati vzroke za napako in pripravljati popravke v postopkih preverjanja in samih silosnih aplikacijah, kjer je do napake prišlo. Ta skupina je daleč največja, saj v njo sodi od 40 do 50 odstotkov organizacij [23]. Integracija aplikacij, razen v nekaj morebitnih izjemah, v tej stopnji še ni izvedena v celoti.

Opredeljenost stopnje [23]:

- Ljudje: S kakovostjo podatkov se navadno ukvarja nekaj tehničnega osebja, ki ima neposreden dostop do baz podatkov. Namesto politik, ki veljajo na nivoju celotne organizacije, osebje odgovorno za kakovost, še vedno uporablja ozko namembne lokalne procese s pomočjo katerih skuša zagotavljati kakovost podatkov.
- Politike: Pravila za obvladovanje podatkov se pojavijo, vendar se popravki še vedno dogajajo zgolj kadar se težave pojavijo. Postopki in vloge pri odpravljanju napak v podatkih so vsaj na nivoju posameznega oddelka standardizirani.
- Tehnologija: Orodja za analizo kakovosti podatkov so na voljo in se uporabljajo na nivoju posamezne silosne aplikacije. Nekateri oddelki že postopoma uvajajo integracijo podatkov, ki pa predstavlja manjšino vseh podatkov potrebnih za integracijo. Kadar prihaja do poizkusov upravljanja matičnih podatkov, se to navadno dogaja v DWH [15].

- **Tveganje:** Tveganje je še vedno visoko saj integracija podatkov, razen nekaterih redkih primerov, še ni izvedena in zato se še vedno lahko pojavljajo razlike med podatki posameznih silosnih aplikacij.

**Koristi:** Prednosti obvladovanja podatkov se kažejo v zelo omejenem področju. Navadno so prednosti opazne na nivoju enega poslovnega področja oz. nekaj posamičnih vodstvenih kadrov, ne občuti pa jih celotna organizacija.

Za prehod v proaktivno stopnjo je potrebno sprejeti novo strategijo na najvišjem nivoju poslovodstva. Potrebno je poenotiti oz. prilagoditi postopke za upravljanje kakovosti podatkov med poslovnimi enotami, kar včasih ni enostavno. Vsak oddelek ima svoja utečena pravila in procedure, ki mu v tem trenutku zadovoljivo služijo in po katerih se ravna. Prav zato je tukaj bistvena podpora vodstva, ki poseže nad parcialne interese in razvoj usmeri v skupno dobro celotne organizacije. Na tem mestu ključno vlogo odigra že omenjena skupina za obvladovanje podatkov, ki predstavlja vezni člen med oddelki in odreja način ter tempo dela. Kot tehnološki pogoj za napredovanje se od skupine za obvladovanje pričakuje vzpostavitev mehanizma za integracijo podatkov. Navadno se uvede storitvena plast SOA, kjer so zbrane vse storitve, preko katerih aplikacije komunicirajo z MDM. Posebna pozornost gre poenotenju mehanizmov za spremljanje kakovosti podatkov, za katere prav tako nosi odgovornost skupina za obvladovanje.

### **2.6.2.3 Tretja stopnja – Proaktivnost**

V tej stopnji se nahaja manj kot 10 odstotkov organizacij [23]. S prehodom v to stopnjo je narejen bistven napredek iz stopnje, kjer je organizacija napake reševala šele v trenutku nastanka, v stopnjo, kjer težave zaznamo in odpravimo preden pride do napak. Ta stopnja pomeni, da ima organizacija operativno transakcijsko MDM-CDI rešitev [15].

Opredeljenost stopnje [23]:

- **Ljudje:** Vzpostavljena je skupina za obvladovanje podatkov, ki medsebojno povezuje oddelke organizacije in je gonilna sila krepitve pozicije sistema MDM. Poslovodje se počasi začinjajo zavedati vrednosti, ki jo MDM prinese v poslovanje, saj se kakovost podatkov zelo izboljša, procesi pa se postopoma stabilizirajo.
- **Politike:** Standardizacija in procesi so vpeljani do te mere, da je morebitne težave moč zaznati preden se te pojavijo. Točno se ve kdo kaj počne in kakšne so zadolžitve, saj so procesi centralizirani in se izvajajo. Proces vodenja kakovosti je poenoten in vpeljan v vseh oddelkih organizacije [15].
- **Tehnologija:** Močan poudarek je na operativni servisni plasti, ki omogoča veliko stopnjo ponovne uporabe storitev in olajša nadaljnji razvoj aplikacij organizacije. Integracije so hitrejše, bolj uspešne in zahtevajo manj vzdrževanja.

- Tveganje in koristi: Tveganja za napake se v času zmanjšujejo, saj se glede na periodično analiziranje podatkov in dopolnjevanje pravil, kakovost podatkov povečuje. Obratno sorazmerno z višanjem stopnje kakovosti podatkov se niža verjetnost sprejemanje napačnih odločitev vodstva organizacije.

Ko imamo operativen MDM in težave v podatkih zaznamo, preden te nastopijo, se seveda vprašamo, kam organizacija sploh še lahko napreduje. Odgovor leži v avtomatizaciji poslovnih procesov. Sedaj, ko so podatki o strankah vedno na voljo s klicem servisa, lahko te storitve vključujemo direktno v poslovne procese.

#### **2.6.2.4 Četrta stopnja – Vodenje**

Zakaj bi uslužbenka na bančnem okencu za odprtje novega bančnega računa morala najprej vnesti novega komitenta v CRM aplikacijo in šele nato v drugi aplikaciji odpreti bančni račun, če lahko oboje stori v enem poslovnem procesu. Take medsebojno ločene poslovne procese lahko združujemo z orodji za modeliranje procesov (ang. Business Process Modeling - BPM). Opredeljenost stopnje [23]:

- Ljudje: Skupina za obvladovanje podatkov ima vso podporo poslovdij organizacije in v svojem delovanju redno sodeluje s poslovnimi uporabniki, razvijalci aplikacij in skrbniki podatkovnih baz.
- Politike: Pred vpeljavo novosti se skrbno preuči posledice in predvidi možne neželene učinke na poslovanje. Za zagotavljanje kakovosti podatkov se nad MDM izvajajo avtomatizirane implementacije politik, poenotenih na nivoju celotne organizacije. Vsa odstopanja od pričakovanih okvirjev se rešujejo sproti in individualno, nikakor pa se ne dopušča nabiranja zahtevkov in nato njihovega paketnega reševanja.
- Tehnologija: Vse funkcionalnosti za zagotavljanje kakovosti podatkov in upravljanja s poslovnimi pravili so poenotene in na voljo kot servisne storitve.
- Tveganje in koristi: Tveganja v tej stopnji postanejo zelo nizka. Ker je poskrbljeno za kakovost podatkov in so v spremembo poslovnih pravil vključene vse zainteresirane strani, so morebitne napake bistveno manj verjetne. Organizacija je za prehod v četrto stopnjo nagrajena s kakovostnimi podatki in pravilnostjo poročil, kar daje bodro osnovo za usmerjanje poslovanja. Ne gre zanemariti niti povezovanja in izmenjave vsebinskega znanja med posameznimi oddelki in tako iskanja novih, boljših pristopov poslovanja.

Opisani model služi kot merilo za umestitev trenutnega IS organizacije v eno izmed stopenj in postavitvev izhodiščne stopnje za pripravo plana uvedbe MDM. Poleg tega model lahko uporabimo za pridobivanje informacij o tem, kako že implementirane komponente lahko

uporabimo v smeri boljšega upravljanja z matičnimi podatki, kot tudi za pridobitev informacije katere komponente mora organizacija še vzpostaviti, da bo njen način poslovanja prešel v bolj dovršeno stopnjo [15].



### 3 Vpeljava projekta

Vzpostavitev in vpeljava sistema MDM v poslovanje je zahteven proces. Potrebno je namreč izbrati najprimernejši način implementacije, vpeljati veliko novih tehnologij in napraviti preskok v mišljenju uporabnikov ter vodstva organizacije [3, 23]. Organizacije navadno nimajo ustreznega znanja in kadrov, zato se velikokrat izkaže za bolj stroškovno učinkovito, da se namesto lastnega učenja na napakah poslužijo zunanjih izvajalcev ali svetovalcev [15].

Glede na to, da so projekti MDM veliki po obsegu in finančnih vložkih, imajo neuspeli projekti zelo negativne posledice za organizacije in udeležence na projektu [3]. Ker projekt poseže globoko v poslovne procese, napačne odločitve lahko zelo otežijo ali celo onemogočijo normalno poslovanje in s tem prinesejo težko popravljive posledice. Sodelujoči na takih projektih v očeh preostalih sodelavcev izgubijo kredibilnost in zaupanje, kar še oteži sodelovanje ter nadaljnje korake pri reševanju nastale situacije.

Ključne razmisleke vodstva bi lahko zbrali okoli naslednjih točk [3, 15]:

- Vizija poslovanja: Organizacija mora imeti jasno vizijo svojega poslovanja in kako se le-ta sklaplja z uvedbo MDM. Poznana mora biti trenutna stopnja zrelosti obvladovanja podatkov in določen cilj, kamor organizacija želi napredovati.
- Identifikacija ključnih sprememb za doseg cilja: Zbrati je potrebno pomembnejše manjkajoče tehnologije, spremembe poslovnih procesov in manjkajoče kompetence kadrov, ki jih bo potrebno zagotoviti med implementacijo. Premostitev razkoraka do ciljne stopnje zrelosti obvladovanja podatkov je lahko zelo velika.
- Finančno breme in donosnost uvedbe projekta: Potrebno je upoštevati tako stroške uvedbe MDM kot tudi stroške z naslova zmanjšane produktivnosti. Ker so ključni kadri globoko vpeti v proces uvedbe, bodo imeli manj časa za tekoče poslovanje. Pomembna je tudi informacija o času, potrebnem za povrnitev investicije. Nesmotrno je vlagati v projekt, če nimamo jasne predstave o tem, kako nam bo investicija pomagala in v kolikšnem času nam bo povrnila vložena sredstva.
- Organiziranost projekta: Ker je uvedba MDM daljši projekt, je smiselno razmisliti o stopenjskem pristopu, ki bi omogočal koriščenje vsaj nekaterih funkcionalnosti čimprej. Na ta način zmanjšamo finančno breme, saj imamo v kasnejših fazah projekta že nekatere koristi predhodnih faz. S stopenjskim pristopom se pomembno zniža tudi verjetnost neuspeha uvedbe. Krajše faze so veliko bolj obvladljive, uporabniki se postopoma uvajajo v nov način poslovanja, hkrati pa organizacija sproti ugotavlja ali

projekt izpolnjuje njena pričakovanja in želje. Določiti je potrebno kriterije, s pomočjo katerih bomo ocenili, ali je posamezna faza uspešno zaključena.

- Identifikacija kritičnih točk uvedbe: Vsaka organizacija ima v svojem poslovanju nekaj ključnih aplikacij, ki podpirajo glavne poslovne procese. Ker so ti procesi bistveni za obstoj, je posegom vanje potrebno posvetiti še posebno pozornost. Načrtovanje in izdatno testiranje sta bistveni zagotovili, da spremembe ne bodo vodile v neželene posledice.
- Uskladitev vzporednih iniciativ: Ker sistem MDM predstavlja bistvo IS organizacije, je vse morebitne druge iniciative potrebno prilagoditi pravilom MDM. Vse nove aplikacije, ki v organizacijo vstopajo vzporedno z uvedbo MDM, je potrebno pripraviti na spoštovanje novih politik obvladovanja matičnih podatkov.

### 3.1 Vzpostavitev delovne skupine

Ko organizacija sklene, da bo v svoje poslovanje vpeljala MDM-CDI, je primeren pristop vzpostavitev manjše delovne skupine, katere naloga je koordinacija opravil, povezanih z vpeljavo [3]. Zaradi kompleksnosti ni dovolj, da to vlogo prevzame ena oseba. Potreben je namreč nabor osebja z dovolj širokim spektrom znanja in avtoritete, da bodo znali poiskati konsenz med željami posameznih zainteresiranih strani in sprejeti najboljšo odločitev tam, kjer konsenz ni dosegljiv. Literatura [3] navaja, da je na začetku projekta primerno izvesti dve do tri dnevno delavnico, na kateri se pod vodstvom delovne skupine sooči poslovne uporabnike in oddelek IT. Dnevni red uvodne delavnice naj bi vseboval naslednje teme [3]:

- Postavitev ciljev rešitve MDM in kako se ti cilji odražajo tako z vidika poslovnih uporabnikov, kot tudi s tehnološkega vidika. Poleg ciljev je potrebno določiti tudi kriterije uspešnosti projekta.
- Kreiranje projektne organizacijske strukture. Za posamezna področja je potrebno določiti odgovorne osebe ter predvideti dela, ki se bodo zaupala zunanjim izvajalcem. Če organizacija nima razpoložljivih vseh potrebnih znanj, je smiselno angažirati svetovalce ali zunanje izvajalce, ki imajo izkušnje na tem deficitarnem področju.
- Sprejeti odločitev o nakupu komercialne rešitve ali iti v lasten razvoj. Navadno organizacije že imajo nekaj znanja z nekaterih področij ali pa že posedujejo programska orodja za nekatere izmed tehnoloških rešitev, ki so del MDM. Na tem mestu je smiselno doreči, katere module bo organizacija razvila v lastni režiji in katere bo dokupila v obliki že obstoječih komercialnih rešitev. Tukaj je potrebna dodatna pozornost, saj se nekatere modularne rešitve do neke mere medsebojno prekrivajo in lahko prihaja do podvajanja funkcionalnosti. Kot primer lahko navedemo pakete za polnjenje DWH, ki navadno že vsebujejo funkcionalnosti za DQ.

- Določitev smernic razvoja (ang. *road map*). Postaviti je potrebno okvirni plan razvoja in postavitev (ang. *deploy*) posameznih komponent [15]. V plan se umesti mejnike (ang. *milestone*) in predpiše roke za njihovo doseganje. V tem koraku že mora biti znano, kateri moduli MDM in katere aplikacije bodo vzpostavljene v določeni stopnji vzpostavitve MDM.
- Kreiranje plana vpeljave (ang. *rollout plan*). Plan podaja načrt, s pomočjo katerega bomo sledili predhodno določenim smernicam razvoja [15]. Gre za stopenjski proces, ki vodi implementacijo od stopnje priprave preko preverjanja koncepta in večkratnih postavitev do končne stopnje operativnega MDM.

Kot enega od rezultatov nekaj dnevne delavnice delovna skupina pripravi visokonivojski stopenjski plan vpeljave MDM, katerega vzorec prikazuje tabela 7 [15]. Vsaka vrstica predstavlja eno od stopenj v planu, kjer so navedeni mejniki za področja podatkov, storitev, politik in integracije. Ko razvoj doseže vse navedene mejnike, je projekt pripravljen za prehod v naslednjo, višjo stopnjo.

**Tabela 7: Predloga plana vpeljave upravljanja matičnih podatkov.**

	Podatki	Storitve	Politike	Integracija
Priprava na uvedbo	Iskanje matičnih podatkov v silosnih aplikacijah. Izbira ali postavitev podatkovnega modela MDM.	Popis obstoječih servisov. Določitev nabora servisov MDM.	Pravila za upravljanje metapodatkov.	Profiliranje podatkov. Osnovna povezovalna pravila matičnih podatkov med aplikacijami.
Preverjanje koncepta	Preverjanje življenjskega cikla matičnih podatkov. Preverjanje kakovosti podatkov.	Implementacija in testiranje servisov za osnovne operacije MDM.	Predelava in preverjanje poslovnih pravil. Pravila za spremljanje kakovosti podatkov.	Dodelava in testiranje pravil za povezovanje matičnih podatkov. Vzpostavitev registra matičnih podatkov. Osveževanje podatkov v registru.
Prva namestitvev	Postavitev in polnjenje	Vzpostavitev servisne plasti,	Politike dostopa do podatkov za	Zagotavljanje osveževanja

	relacijskega modela MDM.	ki omogoča osnovne operacije nad matičnimi podatki v MDM.	uporabniške vloge.	podatkov MDM iz silosnih aplikacij.
Različica 1.0	Možnost urejanja podatkov vozlišča.	Dodelava silosnih aplikacij za uporabo servisov.	Razdelana pravila za obvladovanje podatkov.	Spremembe nad matičnimi podatki sistema MDM se propagirajo v podrejene aplikacije.
Tranzicija	Ukinjanje kopij v silosnih aplikacijah.	Dodajanje novih, kompleksnejših servisov, ki podpirajo poslovne procese.	Centralizacija in standardizacija poslovnih procesov.	Težnja k transakcijskemu načinu zagotavljanja koherence podatkov.
Operativni MDM	Podatki so na voljo samo še kot servis.	Programska oprema kot servis.	Politika upravljanja poslovnih procesov.	Izvedena integracija v vseh aplikacijah, ki uporabljajo podatke o strankah.

Priprava na uvedbo je namenjena analizi podatkov in pripravi pravil za določanje in uparjanje matičnih podatkov. Organizacija pregleda aplikacije in poslovne procese in izbere primeren arhitekturni način ter potencialne dobavitelje programske in strojne opreme.

V stopnji preverjanja konceptov se lahko preizkušajo različni pristopi in ponujeni programski paketi. Ker dosedanje delo nima neposrednega vpliva na poslovanje organizacije, saj ne spreminja delovanja obstoječih aplikacij, se organizacija lahko brez večjih težav odreče neuspehim in zgrešenim iniciativam. Podrobnejša opravila in naloge znotraj te stopnje se po potrebi lahko ponavljajo, dokler njihov rezultat ne doseže pričakovanj. Kot primer lahko navedemo pravila za združevanje zapisov matičnih podatkov. Pravila dopolnjujemo in jih preverjamo, dokler odstotek napačno združenih zapisov ne pade pod vnaprej določeno mejo. V tej stopnji že imamo zametek registrskega vozlišča MDM-CDI, ki služi kot poenoten nabor

strank in organizaciji nudi izdelavo natančnejših poročil o poslovanju in posledično vodstvu omogoča boljše usmerjanje poslovanja.

V stopnji prve namestitve MDM je podatkovni model vozlišča že v celoti izdelan. Spremembe nad matičnimi podatki se iz silosnih aplikacij prenašajo v vozlišče, zato je stanje vozlišča ažurno. Popravki zapisov neposredno v vozlišču MDM trenutno še niso mogoči, ker se spremembe iz MDM še ne prenašajo nazaj v aplikacije.

Različica 1.0 že prevzame vlogo edine verodostojne kopije matičnih podatkov. Vse obstoječe aplikacije se redno ali preko paketnih obdelav usklajujejo s stanjem MDM, nove aplikacije pa morajo uporabljati MDM kot edini vir matičnih podatkov. Vsi moduli sistema MDM so implementirani, politika obvladovanja podatkov pa se izvaja na nivoju celotne organizacije. Sedaj se sanacija vsake napačno sprejete odločitve v predhodnih treh stopnjah izkaže kot draga, saj je sistem MDM že vpet v poslovanje organizacije.

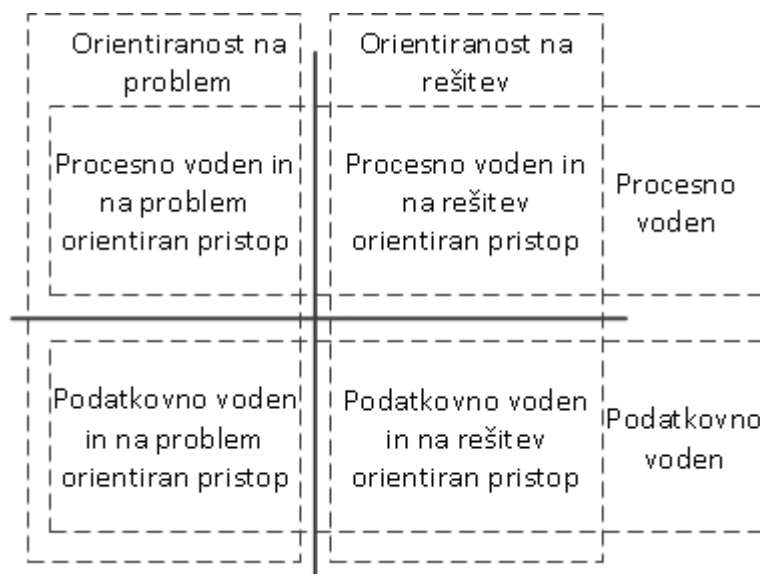
Kadar govorimo o tranziciji, imamo v mislih postopno ukinjanje lokalnih kopij matičnih podatkov v zalednih sistemih in postopno standardizacijo poslovnih procesov. Poti nazaj praktično ni več, ker je MDM globoko ukoreninjen v poslovanje.

V zadnji stopnji se organizacija znebi vseh nepotrebnih aplikacij, kot je na primer silosna aplikacija CRM, kar pomeni, da ima organizacija v jedru svojega IS transakcijski MDM-CDI. Opustitev aplikacije HRM je možna, ker je celotna funkcionalnost aplikacije že zajeta v samem vozlišču MDM-CDI in je ustrezno podprta s servisi. V tem primeru se lahko aplikacijo CRM realizira kot lahki odjemalec (ang. *thin client*), ki je enostavno dostopen preko spletnega brskalnika. Značilnost tankih odjemalcev je, da se izvedba vseh kompleksnih operacij izvaja v oddaljenem strežniku, v našem primeru je to sistem MDM, sam odjemalec pa služi zgolj kot uporabniški vmesnik za komunikacijo z uporabnikom.

## 3.2 Izbira metodologije

Seveda je vpeljava MDM velik integracijski projekt, ki zahteva ustrezno metodologijo [13] in s seboj prinaša precejšno stopnjo tveganja [1]. Izbira pravilne metodologije vpeljave projekta ima na nižanje tveganja velik pomen. Anne Cleven in Felix Wortmann [4] razčlenita štiri načine uvedbe projekta MDM, le-ti pa se medsebojno ločijo v dveh dimenzijah: glede na podatkovno (ang. *data-driven*) ali procesno vodenost (ang. *process-driven*) in glede na orientiranost na problem (ang. *problem-orientated*) oziroma rešitev (ang. *solution-orientated*). Značilnost podatkovno vodenih pristopov je, da za njimi navadno stojijo iniciative oddelkov

IT, medtem ko se pri procesno vodenih projektih zasleduje predvsem poslovne potrebe organizacije in so zato v domeni poslovnih uporabnikov. Problemska orientiranost je mišljena kot pristop od spodaj navzgor, kjer za izhodišče uporabimo obstoječe stanje IS in iščemo rešitve za konkretne težave z matičnimi podatki. Orientacija na rešitev je nasprotno mišljena kot pristop od zgoraj navzdol, ki je osredotočena na želeno končno ureditev matičnih podatkov v IS.



**Slika 19: Štirje pristopi vpeljave projekta upravljanja matičnih podatkov.**

Vsi štirje načini so prikazani na sliki 19 [4], povzemimo pa še njihove prednosti in slabosti:

- Podatkovno voden in na problem orientiran pristop

Glavna značilnost pristopa je, da na podlagi analiziranja obstoječih matičnih podatkov poizkušamo odpravljati napake, ki se v njih pojavljajo. Vodijo nas torej matični podatki, sam proces pa teži k rešitvi odkritih problemov. Analize se izvajajo ločeno v posameznih aplikacijah in ne zagotavljajo celostne rešitve.

Prednosti: Gre za postopke, ki ne zahtevajo veliko vloženega truda. Navadno gre za manjše posege v podatkovnih bazah in dodajanje enostavnejših omejitev na vnosne maske aplikacij. Rezultati so hitro vidni in zajamejo več medsebojno povezanih poslovnih procesov, ki določene podatke uporabljajo.

Slabosti: Proces ni sistematičen, ampak zgolj razreši nekatere zoznane probleme matičnih podatkov. Osredotočenost na odpravo najbolj perečih napak lahko zmanjša pozornost na preostale napake v matičnih podatkih. Ker je proces podatkovno voden, ni poudarka na vpliv kakovosti matičnih podatkov na poslovanje organizacije.

- Podatkovno voden in na rešitev orientiran pristop

Tukaj se opravi sistematična analiza matičnih podatkov, pri tem pa se zasleduje željeno končno stanje matičnih podatkov. Glede na to da smo orientirani na rešitev, imamo izdelane kriterije katerim morajo zadostiti matični podatki na nivoju celotne organizacije.

**Prednosti:** Orientiranost na rešitev predstavlja sistematičen pristop, ki glede na končno pričakovano stanje omogoča ugotovitev dejanskega stanja matičnih podatkov. Ureditev matičnih podatkov zajame več medsebojno povezanih poslovnih procesov, ki določene podatke uporabljajo.

**Slabosti:** Ker zasledujemo željeno končno rešitev, je postopek dolgotrajnejši in zahtevnejši, saj je potrebno pridobiti in poenotiti metapodatke o matičnih podatkih. Postopek je še vedno podatkovno voden, zato ne daje poudarka na vpliv kakovosti matičnih podatkov na poslovanje organizacije.

- Procesno voden in na problem orientiran pristop

V tem primeru kot izhodiščno točko za odpravo napak v matičnih podatkih vzamemo poslovne procese. Na ta način organizacija najprej izpopolni poslovne procese, kjer se dogajajo napake. Šele v drugem koraku se spusti na nivo matičnih podatkov, ki so v teh procesih udeleženi in jih ustrezno korigira.

**Prednosti:** Organizacija išče napake skozi poslovne procese, zato se krepi zavedanje o vplivu kakovosti matičnih podatkov na poslovanje.

**Slabosti:** Sam proces odkrivanja napak zopet ni sistematičen, saj se posvečamo zgolj tistim poslovnim procesom, kjer zaznamo napake. Kadar pristopamo s strani poslovnih procesov se lahko zgodi, da ne odkrijemo napak, ki izvirajo iz matičnih podatkov.

- Procesno voden in na rešitev orientiran pristop

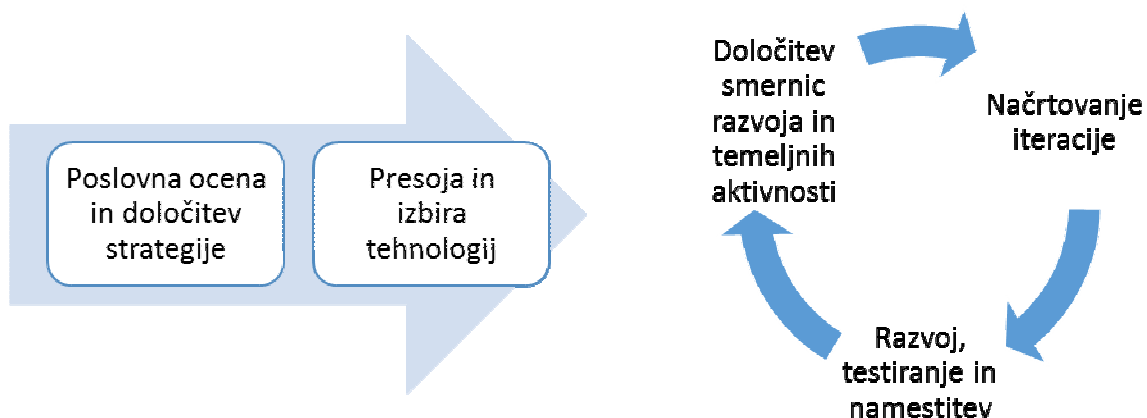
Tukaj zopet govorimo o sistematičnem pristopu, saj je pristop usmerjen proti iskanju rešitve in ne zgolj odpravljanju napak. Ker je procesno voden, izhodišče predstavljajo poslovna pravila. Tak način je primeren takrat, kadar se organizacija zaveda vrednosti kakovostnih matičnih podatkov in je pripravljena investirati veliko časa in sredstev. Na ta način vpeljava MDM ni le še ena iniciativa ampak projekt, ki bo podprl in izboljšal poslovanje.

**Prednosti:** Zagotovljen je poudarek kakovosti matičnih podatkov na poslovanje ter sistematičnost. Projekt ima podporo vodstva in je zato visoko prioritetan.

**Slabosti:** Količina dela, ki ga je potrebno opraviti, je zagotovo večja kot pri preostalih pristopih. Lahko se zgodi, da se zaradi osredotočenosti na poslovna pravila spregleda katero od napak, ki izvira neposredno iz matičnih podatkov.

Zgoraj opisani pristopi se medsebojno seveda ne izključujejo, saj jih je možno medsebojno kombinirati. Kot primer navedimo inicialno vpeljavo MDM v organizacijo. Takrat je smiselno popisati vsa poslovna pravila in pripraviti vizijo zelenega končnega stanja. Za ta namen je procesno voden in na rešitev orientiran pristop verjetno najprimernejši. Ko organizacija že ima operativen MDM, je za morebitne popravke in nadgradnje verjetno bolj smiselno izbrati katerega izmed preostalih treh.

Za učinkovito vpeljavo MDM potrebujemo iterativno metodologijo, saj moramo v doglednem času pripraviti delujoče sklope IS. Iterativnost hkrati zagotavlja boljši nadzor in hitrejši prehod posameznih delov v redno uporabo. Ko govorimo o stopenjskem pristopu uvajanja MDM-CDI, se postavi vprašanje, kako dolge naj bodo te stopnje. Le-te ne smejo biti prekratke, saj je znotraj ene stopnje smiselno imeti vsebinsko zaključene celote. Pri časovno predolgih pa se lahko zgodi, da izgubimo orientacijo na zastavljene cilje, vodstvo pa začne postajati nestrpno, saj kljub velikim stroškom nimajo v rokah nikakršnih oprijemljivih rezultatov. Zato, da razvoj uspe pripraviti oprijemljive rezultate, ki so na voljo v doglednem času, se priporoča, da implementacija ene stopnje ne presega šest do osem mesecev [3]. Razvoj mora v tem času pripraviti rešitev, ki ni uporabna zgolj v eni organizacijski enoti, ampak njene doprinose lahko koristijo vsi oddelki. Pri uvedbi MDM je smotrno, da je metodologija procesno vodena in orientirana na rešitev. Ena takšnih metodologij, ki zajema vse naštet, je Metoda za integrirano okolje znanja (ang. *Method for an Integrated Knowledge Environment* – MIKE2.0). Je odprtokodna metodologija, pri kateri gre za obsežen nabor rešitev za upravljanje informacij. Med njimi pa se nahaja tudi modul MDM. Gre za agilni pristop, ki sestoji iz 5 faz, prikazanih na sliki 20 [32].



**Slika 20: Metoda za integrirano okolje znanja.**

V prvi fazi se izvede poslovno oceno in določi strategijo, druga faza pa je namenjena presoji in izbiri pravilne tehnologije. Obe omenjeni fazi skupaj poimenujemo kot strateški načrt (ang. *strategic blueprint*) in še nista iterativni, saj se izvedeta samo na začetku projekta. Strateški

načrt je nekakšna predstavitev trenutnega stanja ter želenega končnega stanja, opredeljene pa so tudi vmesne stopnje preko katerih želimo doseči končno stanje. V teh dveh fazah so tudi že izbrani ponudniki programske in strojne opreme.

S tretjo fazo se prične iterativni pristop, v katerem si sledijo: najprej faza določitev smernic razvoja ter temeljnih aktivnosti, nato faza načrtovanja iteracije in kot zadnja faza razvoja, testiranja in namestitvev. Smernice razvoja izhajajo neposredno iz strateškega načrta in se pripravljajo sproti za vsako implementacijsko fazo. Gre za preslikavo vizije strateškega načrta v različne implementacijske iteracije. V tej stopnji se določi glavne komponente in njihovo

umestitev v celoten sistem, vodi pa se tudi dokumentacija o stanju projekta, katera je vedno na voljo sponzorjem projekta. Faza načrtovanja poskrbi za načrtovanje arhitekture, podatkovno ter procesno modeliranje vseh funkcionalnosti zajetih v trenutni iteraciji. V skupek razvoja, testiranja in namestitve seveda sodijo vse aktivnosti usmerjene v realizacijo zastavljene vizije, kar vključuje realizacijo programske kode in dokumentacije, testiranje novih funkcionalnosti ter njihovo vključevanje v IS organizacije.

Sama organizacija dela v zadnjih treh fazah omogoča, da se projektno ekipo iteracije razdeli v več skupin in se jih grupira okoli različnih implementacijskih vlog [32]. Na ta način se skupine, ki se periodično srečujejo s podobnimi zahtevami, izpopolnjujejo. Tako iteracije postajajo hitrejše in izdelki bolj kakovostni.

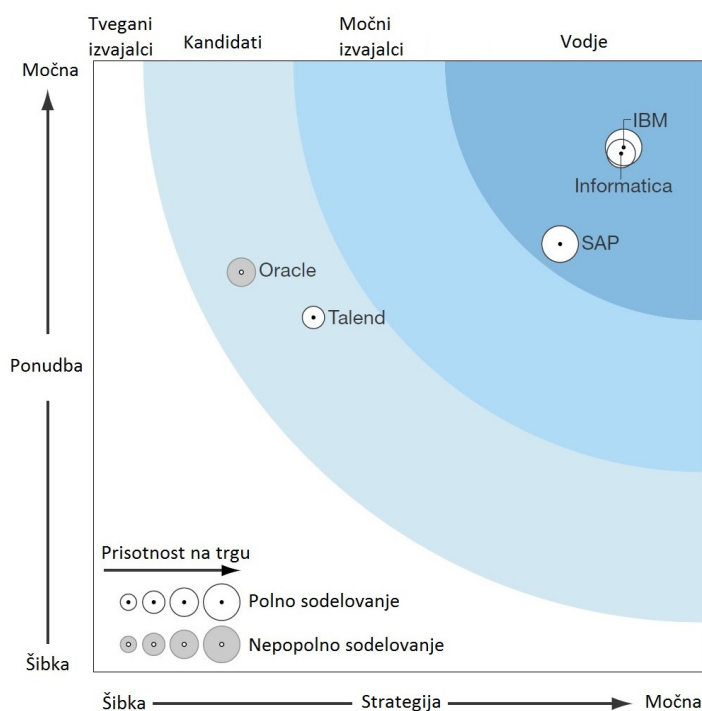


## 4 Pregled ponudbe na trgu

Ko se organizacija odloča za vpeljavo MDM v svoje poslovanje, se navadno znajde pred dilemo, katero rešitev izbrati v množici ponudnikov na trgu. Z nadgradnjo IS in poenotenjem matičnih podatkov želimo pokriti trenutne potrebe, potrebno pa je imeti v mislih tudi prihodnji razvoj. Ne smemo si privoščiti, da bi se z napačno izbiro ponudnika znašli v situaciji, ko bi sistem MDM oviral rast in nadaljnji razvoj organizacije. Ker je vsaka implementacija zgodba zase, je potrebno še posebej skrbno preveriti karakteristike posameznih komercialnih paketov na podlagi nekaj glavnih poslovnih procesov organizacije. Na tem mestu zato ne moremo podati jasne usmeritve, kateri izmed ponudnikov je najboljša izbira, so pa nam lahko v pomoč priporočila različnih analitskih hiš, ki periodično spremljajo dogajanje na segmentu trga MDM-CDI. Na ta način organizacija lažje izbere nekaj ponudnikov, ki jih uvrsti v ožji izbor in v nadaljnjih testiranjih izbere najustreznejšo rešitev. Poleg tehničnih lastnosti posameznih produktov je zelo pomembna tudi vizija in stabilnost ponudnika, ki bo organizaciji nudil tudi ustrezno podporo. Smiselno je, da organizacija že v prvem krogu izbere nabor ponudnikov, ki zadoščajo minimalnim potrebam organizacije. V nadaljevanju si oglejmo kateri so po mnenju analitskih hiš vodilni ponudniki MDM-CDI.

### 4.1 Poročili tržnih raziskav

Podjetje za raziskovanje trgov Forrester je v prvem četrtletnem poročilu za leto 2014 izvedlo



Slika 21: Forrester diagram za upravljanje matičnih podatkov.

oceno petih glavnih ponudnikov rešitev MDM-CDI. Na diagramu, ki ga prikazuje slika 21 [10], so ponudniki razdeljeni v tri dimenzije. Po eni osi diagrama se jih razvršča po obsegu ponudbe njihovih paketov. Tukaj se nahajajo kriteriji kot so: podpora obvladovanju podatkov, integracijske opcije, modeliranje, uparjanje, performance itd. Druga os predstavlja strategijo, pod katero se ocenjujeta strategiji podjetja in produkta s poudarkom na tem, kako ponudnik sledi potrebam svojih kupcev. Tretja delitev bazira na prisotnosti ponudnika na trgu. Slednja je pridobljena iz informacij o: finančnem stanju podjetja, številu novih strank itd.

Kot je razvidno, so se med vodje uvrstili ponudniki IBM, Informatica in SAP, medtem ko sta Talend in Oracle pristala v segmentu kandidatov. Potrebno je poudariti, da je ocena za Oracle pridobljena na nepopolnih podatkih, saj slednji v raziskavi ni želel sodelovati [10].



**Slika 22: Gartnerjev diagram za upravljanje matičnih podatkov.**

Bolj obsežna je raziskava, ki jo je oktobra 2013 opravila analitska hiša Gartner [16]. Pri tej raziskavi se prodajalce rešitev MDM deli v dve dimenziji, prikazani na sliki 22. Prva je sposobnost izvedbe, kjer analitiki ocenjujejo: zmožnost in kakovost temeljnih funkcionalnosti ponujenih rešitev, finančno stabilnost podjetja in vlaganja v nadaljnji razvoj produkta, uspešnost prodaje, hitrost prilagajanja zahtevam strank, zadovoljstvo strank itd. Druga dimenzija je celovitost vizije, ki vključuje: razumevanje potreb strank, marketinške strategije, prodajne strategije, produktne strategije, inovacije itd.

Tako dobljeni prostor je nadalje razdeljen v štiri kvadrante. Ponudba nižnih igralcev je omejena zgolj na nekatere funkcionalnosti in imajo nizko stopnjo inovativnosti ter implementacijske sposobnosti. Kot že ime pove, ima skupina vizionarjev potencial za vplivanje na smer nadaljnjega razvoja celotnega segmenta MDM-CDI, so pa implementacijsko šibkejši, saj svojih vizij navadno nimajo dokončno implementiranih ali pa rešitve še niso dovolj uveljavljene. Izzivalci imajo visoko sposobnost zadovoljiti potrebe svojih strank, saj poznajo vse uveljavljene koncepte MDM, ne uspeva pa jim slediti trendom v industriji, zato se niso prebili med vodje. Vodje veljajo za zanesljive ponudnike, ki so sposobni izpolniti želje svojih strank, njihove rešitve pa sledijo trendom sodobnega poslovanja.

Po mnenju podjetja Gartner sodijo v skupino vodij IBM, Informatica, Oracle in Tibco. Vse ostale ponudnike so uvrstili v skupino nižnih igralcev.

Čeprav sta bili raziskavi opravljeni nad programskimi paketi enakih različic, se njuni izsledki nekoliko razlikujejo. Obe analizi med vodje uvrščata IBM in Informatico, medtem ko se SAP in Tibco med vodji pojavljata zgolj v eni raziskavi. Kot smo že omenili, je Oracle posredoval podatke zgolj podjetju Gartner, zato oceno Forresterja za tega ponudnika lahko smatramo kot nepopolno. Vzroke za nekoliko različne rezultate gre najverjetneje iskati v različnih metodologijah in izboru kandidatov, zato si bomo poglobljeje ogledali vse kandidate kategorizirane kot vodje.

## 4.2 Lastnosti vodilnih ponudnikov

V nadaljevanju so navedene lastnosti tistih ponudnikov in njihovih glavnih MDM-CDI rešitev, ki so kot vodje predstavljeni vsaj v eni izmed omenjenih analiz.

Ponudnik: IBM

Produkt: IBM InfoSphere Master Data Management, verzija 11

- Prednosti: Gre za paket, ki zajema širok spekter strategij upravljanja informacij (ang. *information management*) in velikih podatkov (ang. *big data*) ter je primeren predvsem za velike organizacije [10]. Ponuja zelo dobro integracijske zmožnosti in performance, podjetje pa kaže pri napredni izdaji (ang. *advanced edition*) poznavanje različnih industrij in dobro podporo vodenju podatkov [16].
- Slabosti: Gartner opozarja na zmanjševanje rasti prihodkov iz naslova programske opreme MDM, ki je v leti 2012 znašala zgolj 3,5 odstotka [16]. Standardna izdaja (ang. *standard edition*) je omejena zgolj na registrski implementacijski način in se kot

podpovprečno izkaže pri sledenju uporabniškim potrebam in poročanju. Izbira standardne izdaje bi lahko pomenila resne ovire morebitnim nadgradnjam IS organizacije. Nekateri kupci tudi izpostavljajo zmedo, ki je nastala med implementacijo rešitve [10], kar bi lahko pomenilo oteženo poslovanje organizacije.

Ponudnik: Informatica

Produkt: Informatica Master Data Management, verzija 9.6

- Prednosti: Poleg tega, da uporabniki izpostavljajo izjemno fleksibilnost podatkovnega modela, paket vsebuje tudi dobra orodja za integracijo in zagotavljanje kakovosti podatkov [10, 16]. Različica 9.6 ponuja rešitev MDM, ki je neodvisna ob podatkovne baze, saj za namene uparjanja zapisov uporablja abstrahiran dostop do podatkovne baze [27]. Informatica je v letih 2012 in 2013 veliko vlagala v razvoj in nakupe tehnologij s področja MDM in upravljanja poslovnih procesov, kar daje jasen znak, da ima podjetje resne načrte še okrepiti svoj položaj na trgu. V letu 2012 je podjetje zabeležilo 7 odstotno rast prihodkov na področju upravljanja podatkov o strankah [16].
- Slabosti: Še nepopolna integracija kupljenih tehnologij in rešitev je razlog za previdnost, saj je od uspeha le-te precej odvisna nadaljnja smer razvoja rešitev. Programski paket je nekoliko podhranjen na področju tehnologij za vodenje podatkov [16], vsebuje pa nekatera lastna orodja, za katera skrbniki potrebujejo dodatna usposabljanja [10].

Ponudnik: Oracle

Produkt: Oracle Universal Customer Master, verzija 8.2

- Prednosti: Ponudnik ima širok spekter rešitev MDM in 10 odstotno rast prometa v letu 2012 [16]. Strankam, ki uporabljajo to rešitev, zelo ustrezata podpora različnim implementacijskim načinom in vizija razvoja, ki ponuja rešitve za njihove potrebe. Vozlišče omogoča obdelavo velikega števila podatkov in razširljivost (ang. *scalability*) [16].
- Slabosti: Množica istovrstnih rešitev, ki jih ponuja Oracle, lahko povzroča zmedo pri izbiri prave alternative [10]. Takšen produkt je Fusion MDM, ki še ne dosega občutnejšega tržnega deleža, vendar v izbiro vnaša dodatno kompleksnost in sproža dvome o tem, katera rešitev bo v prihodnje predstavljala glavni produkt in bo kot tak deležen nadaljnjega razvoja. Nekatero organizacije, ki so se odločile za Oracle rešitev, navajajo, da je bila za uspešno implementacijo nujna velika podpora ponudnika [16], kar lahko pomeni nekoliko kompleksnejšo prilagoditev potrebam stranke.

Ponudnik: Tibco

Produkt: Tibco's MDM, verzija 8.3

- Prednosti: Tibco je v letu 2012 zabeležil 25 odstotno rast prihodkov, kar mu daje močan zagon za nadaljnjo širitev tržnega deleža [16]. Ponudnik gradi ekipo za namene prihodnjih implementacij MDM, kar mu daje še dodatno kredibilnost v svojih namenih. Trenutna rešitev ponuja dobro podporo spremljanju kakovosti in obvladovanju podatkov, stranke pa ponudniku dajejo visoke ocene tudi na drugih področjih.
- Slabosti: Filozofija in sama implementacija je še preveč usmerjena v vidike MDM s stališča oddelkov IT, ne posveča pa dovolj pozornosti poslovnim vidikom. Po navedbah analitske hiše Gartner bo Tibco v pripravo rešitev primoran angažirati dodatne poslovne uporabnike. Tudi zaradi velike rasti se ponudnik sooča s pomanjkanjem izkušenih kadrov, kar utegne ogroziti kakovost implementacij in nadaljnjega razvoja.

Ponudnik: SAP

Produkt: Master Data Governance for Customer, verzija 6.1

- Prednosti: SAP ponuja širok nabor rešitev za centralizacijo, konsolidacijo in vodenje podatkov. Orodje se kaže kot fleksibilno, saj omogoča tako upravljanje podatkov SAP ERP, kot tudi razširitve podatkovnega modela za podatke drugih virov [16]. Pozitivna je tudi usmerjenost na vidik poslovne uporabe, saj je orodje prilagojeno tako, da daje kontrolo nad matičnimi podatki v roke poslovnim uporabnikom [10].
- Slabosti: Produkt je neprimeren za konsolidacijski način implementacije, zato se morajo potencialni interesenti odločiti za dodatna paketa SAP Data Services in NetWeaver MDM, ki pa zgublja položaj na trgu [16]. Gartner tudi izpostavlja verjetnost zahtevnejšega implementacijskega cikla, ki bi utegnil odvrnati potencialne kupce.

Oborožena s temi informacijami se organizacija lažje odloči za tisto različico, ki bolje pokriva njene sedanje potrebe in ponuja vizijo razvoja v prihodnosti. Uvedbo MDM v poslovanje ne smemo gledati samo iz perspektive pokrivanja trenutnih potreb organizacije, ampak je potrebno gledati dolgoročno. Vsaka širitev poslovanja ali nabora produktov in storitev se namreč odraža s povečanjem bremena, ki ga mora MDM kot osrednji člen IS organizacije uspešno prenašati.



## 5 Sklep

V magistrski nalogi so prikazani vsi štirje načini implementacije sistema MDM-CDI. Čeprav so prehodi med posameznimi načini možni, je izjemnega pomena, da se organizacija odloči za pravi način implementacije.

Organizacija se bo odločila za registrski način, kadar za njeno poslovanje zadostuje operativna raba matičnih podatkov. To pomeni, da ne obstaja potreba za izdelavo resnejših poročil za namene poslovnega obveščanja. Rezultati poizvedb nad registrom so sicer ažurni, a poizvedbe lahko trajajo predolgo za normalno delo. Tak register je primeren samo za branje podatkov, zato se morebitno urejanje matičnih podatkov še vedno dogaja v silosnih aplikacijah. Registrski način implementacije je po drugi strani idealna rešitev za razvoj pravil za ujemanje zapisov matičnih podatkov. Ker je izvedba registra enostavna in brez poseganja v obstoječe aplikacije, je uvedba poceni, hitra in brez tveganj za tekoče poslovanje. Tovrsten način je primeren za začetek razvoja MDM rešitve, ki se nadalje prelevi v enega izmed preostalih treh načinov implementacije.

Spoznali smo, da je usklajevalni način zelo podoben registrskemu. Ta se izkaže kot prava izbira, kadar organizacija za svoje poslovanje ne potrebuje ažurnih podatkov, BI pa predstavlja podlago za usmerjanje bodočega poslovanja. Ker pri tem načinu MDM vozlišče hrani konkretne vrednosti atributov in ne zgolj referenc na zapise virov, je izvajanje zahtevnih poizvedb bistveno hitrejše kot pri registrskemu načinu. Pomembno je dejstvo, da tudi pri vzpostavitvi usklajevalnega MDM ni nikakršnega poseganja v obstoječe aplikacije. Tako vozlišče je zato cenovno ugodno in za IS organizacije ne predstavlja nikakršnega tveganja. Tudi tukaj nimamo možnosti urejanja matičnih podatkov direktno preko MDM.

V trenutku, ko organizacija koristi tako operativno, kot tudi analitično metodo uporabe in za poslovanje potrebuje ažurne podatke, se pojavi potreba po transakcijskem vozlišču. Ravno tako je transakcijsko vozlišče predpogoj za urejanje matičnih podatkov v samem sistemu MDM. Tukaj se celotno upravljanje z matičnimi podatki prenese iz silosnih aplikacij v novo nastalo vozlišče, kar pomeni kompleksnejšo implementacijo in korenite spremembe v silosnih aplikacijah. Ker implementacija posega tudi v obstoječe poslovne procese, uvedba lahko resno ovira ali celo ogrozi poslovanje organizacije.

Vozlišče MDM, na podlagi kombiniranega načina implementacije, se na spektru implementacijskih načinov nahaja med usklajevalnim in transakcijskim. Gre za kombinacijo obeh omenjenih načinov. Organizacija se zanj odloči, kadar nekaterih silosnih aplikacij ni mogoče hitro prilagoditi za delo s transakcijskim vozliščem. Če organizacije ne moti nekoliko nižja stopnja ažurnosti podatkov in delno podvajanje, potem je to lahko dobra vmesna rešitev v procesu uvajanja transakcijskega vozlišča. Seveda je to lahko tudi končna rešitev, ki ohranja trenutno stanje nekaterih aplikacij in omogoča visok nivo stopnje zrelosti upravljanja matičnih podatkov [23]. Vse nove aplikacije se namreč priklapljujejo na servisno plast, ki je neposredno povezana z vozliščem MDM.

Izbira načina implementacije je, poleg vizije razvoja organizacije, močno povezana z lastnostmi silosnih aplikacij, ki trenutno hranijo matične podatke. Predlagali smo, da se pripravi vizualizacija na podlagi zvezdnih diagramov, ki prikazuje stopnje pomembnosti kriterijev kot so: število atributov, stopnja kompleksnosti usklajevanja, pomembnost sinhronizacije, stopnja kontrole dostopa, kompleksnost servisov in zmogljivost [15]. Tako se organizacija lažje odloči, kateri izmed načinov implementacije je zanjo najprimernejši in ali so za doseg izbrane stopnje zrelosti upravljanja morebiti potrebne vmesne implementacije.

Vpeljava projekta MDM je izredno zahtevna in z dolgoročnimi posledicami za organizacijo [3]. Kot ustrezno metodologijo vodenja in implementacije MDM smo prepoznali MIKE 2.0, ki je agilna, procesno vodena in orientirana na rešitev [4, 32]. Kot rezultat vsake iteracije se pričakuje nova delujoča funkcionalnost v sistemu MDM, iteracija pa ne sme trajati dlje kot šest do osem mesecev [3]. V nasprotnem primeru se lahko zgodi, da implementacija izgubi orientacijo na rešitev, vodstvo pa zaradi dolgotrajnega razvoja in višine stroškov izgublja motivacijo za nadaljevanje projekta. Ker za implementacijo potrebujemo kadre z različnimi kompetencami, je smiselno, da organizacija, namesto učenja na lastnih napakah, pridobi zunanje izvajalce ali svetovalce s področja MDM [15].

Ker je sistem MDM zasnovan modularno, se organizacija lahko odloči, da bo nekatere module ali pa celotno rešitev poiskala na trgu. Izbira ustreznega ponudnika temelji na trenutnih potrebah poslovanja in viziji nadaljnjega razvoja organizacije. Prav zaradi tega je pomembno, da izbrani ponudnik ponuja dolgoročno sodelovanje in razvoj produkta v smeri potreb uporabnikov. Iz rezultatov analiz dveh analitskih hiš Forrester in Gartner smo izluščili ponudnike, ki predstavljajo dolgoročno stabilne partnerje z vizijo razvoja, ki sledi potrebam svojih strank [10, 16]. To so programski paketi podjetij: IBM, Informatica, Oracle, Tibco in SAP, za katere smo tudi predstavili prednosti in slabosti njihovih produktov. V tem delu ne moremo, kot najboljšega, izpostaviti samo enega od navedenih ponudnikov. Ker je vsaka

implementacija z godba zase, je izbira najustrežnejšega ponudnika naloga vsake organizacije posebej. Glede na specifične svojega poslovanja in želene rešitve, je potrebno natančno preučiti, ali so podprte vse zahtevane funkcionalnosti in kako je nova programska oprema kompatibilna z morebitnimi obstoječimi moduli.

To delo podaja veliko informacij, ki zainteresirani organizaciji osvetlijo področje upravljanja matičnih podatkov. V primeru, da se organizacija odloči za nadgradnjo svojega IS s sistemom MDM-CDI, so na voljo usmeritve za izbiro načina implementacije in kaj vse je potrebno opraviti na organizacijski ravni, da projekt pripeljemo do realizacije.



## 6 Priloge

Celoten izračun primerjav za primer uparjanja zapisov. Navedene so primerjave zapisov od A do D pri katerih je za vsak izmed atributov izvedena primerjava z zapisi 1 do 3. Na kocu primerjave vsakega atributa je podana ocena ujemanja po tem atributi, ocene pa so nato seštete v razdelku vsot, ki predstavljajo ocene za primerjavo celotnega zapisa.

Primerjava zapisa A					
Atribut: ime in priimek					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	7,00	11,00	0,36	0,5	0,18
2	5,00	11,00	0,55	0,5	0,27
3	10,00	12,00	0,17	0,5	0,08

Atribut: datum rojstva					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	3,00	8,00	0,63	0,40	0,252
2	1,00	8,00	0,85	0,40	0,34
3	4,00	8,00	0,50	0,40	0,2

VSOTA
0,43
<b>0,61</b>
0,28

Primerjava zapisa B					
Atribut: ime in priimek					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	10,00	11,00	0,09	0,5	0,05
2	10,00	11,00	0,09	0,5	0,05
3	7,00	12,00	0,42	0,5	0,21

Atribut: datum rojstva					

id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	4,00	8,00	0,50	0,40	0,2
2	5,00	8,00	0,38	0,40	0,15
3	0,00	8,00	1,00	0,40	0,4

Atribut: naslov					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	25,00	23,00	-0,09	0,50	-0,04
2	18,00	17,00	-0,06	0,50	-0,03
3	0,00	23,00	1,00	0,50	0,50

Atribut: telefon					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	6,00	9,00	0,33	0,30	0,10
2	9,00	9,00	0,00	0,30	0,00
3	0,00	9,00	1,00	0,30	0,30

VSOTA
0,30
0,17
<b>1,41</b>

Primerjava zapisa C					
Atribut: ime in priimek					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	20,00	11,00	-0,82	0,5	-0,41
2	21,00	11,00	-0,91	0,5	-0,45
3	17,00	12,00	-0,42	0,5	-0,21

Atribut: naslov					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	21,00	23,00	0,09	0,50	0,04
2	22,00	17,00	-0,29	0,50	-0,15
3	24,00	23,00	-0,04	0,50	-0,02

Atribut: elektronski					
----------------------	--	--	--	--	--

naslov					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	4,00	11,00	0,64	0,70	0,45
2					
3	2,00	11,00	0,82	0,70	0,57

VSOTA
0,08
-0,60
<b>0,34</b>

Primerjava zapisa D					
Atribut: EMŠO					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	1,00	13,00		1,00	1,00
2	0,00	13,00		1,00	0,00
3	0,00	13,00		1,00	0,00

Atribut: telefon					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	7,00	9,00	0,22	0,30	0,07
2	9,00	9,00	0,00	0,30	0,00
3	7,00	9,00	0,22	0,30	0,07

Atribut: elektronski naslov					
id	število menjav	dolžina	(d-menjave)/d	utež	ocena
1	0,00	11,00	1,00	0,70	0,70
2					
3	6,00	11,00	0,45	0,70	0,32

VSOTA
<b>1,77</b>
0,00
0,38



## 7 Literatura

- [1] C. Asakiewicz, »Business investments in IT: Managing integration risks«, IT Professional, julij/avgust 2011, str. 41 – 45.
- [2] M. Bayer, »Master Data Management - A Systematic Approach«, DM Review, vol 16, no 10, str. 32-35, 2006
- [3] A. Berson, L. Dubov, »Master Data Management and Customer Data Integration for Global Enterprise«, McGraw-Hill, 2007.
- [4] A. Cleven, F. Wortmann, »Uncovering four strategies to approach master data management«, v zborniku 43rd Hawaii International Conference on System Sciences, 2010, str. 1 – 10.
- [5] J. Dyché, E. Levy, »Customer Data Integration«, John Wiley & Sons, 2006.
- [6] A. Dreibelbis, I. Milman, »Enterprise Master Data Management: An SOA Approach to Managing Core Information«, IBM Press, 2008.
- [7] L. P. English, »Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits«, Wiley, 1999
- [8] L. Garber, »Michael Stonebraker on the Importance of Data Integration«, IT Pro May/June, 1999
- [9] T. Gilb, »Principles of Software Engineering Management, Pearson Higher Education, 1988
- [10] M. Goetz, »The Forrester Wave: Master Data Management Solutions, Q1 2014«, februar 2014, Forrester
- [11] P. Herceg, D. Madison, »Multichannel Customer Contact Management«, IT Professional, maj/junij 2004, str. 33 – 40.
- [12] R. Kimball, M. Ross, »The Data Warehouse Toolkit, The complete guide to dimensional modeling«, Wiley Inc. 2002.
- [13] W. Lam, V. Shankararaman, »An Enterprise Integration Methodology«, IT Professional, marec/april 2004, str. 40 – 48.

- [14] R. Land, I. Crnković, »Oh Dear, We Bought Our Competitor: Integrating Similar Software Systems«, IEEE Software, marec/april 2011, str. 75 – 82.
- [15] D. Loshin, »Master Data Management«, Morgan Kaufmann, 2009.
- [16] B. O'Kane, S. Judah, »Magic Quadrant for Master Data Management of Customer Data Solutions«, Gartner, oktober 2013.
- [17] N. Rowe, »The state of Master Data Management«, Aberdeen Group, maj 2012
- [18] S. Suram, R. Muppala, »Master Data Management-CDI«, v zborniku IEEE Region 10 Conference, 2008, str. 1267 – 1272.
- [19] L. Wang, X. Ming, J. You, »The Steps and Methodology of Identifying Master Data from Business Processes«, v zborniku World Congress on Software Engineering, 2009, str. 329 – 333.
- [20] H. J. Watson, B.H. Wixom, »The current state of Business Intelligence«, IEEE Computer, September 2007, str. 96 – 99.

#### **Internetni viri:**

- [21] C. Bradley, »Master data management workshop«, <http://www.slideshare.net/inforacer/how-to-identify-the-correct-master-data-subject-areas-tooling-for-your-mdm-initiative1>, junij 2014
- [22] P. Chandramohan, »De-duplication Part 1&2«, <http://www.mdmgeek.com/category/matching-algorithm/>, junij 2014
- [23] »The Data Governance Maturity Model«, DataFlux, 2007, [www.accelacomm.com/jlp/microsite/0/80321541/](http://www.accelacomm.com/jlp/microsite/0/80321541/)
- [24] W. W. Eckerson 2002. »Data Quality and the Bottom Line, Achieving business success through a commitment to high quality data.«, TDWI, [download.101com.com/pub/tdwi/Files/DQReport.pdf](http://download.101com.com/pub/tdwi/Files/DQReport.pdf), junij 2014
- [25] Gartner, »Gartner Says Master Data Management Software Revenue to Grow 21 Percent in 2012«, <http://www.gartner.com/newsroom/id/1886314>, januar 2012
- [26] IBM, »Continuing the MDM journey«, <http://public.dhe.ibm.com/common/ssi/ecm/en/imw14603usen/IMW14603USEN.PDF>

- [27] Informatica, »Informatica MDM 9.6 Powers Universal Master Data Management«, junij 2013, <http://www.informatica.com/us/company/news-and-events-calendar/press-releases/06052013-mdm.aspx#fbid=zy-JBFtY7od>
- [28] J. Parkinson, »Time to Clean Up Your Master Data«, CFO.com, maj 2014, <http://ww2.cfo.com/it-value/2014/03/time-clean-master-data/>
- [29] V. Lovicz, »Beyond matching – enterprise golden record management«, [marketing.profitsee.com/acton/attachment/2329/f-015e/1/-/-/-/file.pdf](http://marketing.profitsee.com/acton/attachment/2329/f-015e/1/-/-/-/file.pdf)
- [30] W. McKnight »MDM & CDI ROI and Justification«, Conversion Services International, april 2007, <http://www.csiwhq.com>
- [31] »Master Data Management definition - What is MDM-CDI?«, april 2014, <http://www.executionmih.com/master-data-management/MDM-concept.php>
- [32] »MIKE2.0«, avgust 2014, <http://mike2.openmethodology.org/>
- [33] »W3C Working Group Note«, februar 2004, <http://www.w3.org/TR/ws-arch/#soadefs>