

UNIVERZA V LJUBLJANI
EKONOMSKA FAKULTETA

MAGISTRSKO DELO

**PODATKOVNO MODELIRANJE, PROCESI IN POSLOVNA
PRAVILA ZA ZAGOTAVLJANJE KAKOVOSTI PODATKOV**

Ljubljana, julij 2023

JANJA MALOVRH

IZJAVA O AVTORSTVU

Podpisana Janja Malovrh, študentka Ekonomske fakultete Univerze v Ljubljani, avtorica predloženega dela z naslovom Podatkovno modeliranje, procesi in poslovna pravila za zagotavljanje kakovosti podatkov, pripravljene v sodelovanju s svetovalcem red. prof. dr. Jurijem Jakličem

IZJAVLJAM

1. da sem predloženo delo pripravila samostojno;
2. da je tiskana oblika predloženega dela istovetna njegovi elektronski obliki;
3. da je besedilo predloženega dela jezikovno korektno in tehnično pripravljeno v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani, kar pomeni, da sem poskrbela, da so dela in mnenja drugih avtorjev oziroma avtoric, ki jih uporabljam oziroma navajam v besedilu, citirana oziroma povzeta v skladu z Navodili za izdelavo zaključnih nalog Ekonomske fakultete Univerze v Ljubljani;
4. da se zavedam, da je plagiatstvo – predstavljanje tujih del (v pisni ali grafični obliki) kot mojih lastnih – kaznivo po Kazenskem zakoniku Republike Slovenije;
5. da se zavedam posledic, ki bi jih na osnovi predloženega dela dokazano plagiatstvo lahko predstavljalo za moj status na Ekonomski fakulteti Univerze v Ljubljani v skladu z relevantnim pravilnikom;
6. da sem pridobila vsa potrebna dovoljenja za uporabo podatkov in avtorskih del v predloženem delu in jih v njem jasno označila;
7. da sem pri pripravi predloženega dela ravnala v skladu z etičnimi načeli in, kjer je to potrebno, za raziskavo pridobila soglasje etične komisije;
8. da soglašam, da se elektronska oblika predloženega dela uporabi za preverjanje podobnosti vsebine z drugimi deli s programsko opremo za preverjanje podobnosti vsebine, ki je povezana s študijskim informacijskim sistemom članice;
9. da na Univerzo v Ljubljani neodplačno, neizključno, prostorsko in časovno neomejeno prenašam pravico shranitve predloženega dela v elektronski obliki, pravico reproduciranja ter pravico dajanja predloženega dela na voljo javnosti na svetovnem spletu preko Repozitorija Univerze v Ljubljani;
10. da hkrati z objavo predloženega dela dovoljujem objavo svojih osebnih podatkov, ki so navedeni v njem in v tej izjavi;
11. da sem preveril verodostojnost informacij, ki izhajajo iz zapisov na podlagi uporabe orodij umetne inteligence.

V Ljubljani, dne _____

Podpis študentke: _____

KAZALO

UVOD	1
1 PREGLED LITERATURE	3
1.1 Poslovna pravila.....	3
1.1.1 Teorije poslovnih pravil	4
1.1.2 Smernice določanja poslovnih pravil	6
1.1.3 Razvrščanje poslovnih pravil.....	6
1.1.4 Odločitvene tabele in drevesa.....	9
1.2 Procesi.....	11
1.2.1 Lastnosti poslovnih procesov	11
1.2.2 Modeliranje poslovnih procesov	12
1.2.3 Združitev poslovnih pravil in procesov	15
1.3 Podatkovno modeliranje	17
1.3.1 Lastnosti podatkovnega modeliranja	18
1.3.2 Konceptualni model.....	19
1.3.3 Logični model.....	22
1.4 Podatkovna baza	24
1.4.1 Relacijska baza	24
1.4.2 Normalizacija baze	28
1.5 Kakovost podatkov	29
1.5.1 Teorije kakovosti podatkov	30
1.5.2 Dimenzije kakovosti podatkov statističnih uradov.....	32
1.5.3 Vpliv in izboljšanje kakovosti podatkov	34
1.6 Sinteza spoznanj literature	37
2 OVREDNOTENJE OBSTOJEČE REŠITVE ANP.....	39
2.1 Metodologija zbiranja	39
2.2 Analiza stanja z vidika poslovnih pravil, procesov in podatkovnega modeliranja	40
2.2.1 Analiza z vidika poslovnih pravil.....	40
2.2.2 Analiza z vidika procesov.....	41
2.2.3 Analiza z vidika podatkovne baze	43
2.2.4 Ugotovitve na podlagi analize	43

3	PREDLOG NOVE REŠITVE ANP	44
3.1	Določitev poslovnih pravil.....	44
3.2	Določitev procesov	48
3.3	Podatkovno modeliranje	52
3.4	Ovrednotenje nove rešitve	58
3.4.1	Metodologija	58
3.4.2	Ovrednotenje z vidika prednosti in slabosti nove rešitve.....	59
3.4.3	Ovrednotenje glede na dimenzije kakovosti podatkov	61
	SKLEP.....	62
	LITERATURA IN VIRI.....	64
	PRILOGE	69

KAZALO TABEL

Tabela 1:	Primeri simbolov BPMN po skupinah	14
Tabela 2:	Kardinalnost in opis povezave entitetnih tipov	21
Tabela 3:	Razvrstitev problemov, ki vplivajo na kakovost podatkov	30
Tabela 4:	Odločitvena tabela eno pravilo v eni vrstici.....	47
Tabela 5:	Odločitvena tabela stil križišč	48
Tabela 6:	Primerki povezav med poslovnimi pravili in procesi.....	51
Tabela 7:	Seznam entitetnih tipov za zbiranje podatkov in njihov pomen	54
Tabela 8:	Seznam entitetnih tipov za prevzem surovih podatkov in njihov pomen	55
Tabela 9:	Seznam entitetnih tipov za obdelavo podatkov in njihov pomen.....	55

KAZALO SLIK

Slika 1:	Razvrščanje poslovnih pravil po avtorjih Wan-Kadir in Loucopoulos	7
Slika 2:	Razvrščanje poslovnih pravil po avtorju Rossu.....	8
Slika 3:	Sestavine poslovnega procesa.....	12
Slika 4:	Model za integracijo poslovnih pravil in procesov	17
Slika 5:	Oblikovanje podatkovne baze	19
Slika 6:	Delitev povezave mnogo proti mnogo z vmesno relacijo.....	23
Slika 7:	Okvir za delitev dimenzij kakovosti podatkov	31
Slika 8:	Meja med točno oceno podatka in napačnim podatkom.....	33
Slika 9:	Model dejstev nove rešitve	46
Slika 10:	Del glavnega procesnega modela BPMN nove rešitve.....	50
Slika 11:	Primer ER modela za pripravo adresarja	52

Slika 12: Primer relacijskega modela za pripravo adresarja..... 57

KAZALO PRILOG

Priloga 1: Procesni model urada.....	1
Priloga 2: Model podprocesov nove rešitve za pripravo na zbiranje	2
Priloga 3: Model podprocesov nove rešitve za obdelavo podatkov	3
Priloga 4: ER Model nove rešitve	4
Priloga 5: Relacijski model nove rešitve za zbiranje podatkov.....	5
Priloga 6: Relacijski model nove rešitve za obdelavo podatkov	6

SEZNAM KRATIC

angl. – angleško

ANP – (angl. Labour Force Survey); Aktivno in nekativno prebivalstvo

BPM – (angl. Business Process Management); Management poslovnih procesov

BPMN – (angl. Business Process Modeling Notation); Notacija poslovno procesnega modela

CRM – (angl. Customer Relationship Management); Management odnosov s strankami

CRUD – (angl. Create Retrieve Update Delete); Kreiraj, vrni, popravi in izbriši

DBMS – (angl. Data Base Management System); Sistem za upravljanje zbirk podatkov

DDL – (angl. Data Definition Language); Jezik za definiranje podatkov

DML – (angl. Data Manipulation Language); Jezik za upravljanje s podatki

ER – (angl. Entity Relationship Model); Model entitet povezav

ESS – (angl. European Statistical System); Evropski statistični sistem

IT – (angl. Information Technology); Informacijska tehnologija

OECD – (angl. Organisation for Economic Co-operation and Development); Organizacije za gospodarsko sodelovanje in razvoj

SQL – (angl. Structured Query Language); Strukturirani povpraševani jezik

UVOD

Namen raziskave aktivnega in neaktivnega prebivalstva (angl. Labour Force Survey, v nadaljevanju ANP) je prikaz dogajanja aktivnosti prebivalstva na slovenskem trgu. Med pomembnimi rezultati raziskave sta stopnja brezposelnosti in stopnja delovne aktivnosti. Zbiranje je predpisano s strani Evropske unije in kot uradni izvajalec državne statistike je Statistični urad Republike Slovenije (v nadaljevanju urad) dolžan zagotavljati te podatke. Zbiranje podatkov poteka četrtletno. Enote v vzorcu so izbrane osebe in njihovi družinski člani. Izbrane osebe so v vzorcu eno leto in pol, z vmesnim premorom, kar dodatno otežuje celotno zbiranje podatkov. Podatki se zbirajo tako preko terenskega kot telefonskega anketiranja. Pri ponovljenem anketiranju, se določeni odgovori prevzamejo iz predhodnega anketiranja. Pri obdelavi podatkov se uporabljajo tudi administrativni viri (Rutar & Tomažič, 2020).

Urad je glavni dajalec statističnih podatkov v državi. Kot član evropskega statističnega sistema imenovanega European Statistical System (v nadaljevanju ESS), pri svojem delu upošteva načela kodeksa ravnanje evropske statistike. Kodeks zajema načela in kazalnike, ki služijo za nadzor okolja, procesov in rezultatov in tako prispevajo h kakovosti. Kakovost je predstavljena iz dimenzij ustreznosti, točnosti ocen, pravočasnosti in točnosti objave, dostopnosti in jasnosti, skladnosti in primerljivosti. V kodeksu je tako tudi načelo, da so statistični postopki ustrezni in načelo, da se urad zavezuje h kakovosti, zato nenehno izboljšuje kakovost procesov in storitev. Urad si mora prizadevati, da obstoječe rešitve stalno spremlja, nadzoruje in po potrebi izboljšuje ter pri tem upošteva smernice za zagotavljanje kakovosti in tako prispeva h konkurenčni prednosti (Statistični urad Republike Slovenije, 2017).

Glede na omenjeni načeli ima obstoječa rešitev zbiranja in obdelave podatkov ANP veliko slabosti. Procesi se v večini primerov izvajajo ročno. Za celotno obdelavo je potrebno veliko človeških virov in pri tem lahko kaj hitro pride do napake pri zajemu in obdelavi podatkov. Pri sami obdelavi prihaja do podvajanja aktivnosti. Z vključitvijo več oseb v proces obdelave je večje tveganje za neuspešno zaključitev vseh faz obdelave. Tako obstaja tveganje zaradi možnosti odsotnosti ključnih oseb, težave zaradi medsebojnega usklajevanja in možnosti človeške napake zaradi ročnega izvajanja postopkov. Celotna obdelava je posledično tudi časovno potratna. Zajeti in obdelani podatki so shranjeni na datotečnem strežniku. Vsaka časovna serija ima svoje tabele, kar otežuje dostop in obdelavo mikropodatkov. Obstoječa rešitev ne omogoča revizijske sledi dostopov in obdelave podatkov. V primeru izgube oz. napak v sami obdelavi nimamo na voljo informacij, kdo oz. kateri proces je odgovoren za nastalo situacijo.

Neprestane spremembe na trgu vplivajo na to, da morajo biti informacijske rešitve prilagodljive in spremenljive. Pri tem ima pomembno vlogo določitev poslovnih pravil (Bajec & Krisper, 2005). Ustrezna poslovna pravila so tako vez med poslovnim in infrastrukturnim področjem. Naš primer je odraz tega, da si pri definiranju poslovnih aktivnosti vzamemo premalo časa za razmislek, torej se ne razmisli o pravih stvareh na pravem mestu ob pravem času. Tako se skozi razvoj rešitve kažejo težave, ki bi morale biti predvidene že pred začetkom razvoja (Ross, 2003a), kar ima posledično vpliv na kakovost rešitve.

Prav tako nas primer opozarja na pomembnost podatkovnega modeliranja in shranjevanja podatkov v primerno okolje. Podatkovno modeliranje je oblikovanje vsebine tabel in povezav med tabelami informacijske rešitve. Predstavlja majhen delež celotnega razvoja rešitve, a vendar je ključnega pomena, da se tudi tej fazi nameni dovolj časa za razmislek, kako bodo podatki shranjeni in organizirani v podatkovnem okolju in tako prispevamo k bolj kakovostni rešitvi (Simsion & Witt, 2005, str. 8–11). Med najbolj razširjenimi podatkovnimi bazami je relacijski podatkovni sistem, ki ima prednost tudi v tem, da zagotavlja visoko stopnjo neodvisnosti podatkov. To pomeni, da ob manjših spremembah z minimalnimi posegi vzdržujemo informacijsko rešitev. Zaradi omenjenih lastnosti je, v primerjavi z datotečni strežniki, za shranjevanje podatkov, bolj primerna podatkovna baza (Teorey, Lightstone, Nadeau & Jagadish, 2011, str. 34).

Negativni vpliv na kakovost podatkovne baze imajo dejavniki kot so prazne, nepravilne vrednosti ali nizi, napačna interpretacija podatka v bazi, kršitve enoličnosti zapisa v tabeli, dvojniki v tabeli glede na vse attribute v bazi oz. del ključnih atributov, napačna sintaksa, itd. (Oliveira, Rodrigues & Henriques, 2005, str. 13). Tako je dokazal tudi Godnov (2012, str. 165), da podatkovni model vpliva na kakovost podatkov in sicer od tega kako so postavljeni glavni in tuji ključi, kakšnega podatkovnega tipa so atributi, kakšne so deklarativne omejitve atributov in normalizacija podatkovnega modela, in vse to je povezano s poslovnimi pravili.

Drugače povedano, lahko predvidevamo, da neustrezno opredeljena pravila, procesi, podatkovno modeliranje rezultirajo v slabši kakovosti podatkov, kot bi bila potrebna oz. pričakovana.

Namen magistrskega dela je tako z upoštevanjem pomena opredelitve pravil, procesov in podatkovnih modelov prispevati k razvoju take rešitve, ki bo zagotavljala učinkovito in uspešno zbiranje in obdelavo podatkov in tako prispevala h kakovosti rešitve in s tem k uspešni digitalizaciji.

Raziskovalno vprašanje magistrskega dela je, kako so medsebojno povezani vidiki poslovnih pravil, procesov in podatkovnega modeliranja ter kako ti vidiki vplivajo na kakovost podatkov.

Glede na podano problematiko in namen magistrskega dela so cilji sledeči:

- prikazati značilnosti poslovnih pravil, procesov in podatkovnega modeliranja,
- opredeliti pojem kakovost podatkov,
- podati ključne ugotovitve kako so področja poslovnih pravil, procesov in podatkovnega modeliranja med sabo povezana in kakšen vpliv ima vsako od njih na kakovost podatkov,
- kakšno vlogo ima pri tem podatkovno okolje,
- analizirati obstoječo rešitev zbiranja in obdelave podatkov raziskave ANP glede na omenjena področja ter ugotoviti pomanjkljivosti rešitve,
- podati predlog nove rešitve,
- podati ključne ugotovitve, ki so vodilo za uspešno in učinkovitejšo informacijsko rešitev.

Uporabljene metode dela v teoretičnem delu slonijo na analitičnem pregledu strokovnih in znanstvenih del tuje in domače literature. Uporabljene so deskriptivne metode dela in primerjalna analiza. V empiričnem delu bodo uporabljene metode dela na podlagi analitičnega pregleda interne dokumentacije obstoječe rešitve. Opravljeni so tudi intervjuji z zaposlenimi na uradu. Povezava teoretičnega dela z empiričnem je podano na podlagi logičnega sklepanja glede na pridobljeno znanje in izkušnje v okviru študija in zaposlitve.

V teoretičnem delu iz različnih vidikov opredelimo teorijo poslovnih pravil, procesov in podatkovnega modeliranja. Opišemo lastnosti podatkovne baze, osredotočeno na relacijsko podatkovno okolje. Opredelimo dimenzije kakovosti podatkov, ki so v uporabi na uradu in jih vsebinsko podkrepimo še z dimenzijami kakovosti podatkov drugih avtorjev. Povežemo teorije poslovnih procesov, pravil in podatkovnega modeliranja s teorijo o kakovosti in pri tem uporabimo logično sklepanje. V empiričnem delu naloge obstoječo rešitev ovrednotimo na podlagi teoretičnih spoznanj iz prvega poglavja, opravljenih intervjujev in analize obstoječe tehnične in uporabniške dokumentacije. V zadnjem poglavju naloge predlagamo novo rešitev s postavitvijo poslovnih pravil, procesov in podatkovnim modeliranjem. Rešitev je predlagana glede na teoretična spoznanja v prvem delu. Na koncu je rešitev ovrednotena na podlagi opravljenih intervjujev in logičnega sklepanja.

1 PREGLED LITERATURE

Poslovna pravila določajo obnašanje, delovanje in odločanje v podjetju. Poslovni procesi določajo obnašanje sistema, da zagotovi glavno dejavnost podjetja. Pri tem se upoštevajo poslovna pravila. Pogosto zasledimo literaturo, ki opisuje povezavo med poslovnimi pravili in procesi ter poslovnimi pravili in podatki. Vsak proces spremljajo podatki in metapodatki, ki jih je potrebno shraniti v varovano okolje. Pogosto uporabljeno primerno okolje je podatkovna baza. Kakovost podatkov ima vpliv na uspešnost poslovanja podjetja, torej velja, da morajo biti podatki shranjeni v podatkovni bazi, čim bolj kakovostni.

1.1 Poslovna pravila

Pravila so del našega vsakdanjega življenja doma, na delovnem mestu, v šoli, na prireditvah, itd., ki jih moramo spoštovati in se po njih ravnati, saj tako prispevamo k bolj kakovostnemu življenju in lažjemu funkcioniranju v družbi. Namen upoštevanja pravil je, da nas zaščitijo pred nevarnostmi in usmerjajo vedenje na željeni način. Če pravila upoštevamo, smo nagrajeni, v primeru kršitve pa sledijo ukrepi, kot so opozorila in kazni. Velikost nagrade oz. kazni je odvisna od tega, kako pomembno je povezano pravilo. Podobno velja za poslovna pravila, le da ta veljajo v poslovnem okolju.

Poslovna pravila določajo obnašanje, delovanje in odločanje zaposlenih glede na politiko delovanja, zakonodajo in predpise, ki veljajo v podjetju. Delimo jih na poslovna pravila, ki so del politike delovanja podjetja, aktivnostna pravila, ki določajo obnašanje posameznih aktivnosti

poslovnega procesa in strukturalna pravila, ki določajo pravila za izvajanje logike v delovnih procesih. Pravila temeljijo na podlagi znanja in izkušenj znotraj oddelka, kot tudi v povezavi z zunanjim okoljem. S spremljanjem dogajanja v poslovanju podjetja pridobimo nove izkušnje in nova znanja, kar lahko pripomore k izboljšanju poslovanja podjetja (Kovačič & Bosilj Vukšič, 2005, str. 100–101).

Ross (2003a) pravi, da so poslovna pravila direktiva, ki vplivajo ali usmerjajo delovanje poslovnih procesov na željeni način in jih ščitijo pred nevarnostmi in grožnjami. Torej poslovna pravila so del procesno organiziranega podjetja. Le-ta postavlja v ospredje rešitve temelječe na poslovnih procesih imenovano Management poslovnih procesov (angl. Business Process Management, v nadaljevanju BPM). Lahko rečemo, da so poslovna pravila temelj v procesno usmerjenem delovanju, ki jih informacijska rešitev mora upoštevati. Kompleksnost in spreminjanje poslovnih pravil vpliva na to, kakšna bo informacijska rešitev in kako pogosto jo bo potrebno prilagajati.

Sistem, kjer so tudi v ospredju poslovna pravila, je Management odnosov s strankami (angl. Customer Relationship Management, v nadaljevanju CRM). V CRM so v ospredju rešitve, ki so v čim večji meri prilagojene željam, potrebam strank in upravljanju odnosov med strankami. Prepoznati moramo čim več različnih vrst strank in se truditi, da dobičkonosne obdržimo, poiščemo strategijo za nedobičkonosne in pridobimo nove stranke. Rešitve strmijo k temu, da so čim bolj prilagodljive, hitro odzivne, ekonomične, da delamo vse samo enkrat, pravilno in v čim bolj zgodnji fazi. V uspešnih CRM sistemih so temelj učinkovita poslovna pravila, na podlagi katerih se izvedejo vsi poslovni procesi rešitve (Kumar & Reinartz, 2012, str. 3–15; Ross, 2003b, str. 13–17).

1.1.1 Teorije poslovnih pravil

Osnovna načela poslovnih pravil po avtorju Rossu (2003a, 2003b, str. 85–86) so sledeča:

- Potrebno je preučiti, ali je dovolj dodane vrednosti, da se pravila uporabijo.
- Pravila naj bodo neodvisna od delovnih tokov (angl. Workflow) in postopkov (angl. Procedures).
- Pravila naj bodo zgrajena na podlagi dejstev (angl. *Facts*), dejstva pa na podlagi terminov (angl. *Terms*).
- Pravila naj določi oseba z največ znanja o vsebini iz poslovnega okolja, na katerega se pravilo nanaša.
- Pravila morajo imeti lastnost adaptabilnosti, da jih je možno nadgraditi in spreminjati.
- Vsako poslovno pravilo nekaj stane. Največji strošek pri spremembah je čas za uvajanje nove rešitve pri uporabnikih.
- Bolje manj kakovostnih pravil, kot več slabih.

Poslovna pravila določimo na podlagi poslovnih ciljev, predpisov, politike delovanja, problematike in poslovnih procesov podjetja (Bajec & Krisper, 2005).

Pogosto se poslovna pravila zamenjuje s sistemskimi pravili. Poslovna pravila se nanašajo izključno na poslovanje podjetja in igrajo glavno vlogo v pravilni izvedbi poslovnih procesov. Sistemska pravila pa omogočijo, da se poslovna pravila izvedejo. Kako se le-ta izvedejo, je seveda odvisno tudi od uporabljene informacijske tehnologije (Lam, 2020). Poglejmo si primer poslovnega in sistema pravila: »Izbrana oseba stara 18 let in več mora odgovoriti na vprašanja o zaposlitvi.« je poslovno pravilo, »Če je izbrana oseba stara 18 let in več, naj se prikaže sklop vprašanj o zaposlitvi.« je sistemsko pravilo.

Poslovna pravila lahko primerjamo z delovanjem človeškega telesa. Človeško telo je sestavljeno iz skeleta, mišic in živčevja. Skelet je ogrodje telesa, mišičice telesu dajejo moč, živčevje pa skrbi za pravilno delovanje telesa. Skelet je sestavljen iz kosti, kar v poslovnem svetu predstavlja osnovne koncepte poslovne rešitve imenovane termini. Vezi med kostmi lahko primerjamo z dejstvi, ki se nanašajo na termine in jih med sabo povezujejo. Mišice lahko primerjamo s poslovnimi procesi, živčevje pa s poslovnimi pravili, ki nadzorujejo sistem in skrbijo za pravilno delovanje procesov. Pri zgradbi telesa je ločeno delovanje mišic in živčevja. Podobno morajo biti poslovni procesi ločeni od poslovnih pravil (Ross, 2003b, str. 43–50).

Poglejmo si definicije pomembnejših izrazov, ki jih srečujemo v poslovnih pravilih (Ross, 2003b, str. 53–66):

- Termin je sestavljen iz besede ali besedne zveze in je običajno samostalnik, ki predstavlja najmanjšo skupino iz poslovnega okolja. Primeri terminov so: telefonski anketar, naročilo, spletna anketa ali status. Vsak termin ima točno določeno definicijo. Katalog konceptov je zbirka terminov in definicij. Na ta način si zagotovimo, da vsi deležniki sistema enako razumejo vsak termin.
- Dejstvo je sestavljeno iz glagola in povezuje termine med seboj. Dejstvo ne predstavlja omejitve nad terminom, ampak zgolj osnovo, na podlagi katere določimo pravilo.
- Model dejstev je grafični prikaz terminov in dejstev ter povezav med njimi. Omogoča enostavnejše definiranje pravil. Model je tudi osnova za povezave in organizacijo podatkov v podatkovni bazi. Vsak termin ali dejstvo naj v modelu nastopi samo enkrat in tako zagotovi konsistenčnost. Predstavlja eno izmed težjih področji modeliranja, saj je to osnovni načrt znanja o poslovanju.
- Dogodek je, ko se nekaj zgodi in je pričakovano, da se na ta dogodek odzovemo. Dogodek je tudi odziv na uporabljeno pravilo.
- Pravilo je sestavljeno iz terminov in dejstev. Pravilo nadzoruje potek dogajanja, torej določa, kako se na nek dogodek odzovemo. Pravilo lahko določa zavrnitev, sprožitev druge aktivnosti, opozorilo, obvestilo, itd. Eno pravilo je lahko uporabljeno pri večih dogodkih. V primeru kršitve pravila se kot napaka izpiše besedilo uporabljenega pravila.

1.1.2 Smernice določanja poslovnih pravil

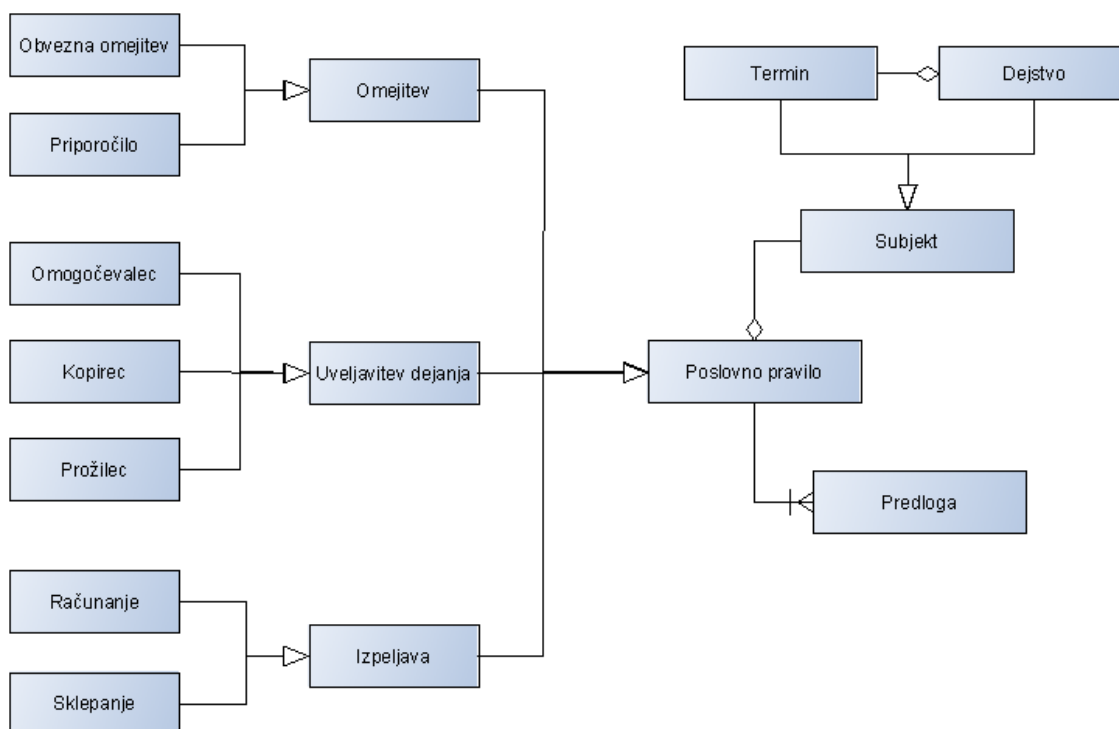
Določanje poslovnih pravil zahteva dobro poznavanje in razumevanje poslovnega sistema. Obstajajo priročniki za določanje poslovnih pravil, katerih namen je, da so pravila čim bolj jasna, nedvomljiva in enako razumljiva vsem vpletenim - poslovnim uporabnikom, analitikom ali razvijalcem. Poglejmo si nekaj pravil avtorja Rossa (2009, str. 23), čigar smernice lahko splošno uporabimo ne glede na govorni jezik:

- Pravilo naj ne vsebuje logike točnih korakov za procesno izvajanje. Takega pravila ni mogoče ponovno uporabiti.
- Pravilo mora biti jasno napisano in razumljivo vsakomur.
- Pravilo mora odgovoriti le na vprašanje, kaj je pravilo. Kako, kje, kdaj, kdo in zakaj bo pravilo uveljavljeno, ne sme biti vključeno v odgovor.
- Pravila morajo vsebovati določene besede, določenim pa se je potrebno ogniti. V pravilu naj bodo vsebovane besede, kot so »mora«, »samo«, izogibamo pa se besedi »lahko«. Izogibamo se odvečnim frazam ali besedam, ki nimajo vpliva na pomen pravila.
- Uporabljen naj bo model dejstev, ki vsebuje termine in dejstva in si zagotovimo, da ne pride do dvoumnosti pravila.
- Pazimo, da so v pravilu vključena vsa dejstva.
- Izogibamo se pravilu, ki se začne z besedo »če«.
- Samostalnik oz. samostalniška zveza, ki določa termin, mora biti v ednini.
- Pravilo, ki vsebuje števno dejstvo, mora pripadati števnemu predmetu (terminu).
- Izogibamo se vezniku »ali«, s katerim opredelimo več možnih izbir. Veznik nadomestimo z besedno zvezo kot je npr. »ko je izpolnjeno vsaj eno izmed naštetega«.
- Izogibajmo se gnezdenju računskih operacij. Pravilo razdelimo na več enostavnejših, med seboj neodvisnih pravil, kjer je vsaka formula vsebovana v svojem pravilu. Na ta način lahko vsako matematično formulo uporabimo večkrat.
- Pravilo naj vsebuje natanko eno omejitev.
- Pravilo mora biti ločeno od procesov.
- V pravilih ne uporabljamo sistemskih ukazov, kot so kreiraj, vrni, popravi in izbriši (angl. Create Retrieve Update Delete, v nadaljevanju CRUD).
- Izogibamo se pravilu, kjer je uporabljena beseda »ko«, ki predstavlja časovno omejitev, ko se nek dogodek zgodi.

1.1.3 Razvrščanje poslovnih pravil

Obstaja več teorij razvrščanja poslovnih pravil, kar nam omogoča bolj poglobljeno razumevanje, lažje interpretiranje in definiranje kakovostnejših pravil. Poglejmo si eno izmed teorij prikazano na sliki 1.

Slika 1: Razvrščanje poslovnih pravil po avtorjih Wan-Kadir in Loucopoulos



Prerejeno po Wan-Kadir & Loucopoulos (2004).

Kot prikazuje slika, je poslovno pravilo sestavljeno iz termina in dejstva. Termin je beseda oz. fraza iz poslovnega okolja. Dejstvo je povezava med dvema terminoma in ga običajno zapišemo kot <subjekt> <povezava> <subjekt>. Iz tega sledi, da lahko uporabimo skupni izraz subjekt za termin kot tudi za dejstvo.

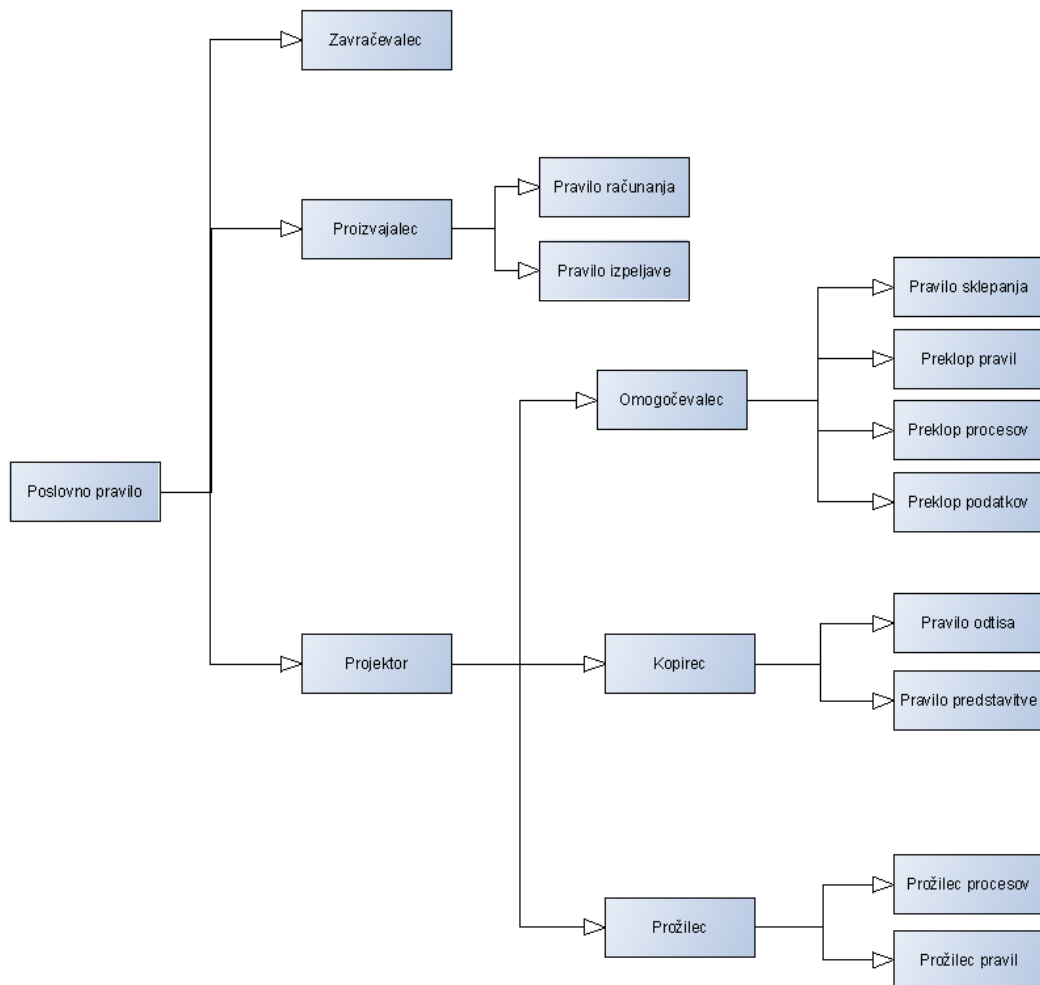
Poslovno pravilo lahko razporedimo v naslednje skupine (Wan-Kadir & Loucopoulos, 2004):

- Omejitev (angl. Constraint). Deli se na obvezno omejitev in priporočilo. Naloga omejitve je, da se preveri dogodek nad predmetom in v primeru kršitve omejitve se ta ne izvede, v primeru kršitve priporočila, pa se izvede opozorilo.
- Uveljavitev dejanja (angl. Action Assertion). Deli se na omogočevalec, kopirec in prožilec. Naloga te skupine je, da ob nekem dogodku ali pod določenimi pogoji sproži dejanje, ki pogosto spremeni stanje subjekta. Naloga omogočevalca je, da omogoči ali onemogoči, da se pravilo oz. nek proces zgodi. Skupina kopirec obstoječe vrednosti kopira in jih ponovno uporabi ob drugem dogodku. Prožilec ob nekem dogodku oz. pogoju sproži neko dejanje.
- Izpeljava (angl. Derivation). Naloga skupine je, da na podlagi izpeljave določi novo dejstvo. Delimo jo na skupino računanja in sklepanja. Skupina računanja na podlagi matematičnih operacij izračuna novo vrednost. Skupina sklepanja pa določi novo vrednost na podlagi logičnega sklepanja.

Vsako skupino pravil lahko zapišemo na podlagi vsaj ene predloge. Npr. predloga za obvezne omejitve je <subjekt> mora [ne sme] <dejstvo> [če velja pogoj] (Wan-Kadir & Loucopoulos, 2004).

Poglejmo še teorijo razvrščanja poslovnih pravil po avtorju Rossu (2003b, str. 142–143) na sliki 2, ki vsebuje podobno delitev kot pri teoriji na sliki 1.

Slika 2: Razvrščanje poslovnih pravil po avtorju Rossu



Prirajeno po Ross (2003b, str. 142–143).

Glede na vrsto kontrole lahko vsako poslovno pravilo razvrstimo v eno izmed treh glavnih skupin (Ross, 2003b, str.141–143):

- Zavračevalec (angl. Rejector) je skupina pravil, ki zavrnejo dogodek v primeru kršitve pravila. Skupina je ekvivalentna skupini omejitev prikazani na sliki 1.
- Proizvajalec (angl. Producer) je skupina pravil, ki izračuna vrednost na podlagi matematičnih operacij ali logičnega sklepanja in je ekvivalentna skupini izpeljave na sliki 1.

- Projektor (angl. Projector) je skupina pravil, ki pravilo vedno sprejme in proizvede nek dogodek. Skupina je ekvivalentna skupini uveljavitev dejanja in se deli na enake podskupine, kot je prikazano na sliki 1. Razlika je le v tem, da se vsaka od teh skupin deli še na nižji nivo:
 - a) Omogočevalec na podlagi sklepanja določi da je nekaj resnično, omogoči oz. onemogoči pravilo, operacijo, proces ali proceduro. V eno izmed podskupin se uvršča tudi skupina pravil, ki kreira ali briše podatke pod določenimi okoliščinami.
 - b) Kopirec ustvari kopijo vrednosti in jo uporabi v drugačnih okoliščinah ali pa vrednost prikaže na različne načine.
 - c) Prožilec ustvari aktivnost, ki proži izvedbo operacije, procesa, procedure ali pravila.

1.1.4 Odločitvene tabele in drevesa

Pri pisanju pravil so nam v pomoč tudi metode, ki nam pravila vizualno poenostavijo, da odpravimo čim več nedoslendnosti in dvoumnosti že v tej fazi. Za pisanja pravil, kjer je pravilo odvisno od več pogojev, je simselno uporabiti tabelarni način. Pogoje zapišemo v obliki tabeliranja, pred tem pa jasno opredelimo koliko od teh pogojev mora veljati (Ross, 2013b).

Poglejmo si primer: Osebo anketiramo, če velja natanko vse od naštetega:

- oseba je stara 16 let ali več,
- oseba je član gospodinjstva izbrane osebe.

Možen zapis pravil pa je tudi v obliki tabel ali dreves, kar imenujemo odločitvene tabele oz. odločitvena drevesa. Takšen prikaz pravil je smiselno uporabiti v primeru, ko se več pravil nanaša na isti subjekt in če je le možno, se ga uporabi pred tabelarnim načinom (Ross, 2013b).

Odločitvene tabele prikazujejo odločitvena pravila, ki podajo odgovor na selektivno vprašanje. To pomeni, da pravila vsebujejo isti termin z več možnimi vrednostmi. Vsaka vrednost določa rezultat, ki ga ne moremo izraziti s formulo. Vizualno predstavitev pravil omogoča enostavnejše in bolj pravilno razumevanje. Posledično jih lažje interpretiramo in v primeru sprememb enostavno nadgradimo. Vrstni red izvajanja poslovnih pravil v odločitvenih tabelah ne sme biti pomemben, saj mora biti logika odločanja deklarativna.

Vizualno je tabela sestavljena iz vrstic in stolpcev. Osnovni elementi odločitvenih tabel so:

- eden ali več dejavnikov ali atributov,
- vsak dejavnik ima vsaj dva možna primerka,
- v primeru vsaj dveh dejavnikov imamo presečne vrednosti,
- rezultati, ki so podani v celicah.

Poznamo več stilov odločitvenih tabel. Pogledali si bomo najenostavnejši stil križišč (angl. Intersection Style) in stil eno pravilo v vrstici (angl. One-Rule-Per_Row Style).

Stil križišč pomeni, da so dejavniki prikazani v matrični obliki. V primeru, ko imamo dva dejavnika, predpostavimo, da ima prvi dejavnik m primerkov, drugi dejavnik pa n primerkov. Odločitvena tabela bo v tem primeru velikosti $m \times n$, kjer so primerki prvega dejavnika predstavljeni v m vrsticah, primerki drugega dejavnika pa v n stolpcih. Celice, kjer se vrstice in stolpci križajo, so rezultati pravil in se imenujejo odločitvene celice. Vsaka celica je rezultat natanko enega pravila. Prednost križišč je v tem, da je enostavno zagotoviti, da ne prihaja do podvajanja primerkov, da zagotovimo vse kombinacije primerkov, torej zagotovimo popolnost poslovnih pravil in vsaka kombinacija primerkov vsebuje rezultat. Takšna tabela bo brez praznih vrednosti. V primeru sprememb pravil, je tabela enostavno prilagodljiva. Zaradi omenjenih lastnosti je stil priporočljiv za uporabo v poslovnem svetu. V primeru večjega števila dejavnikov je stil težko obvladljiv. Priporočilo je, da se uporabi za največ pet dejavnikov (Ross, 2013a).

Stil eno pravilo v vrstici predstavlja vsako pravilo v svoji vrstici. To pomeni, da se z leve strani tabele v vsakem stolpcu nahaja posamezen dejavnik in njegovi primerki. Na desni strani tabele v zadnjem stolpcu se nahajajo odločitvene celice. V primeru dveh dejavnikov imamo tabelo, ki vsebuje tri stolpce. V prvih dveh stolpcih z desne strani sta dejavnika in njuni primerki po vrsticah, v zadnjem stolpcu pa rezultati pravil. Tabele ne vsebujejo križišč, zato pa se primerki dejavnikov, v primeru vsaj dveh dejavnikov, podvajajo, da zagotovimo vse kombinacije dejavnikov. Omenjeni stil ima slabost v tem, da je težko opaziti, če je katera od kombinacij primerkov izpuščena in posledično pridemo do praznih vrednosti pravil. Težko opazno je tudi podvajanje primerkov dejavnikov. Prednost omenjenega stila je, da je tabelo enostavno določiti kljub večjemu številu dejavnikov. Stil je bližji razvijalcem. Tabela, ki vsebujejo samo en dejavnik, predstavlja enostavna pravila, brez križišč in je vizualno vedno enaka, ne glede na uporabljen stil (Ross, 2013a).

Odločitveno drevo je vizualna predstava pravil v obliki vozlišč, vej in listov. Drevo gradimo od zgoraj navzdol. Atributi so vozlišča drevesa. Glavno vozlišče je koren drevesa. Iz vsakega atributa je največ toliko vej, kot je različnih vrednosti atributov, ki gredo v notranjost. Vozlišče, ki ima samo eno vrednost atributov, se imenuje list drevesa. Vozlišče, iz katerega potuje veja v poddrejeto, se imenuje starš, vozlišče v katero se veja steka, je otrok drevesa. Binarna drevesa so v obliki, kjer iz vsakega vozlišča potujeta natanko dve veji, torej samo dve množici atributov (Sörensen & Janssens, 2003).

Odločitvene tabele lahko transformiramo v odločitvena drevesa in nato v poslovna pravila. Lahko pa odločitvene tabele transformiramo direktno v poslovna pravila. Podobno lahko tudi odločitvena drevesa transformiramo v pravila in zmanjšamo redundanco, kajti iz pravil v obliki tabel ali dreves so bolj opazni enaki pomeni pravil. Posledično odpravimo podvajanja in zmanjšamo število poslovnih pravil, ki jih je potrebno upoštevati (Shamim, Hussain & Shaikh, 2010)

1.2 Procesi

Poslovni procesi so logično med seboj povezane aktivnosti, s točno določenimi vhodi in izhodi, katerih namen je, da zagotavljajo glavno dejavnost in uspešno poslovanje podjetja. Procesi spreminjajo vire dosledno, zanesljivo in ponovljivo v izhode za doseg ciljev podjetja (Zairi, 1997). Poslovni procesi so dodana vrednost za podjetje, saj je lahko zaradi uporabe le teh bolj uspešno in bolj konkurenčno na trgu (Abdel-Fattah, Khedr & Aldeen, 2017).

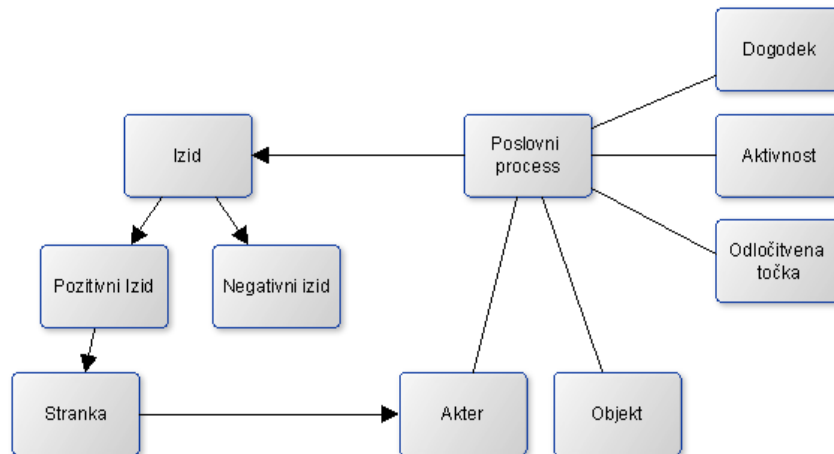
Kot že omenjeno v prejšnjem poglavju s prehodom na procesno urejeno poslovanje podjetja nastopi tudi koncept BPM. Obstaja veliko člankov, ki definirajo BPM. Zairi (1997) navaja, da je namen BPM določiti glavne procese in jih ustrezno dokumentirati, zagotoviti da so postopki ponovljivi, dosledni in kakovostni, procese prilagajati potrebam in željam naših strank, analizirati in meriti uspešnost procesov na podlagi podanih ciljev, procese stalno nadgrajevati in optimizirati ter spreminjati kulturo zaposlenih kot tudi poslovanje podjetja. Po navedbi avtorjev Lee in Dale (1998) je skupno uporabljenim definicijam o BPM, da je strukturirano, analitično, navzkrižno funkcionalno in je koncept, s katerim izboljšamo poslovne procese za doseg ključnih ciljev podjetja.

1.2.1 Lastnosti poslovnih procesov

Poslovni proces je zbirka povezanih aktivnosti, dogodkov in odločitvenih točk (slika 3). Aktivnost ponazarja časovno dlje časa trajajočo eno ali več opravil skupaj. Dogodki ponazarjajo nekaj kar se zgodi med samim izvajanjem poslovnega procesa in vpliva na izvajanje aktivnosti. Dogodek se zgodi v trenutku. Potek izvedbe procesa določajo odločitvene točke. V vsak poslovni proces so vključeni akterji, to so ljudje, organizacije, programski sistemi in objekti v obliki fizičnih ali informacijskih virov. Akterji so lahko notranji, ki delujejo znotraj organizacije, ali pa so zunanji, ki delujejo izven organizacije. Kot fizični vir lahko nastopa oprema, izdelek, dokumentacija, kot informacijski vir pa zapisi in dokumentacija v digitalni obliki. Vsak poslovni proces ima pozitiven ali negativen izid. Pozitiven izid pošlje ali dostavi rezultat za stranko. Negativni izid procesa pomeni, da se zgodi zavrnitev aktivnosti oz. se ne zgodi nič (Dumas, La Rosa, Mendling & Reijers, 2013, str. 3–6; OMG, 2011, str. 29).

Poslovne procese lahko delimo na temeljne, podporne in managerske procese. Temeljni procesi so tisti, ki pokrivajo in zagotavljajo glavne aktivnosti in posledično povečajo dodano vrednost podjetja. Cilj temeljnih procesov je zadovoljiti potrebe kupcev. Podporni procesi nudijo podporo temeljnim procesom. Na dvig dodane vrednosti podjetja vplivajo le posredno. V skupino podpornih procesov lahko uvrščamo kadrovske, finančne, infrastrukturne, razvojne področje. Managerski procesi upravljajo poslovanje podjetja, določajo pravila in usmeritve tako za temeljne kot podporne procese. V to skupino uvrščamo strateško načrtovanje, proračune, skladnost delovanja in krizni management (Dumas, La Rosa, Mendling & Reijers, 2013, str. 41; Kovačič & Bosilj Vukšić, 2005, str. 31).

Slika 3: Sestavine poslovnega procesa



Prيرهjeno po Dumas , La Rosa, Mendling & Reijers (2013, str. 8).

Povezave med procesi lahko uvrstimo v kategoriji zaporedje: dekompozicija in specializacija. Procesna spadata v kategorijo zaporedja, ko sta povezana zaporedno in izhod enega procesa predstavlja vhod v naslednji proces. Dekompozicija procesov pomeni, da proces razdelimo na podprocese. Povezave med procesom in podprocesom so navpične in hierarhično urejene. Specializacija pomeni, da obstaja več različnih možnosti splošnega procesa. Splošni proces se deli glede na različne potrebe izdelkov, storitev in uporabnikov. Povezane aktivnosti, ki jih ni mogoče več členiti na podprocese, imenujemo delovni proces ali postopki (Dumas , La Rosa, Mendling & Reijers, 2013, str. 42–44; Kovačič & Bosilj Vukšić, 2005, str. 179–180).

1.2.2 Modeliranje poslovnih procesov

Modeliranje poslovnih procesov pomeni vizualno predstavitev poslovnih procesov tako v fazi posnetka stanja (angl. as-is), kot v fazi načrtovanja izboljšave procesov (angl. to-be). To pomeni, da ponazarjamo elemente procesov, kot so aktivnosti in njihove odvisnosti, kako poteka tok podatkov, kakšne so vloge akterjev in njihove cilje (Lindsay, Downs & Lunn, 2003). Po mnenju Juan in Yuan (2013) je modeliranje poslovnih procesov najpomembnejša naloga v procesno organiziranih podjetjih. Lahko rečemo, da je le-to ena izmed ključnih faz pri informatizaciji poslovanja. Z modeliranjem pripomoremo k optimizaciji procesov, vplivamo na kakovost poslovnih procesov in posledično na kakovost podatkov. Z modeliranjem tudi dosežemo, da so procesi čim bolj točno razumljivi vsem udeležencem, od poslovnih uporabnikov, analitikov do razvijalcev.

Modeliranje poslovnih procesov lahko razvrstimo glede na osrednji konstrukt modeliranja. Lahko je osrednji konstrukt osredotočen na procese, ki so prikazani kot mreža med sabo povezanih aktivnosti. Modeliranje je lahko objektno usmerjeno ali je osredotočeno na vire prikazano v omrežju (zur Muehlen & Indulska, 2010).

Poznamo različne standarde modeliranja. Med bolj razširjenimi standardi, osredotočenimi na procese, je Business Process Modeling Notation (v nadaljevanju BPMN). Standard podpira veliko orodji na trgu in je enostaven za uporabo. Obenem omogoča širok spekter elementov za modeliranje tudi kompleksnih procesov. Lahko rečemo, da BPMN služi kot specifikacija, ki je prenosljiva med organizacijami in je razumljiva širšemu krogu zaposlenih. Vendar to ne pomeni, da je BPMN najboljša izbira za vsak proces in za vsako organizacijo. V primeru napačno izbranega standarda modeliranja lahko privede do napačne interpretacije procesov bolj kompleksnih in težje razumljivih procesov, potrate časa in neučinkovite izvedbe procesov (Abdel-Fattah, Khedr & Aldeen, 2017; Krogstie, 2013; OMG, 2011).

Standard BPMN določa nabor grafičnih simbolov s točno določenim namenom. Kronološko se je standard nadgrajeval in obstaja več različic. Modele BPMN lahko razdelimo v tri podkategorije (OMG, 2011, str. 21–27):

- Zasebne in javne procese. Lahko jih poimenujemo tudi kot delovni tok ali orkestracija. Zasebne procese dalje ločimo na neizvršljive, ki služijo zgolj za dokumentacijo, in izvršljive, ki so izvedljivi. Zasebni procesi se izvajajo znotraj poslovne enote oz. znotraj bazena, če se izrazimo simbolno. Javni procesi so tisti, ki komunicirajo med različnimi procesi oz. deležniki. Aktivnosti in dogodke v tem primeru imenujemo stične točke.
- Koreografije. To je proceduralna pogodba med vsaj dvema deležnikoma. Koreografijo sestavlja množica aktivnosti, dogodkov in odločitev, podobno kot to velja za procese. Razlikuje se v tem, da vsaka aktivnost predstavlja izmenjavo sporočil med vsaj dvema deležnikoma.
- Kolaboracije. V to skupino spadajo javni ali zasebni procesi in določajo sodelovanje med vsaj dvema poslovnima subjektoma, deležnikoma kolaboracije. Skupina kolaboracije pomeni izmenjavo sporočil med bazeni, kjer je vsak poslovni subjekt v svojem bazenu. Izmenjava je prikazana s tokom sporočil, sami bazeni pa so lahko tudi prazni. V kolaboracije so lahko vključeni tako bazeni, procesi ali koreografije.

Notacijo BPMN zaradi lažje uporabe in interpretacije delimo v pet glavnih skupin (OMG 2011, str. 29–41):




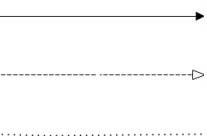
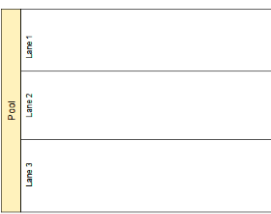

- Elementi toka procesa (angl. Flow Objects). To je glavna skupina BPMN, ki jo sestavljajo elementarni dogodki, aktivnosti in razvejišča (angl. Gateways).
- Povezovalni elementi (angl. Connecting Objects). Namen skupine je, da povezuje različne elemente med sabo v neko celoto. Skupino sestavljajo elementi, kot so potek izvajanja, potek sporočila, asociacija.
- Plavalne steze (angl. Swimlanes). V to skupino uvrščamo elemente, kot so bazeni, ki predstavljajo organizacijske enote oz. glavne vloge v organizaciji in steze, ki delijo notranjost bazena oz. procesa, glede na vloge, ki jih zaposleni opravljajo.
- Artefakti (angl. Artifacts). Namen skupine je, da ponazarja dejstva oz. informira, na sam potek procesa pa nimajo vpliva. Artefakti dodatno prispevajo k izboljšanju razumljivosti procesov za uporabnike. Sestavljata jo elementa skupina in opomba, lahko pa vključimo

poljubne lastne artefakte. Namen skupine je, da združuje elemente v kategorije za dokumentiranje in analizo, kjer opombe dodatno pojasnjujejo elemente v procesu.

- Podatki. V skupino spadajo elementi, kot so podatki na vходу, izhodu, shranjeni podatki.

Podrobneje pogledjmo značilnosti nekaterih glavnih elementov BPMN. Nekaj osnovnih simbolov po posameznih sklopih prikazuje tudi tabela 1.

Tabela 1: Primeri simbolov BPMN po skupinah

Elementi BPMN	Opis
	Dogodki: začetni, osrednji, končni, sporočila, časovnik, povezava, prekinitvev, razveljavitev
	Aktivnosti: naloga, podproces, transakcija
	Razvejišča: ekskluzivna, vključujoča, vzporedna, kompleksna
	Povezovalni elementi: potek izvajanja, potek sporočila, asociacija
	Plavalne steze: bazeni in poli
	Artefakti: Skupina in opomba

Prيرهjeno po OMG (2011, str. 29–41).

Glede na pozicijo v samem procesu poznamo začetne, končne in vmesne dogodke. **Dogodki** so kot sprožilci ali kot rezultati procesa. Simbolno so prikazani v obliki kroga. **Aktivnosti** prikazujejo opravila organizacije in se delijo na primitivne in sestavljene aktivnosti. Primitivne aktivnosti so opravila, kot so sprožitev aplikacije, ročno opravilo, klic skripte, pošiljanje sporočila, prejetje sporočila in klic poslovnega pravila. Med sestavljene aktivnosti uvrščamo

elemente, kot so podprocesi, ki delijo klice na manjše procese, ki so vključeni v sestavljeno aktivnost. Modeliranje z elementi sestavljene aktivnosti uporabimo pri kompleksnejših in obsežnejših procesih, kjer modeliramo v fazah z vključitvijo podprocesov do najnižje ravni, primitivnega procesa, ki je na ravni opravila in se lahko implementira. Aktivnosti so simbolno prikazane s pravokotniki. V skupino **odločitvenih točk**, ki smo jo omenili v predhodnem poglavju, spadajo razvejišča, ki nadzirajo potek toka v procesu, da omogočijo ali zavrnejo prehod. Določajo, kako se dogodki in aktivnosti med samo združujejo oz. delijo. Simbolno so razvejišča prikazana v obliki diamanta. Razvejišča so lahko ekskluzivna, kjer je možna izbira samo ene izmed danih poti, vključujoča, kjer je lahko vsaka izmed poti možna izbira, vzporedna, kjer se vse poti izvajajo vzporedno ali kompleksna, kjer je večje število razvejišč in pravilo, katere poti so omogočene. Prav tako, kot je namen razvejišč, da se pri izvajanju tokovi delijo, se lahko tokovi tudi združujejo. Med **povezovalne elemente** uvrščamo potek izvajanja, ki točno določa vrstni red izvedbe aktivnosti v procesu. Element poteka sporočil določa izmenjavo sporočil med izključno poslovnima subjektoma, torej je ta element izključno del kolaboracije. Element asociacija določa izključno povezavo med artefakti in ostalimi elementi. Znotraj bazena lahko nastopajo vsi do sedaj omenjeni elementi. Če elementov ni, govorimo o tako imenovani črni škatli. V tem primeru bazeni služijo zgolj za predstavitev deležnikov kolaboracije. Znotraj bazena lahko nastopajo steze, ki delijo bazen glede na vloge. Elementi toka procesa so med stezami povezani z elementi poteka izvajanja. Povezave med bazeni pa so možne izključno z elementi poteka sporočila (Flowers & Edeki, 2013; OMG, 2011).

Z večanjem števila uporabljenih simbolov BPMN se stopnjuje kompleksnost in razumljivost procesnih modelov. Prepoznavanje različnih vzorcev procesnega toka je metoda, ki nam lahko pri tem zelo koristi. Tip vzorca določajo razvejišča, torej imamo lahko ekskluzivne, vključujoče, vzporedne in kompleksne vzorce procesnega toka. S postopkom prepoznavanja vzorcev lahko ponovno uporabimo strukturo procesnega toka in zagotovimo boljšo razumljivost in lažjo analizo procesa (Juan & Yuan, 2013).

1.2.3 Združitev poslovnih pravil in procesov

Razlika med modeliranjem poslovnih pravil in procesov je v tem, da so poslovna pravila deklarativno usmerjena, določajo, kaj naj se zgodi, poslovni procesi pa proceduralno usmerjeni, določajo, kako naj se zgodi. Pravila so prikazana običajno v tekstovni obliki, procesi pa grafično s širokim nabor simbolov. Integracija poslovnih pravil in procesov pomeni prilagajanje modelov glede na standarde, ki jih posamezna metoda ima, če je seveda to smiselno (Cheng, Sadiq & Indulska, 2011; zur Muehlen, Indulska & Kittel, 2008).

Pomembno je, da se zavedamo prednosti in slabosti, če so poslovna pravila integrirana v poslovni proces ali ne. Če je pravilo integrirano v poslovni proces, se pogosto spreminja in je vključeno v več poslovnih procesih. Ob vsaki spremembi pravila je le-to potrebno nadgraditi v vseh poslovnih procesih, kjer to pravilo nastopa. Pravilo, ki je neodvisno od procesov, je modelirano samo enkrat in je lahko vključeno v več procesov, neodvisno. Ob spremembah pravilo

nadgradimo samo na enem mestu. Slabost ločenega modeliranja je, da lahko privede do kršitve skladnosti med pravili in procesi v manjkajočih pravilih za poslovni proces (Wang, Sadiq & Indulska, 2014).

Odločitev za integracijo poslovnih pravil v proces ni enostavna. V pomoč pri odločitvi so nam naslednji dejavniki (Wang, Sadiq & Indulska, 2014; zur Muehlen, Indulska & Kittel, 2008):

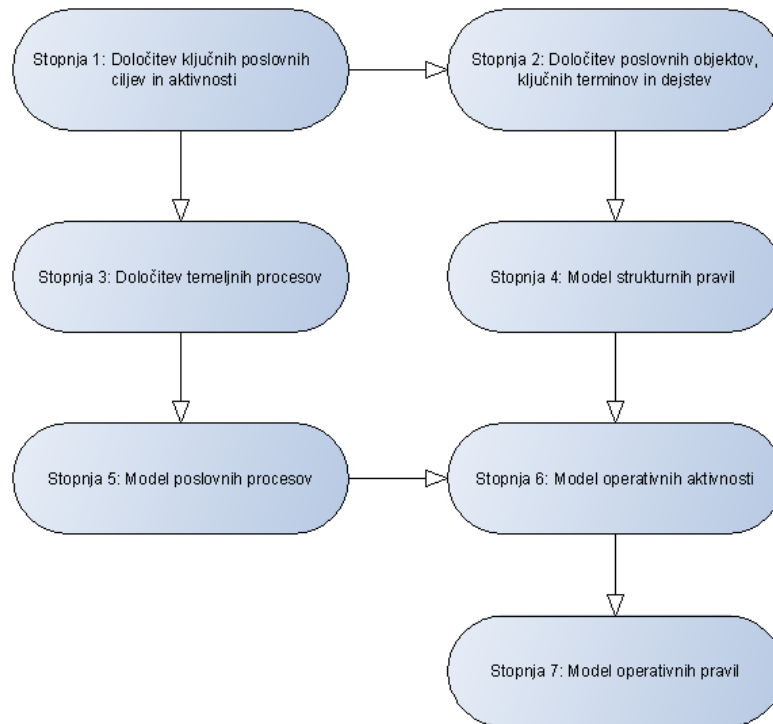
- Dostopnost: ali uporabniki lahko spreminjajo pravila.
- Vidik sprememb: kako hitro in kako kompleksno je prilagajanje komponent pravil v procesu.
- Prilagodljivost: v kolikšnem času in s kakšnim vložkom spremenimo proces, ko se spremeni pravilo.
- Zavedanje: ali se uporabniki zavedajo posledic spreminjanja pravil in jih je mogoče napovedati.
- Kritičnost: kako pomembno je, da se pravilo upošteva. Ali je pravilo nujno za uspešno izvedbo procesa, ali je manjšega pomena.
- Izraznost: ali imamo enostavnejše ali bolj kompleksno pravilo, ali je pravilo lažje razumljeno s simboli ali v tekstovni obliki.
- Odgovornost upravljanja: kdo je zadolžen za skladnost med poslovnimi pravili in procesi.
- Odgovornost za implementacijo: kdo je odgovorna oseba za implementacijo pravila v procesih, ali gre za poslovnega uporabnika ali za osebo z več znanja o informacijski tehnologiji (angl. Information Technology, v nadaljevanju IT).
- Stopnja spremembe: kako pogosto se poslovno pravilo lahko spremeni.
- Ponovna uporabnost: ali obstaja možnost, da se isto pravilo uporabi večkrat v različnih kontekstih.
- Vir pravil: kakšen je vir za pravila, ali gre za zunanje predpise, zakonodajo, ali notranje vire, katere določa oz. predpisuje podjetje.
- Obseg vpliva: kakšen bo obseg vpliva pravila, ali govorimo o vplivu pravila na dejavnost, oddelek, celoten proces ali na celotno delovanje podjetja.

Na podlagi informiranja in zavedanja o posledicah se lažje odločimo ali pravilo integriramo v proces ali ne. Poslovna pravila in procese lahko integriramo, če velja da komponente pravil lahko z minimalnimi vložki vključimo v proces in ob spremembi pravil proces tudi nadgradimo, da lahko nadziramo vpliv sprememb pravil na procese, da je pravilo lahko predstavljeno v notacijo procesnega modeliranja, skladnost pravila in procesov sta zagotovljena, oseba odgovorna za implementacijo ima osnovna IT znanja, pravilo se spreminja redko in je uporabno le enkrat, gre za zunanji vir kot so regulative, predpisi, da omogočimo sledljivost in gre za manjši obseg vpliva pravila v organizaciji (Wang, Sadiq & Indulska, 2014; zur Muehlen, Indulska & Kittel, 2008).

Za pristop k integraciji pravil in procesov lahko uporabimo model, kot prikazuje slika 4. Integracija se izvede v sedmih stopnjah. Najprej se določijo ključni poslovni cilji in aktivnosti podjetja, kar je pomembno iz vidika poslovnih pravil kot procesov. Prva stopnja je podlaga, da lahko v drugi stopnji določimo ključne poslovne objekte, termine in dejstva, in v tretji stopnji

temeljne procese. Na podlagi terminov in dejstev določimo model strukturnih pravil v četrti stopnji, ki predstavljajo objekte in povezave med njimi. Sočasno pa v peti stopnji modeliramo temeljne poslovne procese, kjer prepoznamo povezave med procesi. Na podlagi četrte in pete stopnje se v šesti stopnji določijo operativne aktivnosti, ki vsak proces razstavi do podrobnosti. Za zaključek definiramo še operativna pravila, ki nadzorujejo obnašanje virov (zur Muehlen, Indulska & Kittel, 2008).

Slika 4: Model za integracijo poslovnih pravil in procesov



Prirjeno po zur Muehlen, Indulska & Kittel (2008).

1.3 Podatkovno modeliranje

Podatkovno modeliranje ima po mnenju številnih avtorjev izredno pomembno vlogo pri izgradnji aplikativnih rešitev (Elmasri & Navathe, 2016, str. 59; Hoberman, 2009, str. 37; Teorey, Yang & Fry, 1986, 2011, str. 10). Rezultat podatkovnega modeliranja je prikaz ključnih podatkov in povezav med njimi, z upoštevanjem poslovnih pravil. Namen je zagotoviti ključne podatke za poslovne uporabnike kot tudi za programske rešitev, ki te podatke potrebuje in zagotoviti uspešno izvedbo poslovnih procesov. Na ta način omogočimo, da so podatki pravilno shranjeni v okolju, ki je skrbno varovano.

Podatkovno modeliranje je načrtovano shranjevanje podatkov podjetja za skrbno in pravilno ravnanje. Pravilna uporaba podatkov omogoči prave informacije, kar pomeni dodano vrednost za podjetje. Pri slabem načrtovanju shranjevanja podatkov se lahko zgodi, da so podatki neustrezno, pomanjkljivo shranjeni in posledično pri uporabi shranjenih podatkov pridobimo

napačne informacije. Na podlagi omenjenih dejstev je pravilno podatkovno modeliranje pomembna faza oblikovanja podatkovnega okolja (Coronel, Morris & Rob, 2013, str. 46).

Pomembna vloga uporabe podatkovnih modelov je v tem, da se izboljša medsebojna komunikacija različnih deležnikov, tako v sami fazi načrtovanja in izgradnje podatkovnih modelov kot tudi po zaključku modeliranja. Pri modeliranju bomo uspešni, če bomo dosledni in natančni pri uporabi pravil. Le tako bodo deležniki enako razumeli ključne koncepte, povezave med njimi od najvišje ravni abstrakcije do najnižje, torej do izgradnje modela v podatkovni bazi. Obenem je vloga podatkovnih modelov tudi v tem, da boljše razumemo aplikativne rešitve in poslovanje podjetja. Modeli lahko služijo kot pripomoček pri analizi vpliva sprememb obstoječe aplikativne rešitve in kot dobra dokumentacija za izobraževanje novo zaposlenih.

1.3.1 Lastnosti podatkovnega modeliranja

Rezultat procesa podatkovnega modeliranja je podatkovni model. Modeliranje poteka v treh fazah, ki so med sabo tesno povezane in pomembne za oblikovanje končne podatkovne baze, kar prikazuje slika 5.

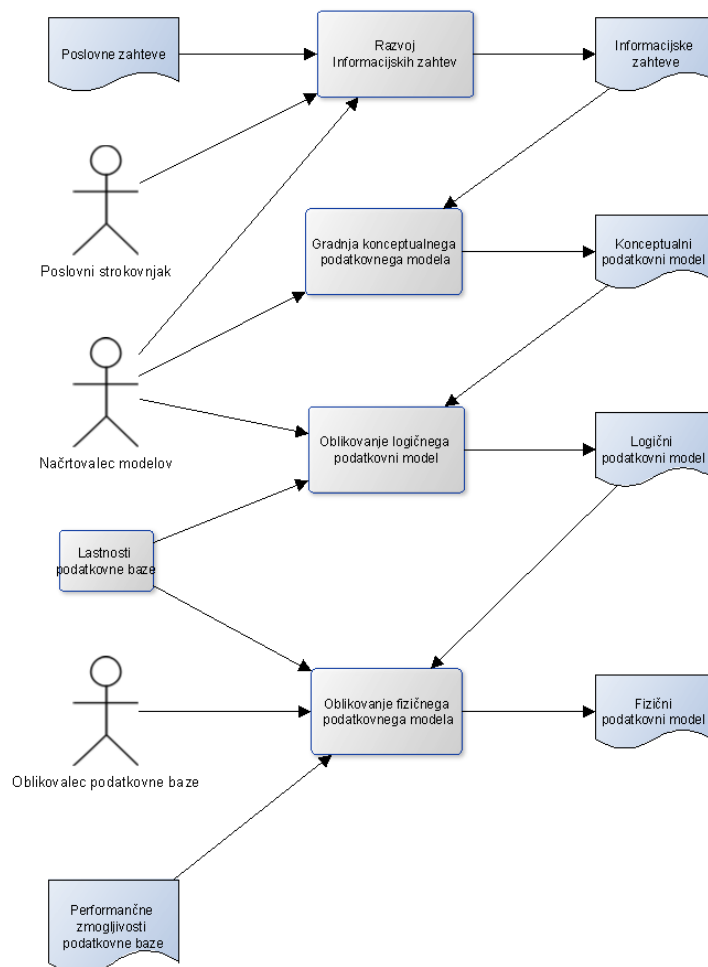
V prvi fazi se na podlagi poslovnih pravil določijo informacijske zahteve, ki so osnova za razvoj konceptualnega modela (angl. Conceptual Data Model). Le-ta odraža dožemanje stvarnega sveta, torej podatkov v grafični obliki. Prikazuje zbirke in povezave med ključnimi poslovnimi koncepti. V drugi fazi sledi logični model (angl. Logical Data Model) ali izvedbeni oz. implementacijski model (angl. Implementation Data Model), ki prikazuje organizacijo podatkov v podatkovni bazi. Osnova za izgradnjo modela v tretji fazi, ki ga imenujemo fizični model (angl. Physical Data Model), je konceptualni model. Fizični model je prikaz dejanskega stanja objektov, ki so shranjeni v podatkovni bazi, z vključenimi vsemi objekti in lastnostmi, ki so uporabljene pri fizični postavitvi podatkovnega okolja za shranjevanje podatkov. Torej fizični model odraža fizično in tehnično implementacijo sistema. Fizični model omogoča dodatne možnosti kot so particioniranje, indeksiranje in združevanje podatkov. Izgradnja konceptualnega, logičnega ali fizičnega podatkovnega modela je naloga IT strokovnjaka z znanjem modeliranja oz. oblikovanjem podatkovne baze (Li & Chen, 2009, str. 4; Navathe, 1992).

Za isti poslovni proces je možnih več različic podatkovnega modela. Odvisno je od razumevanja, kateri so ključni podatki, kako jih organizirati in katera je uporabljena tehnika modeliranja. V želji, da je podatkovni model čim boljši, se osredotočimo na ključne pokazatelje kakovosti po Simsionu in Wittu (2005, str. 10–15):

- popolnost: so vsebovani vsi ključni podatki,
- brez podvajanja: vsebina se ne podvaja, če ni nujno potrebno,
- uveljavljanje poslovnih pravil: poslovna pravila so vključena v model in so pravilno interpretirana,
- ponovna uporabnost podatkov: podatki so uporabni tudi v primeru nepredvidenih, naknadnih procesnih zahtev,

- prilagodljivost: ob spremembi poslovnih pravil je model prilagodljiv z minimalnimi posegi,
- preprostost: poslovni model je enostaven, brez odvečnih tabel in povezovanj,
- razumljivost: poslovni model je razumljiv za poslovne uporabnike, analitike, razvijalce,
- zmogljivost: podatkovni model modeliran tako, da zagotovi odzivnost dostopanja do podatkov v podatkovni bazi,
- pravo ravnovesje: predhodno omenjeni pokazatelji so si lahko nasprotujoči, zato je potrebno poiskati pravo ravnovesje med omenjenimi lastnostmi, za pridobitev čim bolj kakovostnega modela.

Slika 5: Oblikovanje podatkovne baze



Prirjeno po Simsion & Witt (2005).

1.3.2 Konceptualni model

Konceptualni model je začetna faza načrtovanja podatkov v bazi. Osnovni gradniki so entitete, atributi in povezave. Model gradimo na podlagi poslovnih pravil in informacij zbranih v fazi analize in načrtovanja poslovne rešitve. Lahko rečemo, da gradniki predstavljajo samostalnike in glagole, pridobljene iz poslovnih pravil. Poznamo različne vrste konceptualnih modelov, ki pa

imajo skupne lastnosti. Med bolj enostavne za uporabo in tudi pogosto uporabljene uvrščamo model entitet povezav (angl. Entity Relationship Model, v nadaljevanju ER model). Omenjena tehnika je enostavna za razumevanje, saj v fizični obliki pomeni povezovanje dvodimenzionalnih tabel. V našem primeru se tako osredotočimo na ER model. V nadaljevanju si pogledimo lastnosti tega modela.

Sestavine ER modela so entitetni tip, entiteta, atribut, enolični identifikator in povezava. Entitetni tip predstavlja subjekte, o katerih želimo shranjevati podatke, in imajo enake lastnosti. Primer ek entitetnega tipa imenujemo entiteta. Torej entitetni tip je tabela, v kateri bomo shranjevali podatke, vsak zapis v tabeli pa predstavlja entiteto. Primerki entitet so lahko anketar, vprašalnik. Vsi anketarji predstavljajo entitetni tip anketar oz. vsi vprašalniki predstavljajo entitetni tip vprašalnik. Za vsak entitetni tip izberemo informacije, za katere menimo, da so tako pomembne, da jih shranjujemo v podatkovni bazi. Vsebina je shranjena v atributih. Poznamo numerične, znakovne ali datumske attribute.

Mejne vrednosti, ki so dopustne za vsak atribut, imenujemo domena. Glede na množico vrednosti, ki jih entitete lahko zasedejo, attribute delimo na (Coronel, Morris & Rob, 2013, str. 174 - 178; Elmasri & Navathe, 2016, str. 64–66):

- Elementarne: vrednost atributa se ne more deliti na več možnih atributov. Primer elementarnega atributa je spol.
- Sestavljene: vrednost atributa je sestavljena iz več nizov, ki se lahko delijo hierarhično na več elementarnih atributov. Primer sestavljenega atributa je naslov, ki se lahko deli na ulico, hišno številko, pošto. Sestavljen atribut uporabimo, ko se elementarna vrednost atributa ne uporablja.
- Enovrednostne: atribut entitete ima natanko eno vrednost. Primer enovrednostnega atributa je EMŠO, ki je identifikacijska številka osebe.
- Večvrednostne: atribut entitete lahko vsebuje niz vrednosti. V tem primeru omejimo zgornjo mejo možnih vrednosti atributa. Primer večvrednostnega atributa je telefonska številka gospodinjstva, kjer ima lahko entiteta gospodinjstvo več telefonskih števil, ni pa nujno. Večvrednostne attribute je priporočljivo deliti na več enovrednostnih znotraj entitetnega tipa ali pa se prav tako večvrednosti atribut deli na enovrednostne, pri čemer se kreira nov entitetni tip z dvema atributoma. Prvi atribut vsebuje skupine enovrednostnih (npr. tel1, tel2), ki smo jih dobili iz večvrednostnih atributov. Drugi atribut pa vsebuje dejansko vrednost za vsako novo skupino.
- Izpeljane: atribut je izračunan na osnovi enega ali več elementarnih atributov. Izpeljanega atributa običajno ne shranjujemo fizično, saj nastopi težava, ko se elementarni atribut spremeni. Posledično se mora spremeniti tudi izpeljani. Iz tega sledi, da je priporočljivo shranjevanje samo osnovnega atributa oz. atributov, iz katerega se potem lahko izračuna izpeljani atribut. Algoritem izpeljanega atributa se vključi kot del programske rešitve.
- Prazne vrednosti (angl. NULL Values): posebno pozornost je potrebno nameniti praznim vrednostim. Delimo jih na vrednost - ni možno (angl. Not applicable), ko vemo, da vrednost

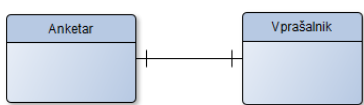
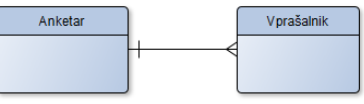
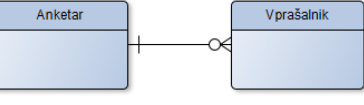
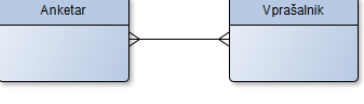
ne more obstajati in na vrednost - neznano (angl. Unknown), ko vemo, da vrednost lahko obstaja, vendar je ne poznamo. Primer atributa je poklic. Če imamo šoloobveznega otroka, potem je vrednost atributa - ni možno, če pa je oseba zaposlena, vendar ne vemo poklica, potem je vrednost atributa - neznano.

Ključni atribut je samo en atribut ali pa je sestavljen iz več atributov skupaj in enolično določa entitetni tip. To pomeni, da se vsaka vrednost ključnega atributa pojavi v entitetnem tipu natanko enkrat in tako lahko med sabo ločimo entitete. Ključni atribut imenujemo enolični identifikator entitetnega tipa. Entitetni tip ima lahko več kombinacij ključnih atributov, ali pa ključnega atributa sploh nima. Če ključnega atributa ni, to imenujemo šibek entitetni tip (Elmasri & Navathe, 2016, str. 69), za katerega velja, da ne more obstajati brez entitetnega tipa.

Povezava združuje entitetne tipe, ki so med seboj odvisni. Povezavo opišemo z glagolom, ki smiselno opredeli odnos med entitetnima tipoma. Kardinalnost pomeni tip povezave, torej kakšna je stopnja povezljivosti dveh entitetnih tipov. Primerki povezav med entitetnima tipoma prikazuje tabela 2. Entitetna tipa sta lahko v povezavi:

- ena proti ena: ena entiteta prvega entitetnega tipa je v povezavi natanko z eno entiteto drugega tipa,
- ena proti mnogo: ena entiteta prvega entitetnega tipa je v povezavi z več entitetami drugega tipa,
- mnogo proti mnogo: več entitet prvega entitetnega tipa je v povezavi z več entitetami drugega entitetnega tipa.

Tabela 2: Kardinalnost in opis povezave entitetnih tipov

Kardinalnost povezave	Opis
	En anketar izpolni natanko en vprašalnik izbrane osebe.
	En anketar izpolni več vprašalnikov izbranih oseb.
	En anketar lahko izpolni več vprašalnikov izbranih oseb ali pa nobenega.
	En anketar izpolni več vprašalnikov izbranih oseb, en vprašalnik izpolnjuje več anketarjev.

Vir: lastno delo.

Kardinalnost povezave ena se lahko nadomesti tudi s kardinalnostjo ena ali nič, kar pomeni, da je ena ali pa ni nobene entitete v povezavi z drugo entiteto. Entiteta tipa sta v povezavi dvosmerno, torej prvi entitetni tip v povezavi z drugim entitetnim tipom in obratno, kjer pa je lahko kardinalnost teh dveh povezav med seboj različna.

Pri modeliranju obstajajo različne tehnike. V našem primeru je uporabljena Crow's Foot notacija, kjer niso navedeni atributi. Entitetni tip je prikazan v obliki pravokotnika. V zgornjem delu pravokotniku je navedeno ime podatkovnega tipa, v spodnjem delu je prostor za attribute, kjer je ključni atribut označen.

V fazi konceptualnega modela imajo pomembno vlogo poslovna pravila, saj ta določajo komponente ER modela, torej entitetne tipe, attribute, glavni ključ, povezave med njimi in kardinalnost (Coronel, Morris & Rob, 2013 str. 180).

1.3.3 Logični model

Logični model prikazuje lastnosti objektov v bazi in povezave med njimi. Logični model je neodvisen od podatkovnega okolja. Objektom se določijo atributi in njihove lastnosti. Primeri logičnega modela so hierarhični, mrežni, relacijski in objektni.

Logični model je razširitev konceptualnega modela in predstavlja logično postavitve objektov v podatkovni bazi. V našem primeru smo se med konceptualnimi modeli osredotočili na ER model, zato se v tem poglavju osredotočimo na relacijski model. Velja namreč, da se ER transformira v relacijski model, ki je pomemben mejnik v predstavljanju in shranjevanju večje količine podatkov (Codd, 1983).

Logični model opredeljuje povezave med relacijami, ki so predstavljene v obliki tabel. Relacije so preslikave entitetnih tipov konceptualnega modela. Le-te so sestavljene iz množice, kjer so vrstice n-terice relacije oz. entitete konceptualnega modela, stolpci pa predstavljajo attribute relacije oz. entitete (Teorey, Lightstone, Nadeau & Jagadish, 2011, str. 2). Vrstni red vrstic ali stolpcev je nepomemben. Vsak atribut ima točno določeno vrsto in dolžino podatkovnega tipa. Vrednosti atributov so omejene z domenami. Atribut ali več atributov skupaj predstavljajo različne vrste ključev relacij, katerih namen je, da skrbijo za enoličnost zapisov in povezavo med njimi. Vse omenjene lastnosti relacij in povezav določajo poslovna pravila.

Lahko rečemo, da so bistvene razlike logičnega od konceptualnega modela v drugačnem poimenovanju elementov, v točno določenem naboru atributov za vsako relacijo, v določitvi ključev relacije, kot so primarni, tuji ključ, v določitvi odvisnosti med relacijami oz. nadomestitvi povezave mnogo proti mnogo, če ni bila odpravljena v fazi konceptualnega modeliranja (Fiorina, 2022).

V nadaljevanju si pogledjmo pomen zgoraj omenjenih lastnosti logičnega modela.

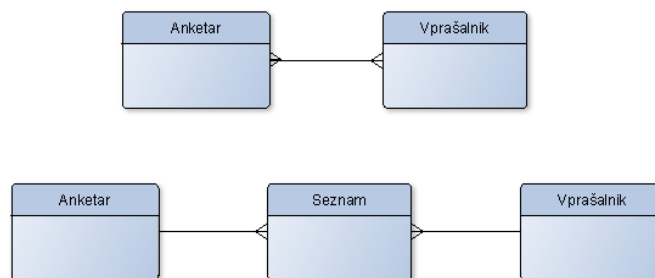
Namen ključev je, da lahko na podlagi znane vrednosti atributa ali atributov določijo vrednosti ostalih atributov relacije. Attribute, ki so določeni na podlagi drugih znanih atributov, imenujemo odvisni atributi (Coronel, Morris & Rob, 2013, str. 121–122). Kot smo že omenili, poznamo različne vrste ključev.

Primarni ključ je enolični identifikator relacije. Lahko ga predstavlja samo en atribut ali pa je sestavljen iz več atributov. Ključni atribut konceptualnega modela je tako primarni ključ v logičnem modelu. Namen primarnega ključa je, da zagotavlja celovitost relacije, da enolično določa zapis relacije, torej so vrednosti atributov brez podvajanj in ne vsebujejo praznih vrednosti (Coronel, Morris & Rob, 2013, str. 123). Če relacija vsebuje ključni atribut, to imenujemo naravni primarni ključ. V primeru, da relacija ne vsebuje ključnega atributa, pomeni, da ga ustvarimo umetno, kar imenujemo surogatni primarni ključ. Običajno je umetno ustvarjeni ključni atribut kar sekvenca, ki se z dodajanjem zapisov povečuje za ena. Kandidat za ključ imenujemo vsak atribut ali skupino atributov, ki lahko enolično določajo relacijo. Prednost pri izbiri primarnega ključa imajo atributi, katerih vrednosti niso pogosto spremenljive. Tako imajo prednost številski atributi, ki označujejo identiteto in obenem zagotavljajo varovanje identitete subjekta (Coronel, Morris & Rob, 2013, str. 242–243).

Tuji ključ je atribut, ki predstavlja primarni ključ druge relacije. Relacija, ki vsebuje tuji ključ je podrejena relaciji, ki vsebuje ta atribut kot primarni ključ. Relacijo s tujim ključem imenujemo otrok, nadrejena relacija s primarnim ključem pa je mama. Relacija lahko vsebuje več tujih ključev in je v povezavi z več nadrejenimi relacijami. Povezava mama-otrok je pomembna, saj na ta način zagotovimo, da v relaciji otrok nastopajo izključno veljavne vrednosti, ki so tudi v relaciji mama in tako poskrbimo za referenčno celovitost entitet.

Relacije so tako kot pri konceptualnem modelu, v povezavi med sabo ena proti ena, ena proti mnogo ali mnogo proti mnogo. Pri povezavi ena proti ena velja, da sta povezani tako, da je ena vrstica prve relacije v povezavi z natanko eno vrstico druge relacije. Povezava ena proti mnogo pomeni, da je ena vrstica prve relacije v povezavi z eno ali več vrsticami druge relacije. Velja pravilo, če sta relaciji v povezavi mnogo proti mnogo, potem se ustvari nova, vmesna relacija, ki poskrbi, da se povezava deli na ena proti mnogo za povezavo prve in vmesne relacije in mnogo proti ena za povezavo vmesne in druge relacije (slika 6).

Slika 6: Delitev povezave mnogo proti mnogo z vmesno relacijo



Vir: lastno delo.

1.4 Podatkovna baza

Podatkovna baza je zbirka podatkov, ki so shranjeni v varovanem in nadzorovanem okolju. Sistem za upravljanje zbirk podatkov (angl. Data Base Management System, v nadaljevanju DBMS) omogoča shranjevanje definicij o podatkih in povezavah med njimi v podatkovnem slovarju, shranjevanje samih podatkov, poglede, sintakso za delo s podatki, omejuje in nadzoruje podatke in transakcije nad objekti. Obenem omogoča sočasen dostop do podatkov, z različnimi pravicami, za uporabnike ali aplikativne rešitve. Prednost uporabe DBMS je v tem, da je lažja izmenjava in dostop do večje količine podatkov, izboljšana varnost podatkov, boljša integracija podatkov, celovit pregled nad delovanjem podjetja, da so zmanjšana podvajanja podatkov v bazi, da omogoča pridobivanje kakovostnejših podatkov in tako tudi boljše odločanje in hitrejše sprejemanje odločitev (Coronel, Morris & Rob, 2013, str. 42–43). DBMS se med sabo nekoliko razlikujejo glede na način delovanja in v sintaksi programskega jezika, ki je prilagojen glede na proizvajalca podatkovne baze kot so Oracle, Microsoft SQL Server, MySQL.

Relacijska baza je fizični model, ki poskrbi za dejansko shranjevanje podatkov in metapodatkov. Shranjevanje podatkov je naslednja faza po logičnem modelu. Načrtovanje baze je ključnega pomena, saj lahko odzivni čas obdelave podatkov ali pridobivanja podatkov bistveno zmanjšamo, če so podatki, postopki kakovostno načrtovani in so tako podatki hitreje na voljo uporabnikom (Lightstone, Teorey & Nadeau, 2007, str. 5).

1.4.1 Relacijska baza

Osnovni objekt relacijske baze je dvodimenzionalna tabela. V prvi vrstici so nanizani atributi, v ostalih vrsticah pa vrednosti za attribute. Osnovna enota vrstice je polje, ki predstavlja vrednost entitete za določen stolpec tabele. Poleg tabele obstajajo še drugi objekti relacijske baze. Nekatere objekti se nanašajo direktno na tabelo, kot so primarni, enostavni, sestavljeni, surogatni in tuji ključ, omejitve na tabeli in indeks. Samostojni objekti so pogled, sekvenca, prožilec, shranjena procedura (v nadaljevanju procedura), ki so v povezi s tabelo.

Bazo se zgradi tako, da se logični model integrira v fizični model s predpisano sintakso, imenovano strukturirani povpraševani jezik (angl. Structured Query Language, v nadaljevanju SQL) in omogoča širok spekter funkcionalnosti. Jezik SQL se tako deli na naslednje kategorije (Fredrick, brez datuma):

- jezik za definiranje podatkov (angl. Data Definition Language, v nadaljevanju DDL), ki omogoča kreiranje, brisanje, spreminjanje objektov, kot so tabele, pogledi, procedure, sprožilci,
- jezik za upravljanje s podatki (angl. Data Manipulation Language, v nadaljevanju DML), ki omogoča vstavljanje, brisanje, spreminjanje, poizvedovanje med podatki, ki so shranjeni v tabelah, pogledih,
- jezik za poizvedovanje nad podatki (angl. Data Query Language), ki vrne podatke iz objektov in hkrati lahko izvede različne matematične operacije nad podatki, jih uredi, združuje,

- jezik za nadzor nad podatki (angl. Data Control Language), ki z dodeljevanjem pravic dostopov oseb nadzoruje podatke in objekte,
- jezik za nadzor nad transakcijami (angl. Transaction Control Language) - transakcije potrdimo ali pa ovržemo in vsebine tabel povrnemo v prejšnjo stanje.

Glede na omenjena dejstva SQL omogoča kreiranje, spreminjanje in brisanje objektov, kot so tabele, pogledi, ključi, indeksi, prožilci, procedure, funkcije, vstavljanje in brisanje podatkov v objektih, poizvedovanje nad podatki in urejanje pravic v podatkovni bazi. Izvajamo lahko selekcijo (izberemo in pregledujemo samo del podatkov), projekcijo (izberemo in pregledujemo samo določena polja), urejanje podatkov, podatke grupiramo v skupine in uporabimo matematične funkcije, kot so vsota, povprečna vrednost, maksimalna vrednost, število enot, uporabimo zanke nad podatki, združevanje dveh ali več tabel po ključu, lahko pridobimo unijo ali razliko tabel (vrstice, ki so v eni tabeli in jih v drugi ni) in še bi lahko naštevali.

Poglejmo lastnosti nekaterih objektov relacijske baze.

Pogled je objekt, ki je vizualno tabela, fizično pa je shranjena samo poizvedba nad fizičnimi tabelami ali drugimi pogledi. Pogled zasede samo prostor za shranjevanje poizvedbe, kar je bistveno manj od shranjevanja podatkov. Nad pogledom lahko izvajamo poizvedbe enako kot nad tabelo, s to razliko, da so poizvedbe nad pogledom časovno bolj potratne, saj je potrebna sprožitev dodatne poizvedbe, ki kreira pogled. Fizična izvedba nad pogledom je identična gnezdenju poizvedb.

Uporabnost pogleda je v tem, da za različne skupine uporabnikov oz. aplikacij pripravimo individualni pogled nad istimi podatki, shranjenimi v eni ali več tabelah oz. pogledih. Pogled vsebuje nabor določenih podatkov, katere lahko grupiramo, uporabimo različne računske operacije, primerjamo, omejimo podatke samo nad določenimi vrsticami tabel oz. uporabimo poljubno poizvedbo, ki ob spremembi podatkov v fizičnih tabelah, brez proženja dodatnih procedur, avtomatsko osveži tudi podatke v pogledu. Pogled je uporaben tudi za kompleksne poizvedbe, saj lahko del poizvedbe nadomestimo s pogledom in poizvedbo poenostavimo (Kroenke & Auer, 2010).

Indeks je objekt, ki lahko bistveno pohitri pridobivanje podatkov iz tabele in ima zaradi tega pomembno vlogo pri postavitvi baze. Lahko ga enačimo s kazalom knjige, kar pomeni, da se vsebina tabele strukturira in shrani sklic glede na enega ali več ključnih stolpcev tabele, ki so pogosto uporabljeni v poizvedbah. Pri uporabi indeksiranih stolpcev v poizvedbi se ta hitreje izvede, saj ni potrebnega pregleda celotne tabele, ampak za to uporabi indeks. Le-ta hrani informacijo, kje se vsebina nahaja. Indeks je identifikator zapisov (angl. record identifiers) (Lightstone, Teorey & Nadeau, 2007, str. 15–28).

Indeks lahko delimo na (Davidson & Moss, 2012, str. 160–170):

- enostavni - sestavljen iz enega stolpca tabele,
- sestavljen - sestavljen iz dveh ali več stolpcev tabele,

- enolični - vse vrednosti so enolične glede na stolpec/ce, ki predstavlja/jo indeks,
- gručast - indeks fizično sortira tabelo glede na vrednosti stolpca/ev, ki predstavlja/jo indeks,
- ne-gručast - indeks je fizično neodvisen od tabele in podatki v tabeli fizično niso sortirani, sam indeks pa je sortiran in hrani informacijo, kje se vsebina nahaja, ločeno od dejanske tabele.

Na eni tabeli lahko postavimo več indeksov, vendar vsak indeks zasede določen prostor in upočasnjuje izvajanje ukazov za posodabljanje vrstic tabele, kot so: vstavi, popravi, briši, zato mora biti njihova uporaba preiščljena. Obstajajo pravila, kdaj ga uporabiti, katere vrste in katere kombinacije indeksov so sploh dovoljene. Pravilo je, da je na vsakem primarnem ključu postavljen enolični indeks. Kandidati za indeks so tudi tuji ključi, po katerih se pogosto združujejo tabele in se ne spreminjajo pogosto ter stolpci, ki pogosto nastopajo v »where« pogoju. Vsaka tabela ima lahko samo en gručasti indeks, saj je fizično lahko sortirana samo po eni kombinaciji. Stolpec/ce je smiselno uporabiti v indeksu, če vsebuje/jo okvirno vsaj 30 % enoličnih vrednosti glede na vrednosti v celotni tabeli (Lightstone, Teorey & Nadeau, 2007, str. 180). Pri obsežnejših obdelavah, pred začetkom posodabljanja vrednosti tabele, je priporočljivo indekse odstraniti in jih po koncu obdelave ponovno vzpostaviti. Tako pohitrimo čas izvajanja posodobitve (Davidson & Moss, 2012, str. 160–170; Lightstone, Teorey & Nadeau, 2007, str. 55–58).

Procedura je program v SQL sintaksi, ki je shranjen v podatkovni bazi. V proceduri so tako shranjeni ukazi, ki se izvedejo z zaporednim proženjem, kot si sledijo ukazi. Procedura podpira standardne SQL ukaze. Ta lahko sprejme vhodne parametre, na katere se sklicuje ob izvajanju, ali vrne izhodne parametre ob koncu izvajanja. Naštujemo lahko kar nekaj prednosti v uporabi procedur. Prednost je enkapsulacija podatkov, da skrijemo in zaščitimo kodo pred zunanjim okoljem in vplivi. Koda je dostopna in se izvaja kot del procesov oz. aplikacije in je nedostopna uporabnikom. Prednost je dinamičnost, saj lahko uporabljamo dinamične stavke z uporabo vhodnih parametrov. Uporaba procedur zagotavlja tudi varnost, saj uporabnikom dodelimo dostop za proženje procedure, ne potrebujejo pa dostopa in pravic do samih virov in tako dodatno zaščitimo podatke pred nenadzorovanimi transakcijami. Prednost je učinkovitost, saj z uporabo parametrov kodo poenostavimo in je večkrat uporabna. V primeru težav, kot je počasno delovanje procedure zaradi povečanja števila podatkov, jo je enostavno prilagoditi in izboljšati, brez posega v procese oz. aplikacijo (Davidson & Moss, 2012, str. 613–627).

Lahko je vgrajena funkcija, ki je že vnaprej določena in jo ne moremo spreminjati ali pa jo sprogramiramo sami, kar lahko imenujemo lastna funkcija. Klic funkcije se izvede znotraj poizvedbe. Funkcija sprejme vhodne parametre ali pa jih nima, kot rezultat pa vedno vrne vrednost. Vgrajene funkcije lahko delimo na tekstovne, matematične, datumske, agregatne, sistemske funkcije, metapodatkovne funkcije itd. Pomembno je, da so vhodni parametri pravega podatkovnega tipa. Vzemimo za primer vgrajeno funkcijo CONCAT, ki združi več nizov v en niz. Tukaj je obvezno, da so vhodni parametri znakovnega tipa, vrednost, ki jo vrne, pa je prav tako znakovna.

Prožilca je procedura, ki se izvede samo ob sprožitvi nekega dogodka. Prožilce običajno uporabimo za izvedbo kompleksnejših pravil nad podatki, avtomatsko proženje ukazov, vzdrževanje revizijskih sledi, nadzor nad transakcijami, primerjava podatkov pred in po transakcijah. Lahko se izvede tako na DML ali DDL ukaze. DML prožilci se prožijo po izvedbi ali pa namesto ukazov: vstavi, spremeni, briši. Na eni tabeli ali pogledu je lahko več prožilcev hkrati. Primer prožilca je, da se ob vstavljanju vrstice v tabelo, avtomatsko sproži tudi sekvenca, ki se avtomatsko povečuje za ena in tako enolično določa vrednost vrstice v tabeli. DDL prožilci se uporabljajo za revizijsko sled, nadzor in upravljanje sprememb nad objekti. Lahko se prožijo po ukazih, kot so kreiraj, briši objekt v podatkovni bazi. Primer DDL prožilca je, da ga postavimo na podatkovno bazo in v primeru, da nekdo briše oz. spreminja objekt, to preprečimo ali vsaj zabeležimo transakcijo in storilca (Ben-Gan, 2012, str. 366 - 370; Masood-Al-Farooq, 2014, str. 131–137).

Particioniranje tabele pomeni, da tabelo razdelimo na manjše, približno uravnotežene dele, imenovane particije, glede na vrednosti podatkov v določenem stolpec v tabeli. Pogost primer delitve tabele je gledanje na časovno komponento, kot je npr. delitev po letih. Prednost pri uporabi particije je, da pri poizvedbah ne dostopamo do celotne tabele, ampak samo do podatkov, ki se nahajajo v particiji. Ko se podatki vstavljajo v tabelo, se avtomatsko razdelijo po particijah glede na delitveni stolpec. Vsaka vrednost lahko pripada natanko eni particiji. Particije so uporabne predvsem pri velikih količinah podatkov, kjer shranjevanje podatkov poteka periodično dlje časa.

Urejanje pravic nad tabelami in ostalimi objekti igra pomembno vlogo pri postavitvi podatkovne baze, saj tako omejimo dostop in pravice uporabnikom in na ta način preprečimo marsikatero nevarnost, kot je izguba podatkov, izvajanje transakcije, ki pokvari podatke, dostop do občutljivih podatkov napooblaščenih oseb. Pravice ureja administrator podatkovne baze, ki je oseba z najvišjimi pravicami v podatkovnih bazah. Le-ta lahko upravlja in izvaja vse DDL ukaze nad objekti. Uporabnik mora imeti urejen dostop do podatkovne baze in ustrezne pravice za delo z objekti. Pravice se ureja na nivoju uporabnika ali vloge. Privzeto že obstajajo vloge, kot so vloge za pisanje, branje, administrator baze, lahko pa jih individualno kreira administrator baze in jim dodeli poljubne pravice. Vsakemu uporabniku določi seznam vlog, ki mu pripadajo. Na ta način je urejanje pravic sistematično, pregledno in enostavno.

Metapodatki so shranjeni v katalogu v podatkovni bazi. Metapodatki opisujejo vsak objekt in njihove lastnosti. Za tabelo imamo opisano strukturo, informacijo o stolpcih, število vrstic, velikosti tabele, opisane omejitve na tabelah, seznam indeksov, kardinalnost tabele z ostalimi tabelami, SQL poizvedbo, ki je bila uporabljena za kreiranje tabele in še bi lahko naštevali. Z metapodatki pridobimo opis vsebine in kopijo kode za postavitve enake podatkovne baze na drugem strežniku.

1.4.2 Normalizacija baze

Normalizacija je pomemben proces fizičnega oblikovanja relacijske baze. Izraz pomeni, da na podlagi pravil organiziramo podatke v bazi tako, da prihaja do čim manj podvajanj podatkov kar nam omogoča lažje vzdrževanje podatkovne baze. Obstajajo različne stopnje normalizacije: prva normalna forma, druga normalna forma, tretja normalna forma, Boyce Codd normalna forma, četrta normalna forma, itd.. Splošno pravilo pravi, da naj bi bila kakovostna relacijska baza normalizirana vsaj do tretje normalne forme.

Postopek normalizacije je, da kompleksne entitetne tipe delimo na enostavnejše entitetne tipe, ki predstavljajo ožji koncept. Postopek poteka do nivoja, da vsaka tabela v bazi predstavlja natanko en koncept, s pripadajočimi atributi za opis koncepta (Davidson & Moss, 2012, str. 130).

Za lažje razumevanje procesa normalizacije si pogledjmo še pomen dveh izrazov: atomična vrednost in funkcionalna odvisnost.

Atomična vrednost pomeni, da je vrednost razstavljena na najmanjši del, ki ima enoten pomen. Pri delitvi vrednosti se vsebinski pomen delov ohranja. Postopek lahko primerjamo z delitvijo molekule na atome. Torej atomične vrednosti ni mogoče uporabiti tako, da bi se sklicevali samo na del vrednosti. Atomična vrednost je lahko sestavljena iz več nizov, če se pri relacijah vedno sklicujemo na celoten niz (Davidson & Moss, 2012, str. 11).

Funkcionalna odvisnost atributov pomeni, da so vrednosti atributov med sabo vedno odvisne. Atribut Y je funkcionalno odvisen od atributa X, če velja, da pri točno določeni vrednosti atributa X dobimo vedno točno določeno enako vrednost atributa Y. Iz tega sledi, da v tabeli ne moremo imeti podvojenih vrednosti glede na atributa X in Y (Kent, 1983).

Proces normalizacije si sledi po stopnjah oz. normalnih formah. Tabela je v prvi normalni formi, če velja, da je točno določeno število atributov, da je v vsakem polju natanko ena atomična vrednost in da lahko določimo ključne attribute, enolični ključ tabele. Če vrednost ni atomična, moramo atribut razdeliti na več atributov, da zagotovimo prvo normalno formo (Ben-Gan, 2012, str. 7).

Poglejmo si primer, ki krši prvo normalno formo. Imamo več atributov enake vsebine, običajno imajo atributi enako osnovno poimenovanje, različna je samo pripona, kot npr. telefonske številke oseb (tel1, tel2, tel3). Osebe imajo lahko različno število telefonskih števil in zato je lahko veliko praznih vrednosti, torej nimajo vse vrstice enakega števila atributov. Vzdrževanje takšne tabele je oteženo. Če moramo popraviti telefonsko številko osebe, moramo vedno pogledati vrednosti vseh treh atributov, ki vsebujejo telefonske številke. Če ima osebe štiri različne telefonske številke, četrte ne moremo zapisati. Rešitev za odpravo praznih vrednosti in lažje vzdrževanje tabele je, da se kreira nova tabela in vsebino atributov telefonskih števil prenese v en atribut tel, kjer bodo vse telefonske številke. Tabela bo vsebovala tudi dodatni atribut, ki označuje vrstni red telefonske številke. Prednost takšne oblike tabele z neomejeno

kardinalnostjo stolpce je tudi v tem, da ima oseba lahko zavedenih neomejeno število telefonskih števil.

Drugo in tretjo normalno formo določimo na podlagi funkcionalnih odvisnosti ključnih in neključnih atributov. Tabela je v drugi normalni formi, ko velja, da je tabela v prvi normalni formi in so neključni atributi v celoti odvisni od ključnih atributov, torej odvisni od primarnega ključa. Kršitev druge normalne oblike lahko nastopi samo, ko je primarni ključ sestavljen iz ključa. Tehnično to pomeni, da z nastopom kršitve prve normalne forme, ko ustvarimo novo tabelo, neključne attribute, ki so delno odvisni od primarnega ključa, skupaj z determinanto prestavimo v novo tabelo oz. tabele. Determinanta je tako nov ključni atribut v novi tabeli. Determinanta ostaja tudi v originalni tabeli zaradi povezljivosti tabel (Kent, 1983).

Tabela je v tretji normalni formi, ko velja, da je tabela v drugi normalni formi in neključni atributi niso funkcionalno odvisni od drugih neključnih atributov. To pomeni, da v tretji normalni formi odpravimo ponavljajoče se vrednosti neključnih atributov in ustvarimo nove tabele. Vsak neključni odvisni atribut prestavimo v svojo tabelo (Kent, 1983).

Boyce Codd normalna forma združi pravila druge in tretje normalne forme. To je v praksi najbolj pogosto uporabljena forma. Tabela je v Boyce Codd normalni formi, ko velja, da je tabela v prvi normalni formi, vsi atributi so odvisni od primarnega ključa in vsaka determinanta je ključ svoje tabele. S tem zagotovimo, da imamo v tabeli izključno funkcionalno odvisnost med ključnimi in neključnimi atributi (Davidson & Moss, 2012, str. 144–152).

Če povzamemo, je prednost normalizacije, da zmanjša funkcionalno odvisnost stolpcev v tabeli in minimizira redundanco. Tako zmanjšamo porabo prostora, poenostavimo posodabljanje, brisanje podatkov in poizvedovanje med podatki. Članek po Wangu, Duju in Lehmannu (2010) je lep pokazatelj poteka in prednosti normalizacije, zmanjševanja redundance in porabe prostora podatkovne baze z višanjem stopnje normalizacije, do tretje normalne forme, na manjših tabelah. V realnosti imamo opravka z bistveno večjimi tabelami in v teh primerih je prostorska učinkovitost še toliko bolj izrazita. Z višanjem stopnje normalizacije se zvišuje potreben čas za izvedbo transakcij nad podatkovno bazo. Normalizacija ustvarja dodatne tabele in tako je povezovanje tabel vedno bolj kompleksno. Normalizacija je optimalna do tistega nivoja, ko je izvajanje transakcij poenostavljeno in prostorsko minimizirano do nivoja, ko je izvedba transakcij še vedno dovolj hitra.

1.5 Kakovost podatkov

Kakovost podatkov je že v samem uvodu predstavljena kot pomemben dejavnik, ki zagotavlja verodostojne informacije za uporabnike. Vsak podatek doseže neko raven kakovosti. Ko govorimo o kakovosti podatkov, imamo v mislih, da so le-ti visoko kakovostni. S kakovostjo podatkov se je že v preteklosti ukvarjalo veliko avtorjev. Tema je aktualna še danes.

1.5.1 Teorije kakovosti podatkov

Kakovost podatkov je odvisna od okolja in namena podatkov, kar je eden izmed razlogov, da obstaja več definicij. Podatki so visoko kakovostni, če zadovoljujejo potrebe uporabnikov in jih lahko uporabimo za načrtovanje in odločanje (Alizamini, Pedram, Alishahi & Badie, 2010). Lahko tudi rečemo, da so podatki kakovostni, če ustrezajo predvideni uporabi, torej da so natančni, pravočasni, ustrezni, popolni, razumljivi in zaupanja vredni (Olson, 2003, str. 24). Kakovost podatkov pomeni, da izboljšamo poslovne procese informacijskega sistema, kjer se podatki nahajajo (Sidi in drugi, 2012).

Kakovost ni odvisna samo od natančnosti oz. točnosti ocen, ampak je odvisna od več dejavnikov in ključno je, da te dejavnike prepoznamo. Dejavnike, ki vplivajo na kakovost, imenujemo dimenzije kakovosti podatkov. Nabor le-teh je odvisen od vsebine podatkov, okolja in podrobnosti delitve. Nekatere izmed dimenzij so skoraj vedno prisotne. To so točnost ocen, veljavnost, popolnost, konsistentnost in pravočasnost (Sebastian-Coleman, 2013, str. 40). V želji, da dosežemo čim višjo kakovost, je pomembno, da izluščimo ključne dimenzije in poznamo njihove definicije. Za lažje razumevanje dimenzij kakovosti in določanje ravni kakovosti na podatkih, moramo poznati ključne situacije oz. probleme, ki negativno vplivajo na kakovost podatkov (tabela 3).

Tabela 3: Razvrstitev problemov, ki vplivajo na kakovost podatkov

Vrsta vira podatkov	Kategorija	Opis
Enojni vir	Podatkovna baza	Pomanjkanje omejitev enoličnosti
		Ni referenčne celovitosti
		Pomanjkanje omejitev na poljih
	Instanca	Napake v sintaksi
		Dvojniki
		Vrednost izven domene
Več virov	Podatkovna baza	Različno oblikovani podatkovni modeli, ni povezljivosti med viri
		Konflikti pri poimenovanju polj
	Instanca	Prekrivajoči podatki
		Vsebinsko neujemanje
		Časovno neujemanje

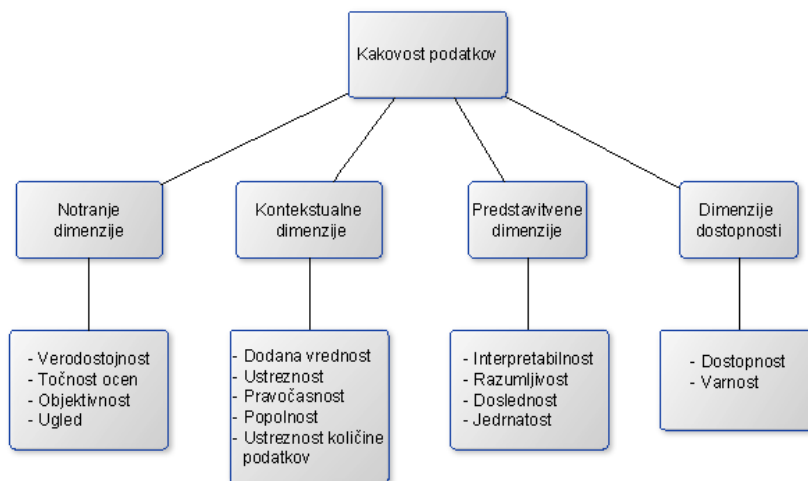
Prerejeno po Man, Wei, Gang & Juntao (2010).

Ena izmed možnih razvrstitev problemov je glede na to, ali skozi proces zajema in obdelave uporabljamo samo en vir ali več virov hkrati. Problemi se lahko nanašajo na tabele podatkovne baze ali pa samo na instanco tabele. Pogosti problemi so pomanjkanje omejitev na tabelah od enoličnosti, referenčne integritete, brez povezljivosti med viri, konflikti pri poimenovanju polj, ko poimenovanje med viri ni poenoteno. Pogoste težave na instanci tabele so napake v sintaksi, dvojniki, vrednosti izven domen polj. Pri povezovanju več virov so težave prekrivajoči se podatki, ko so vrednosti glede na ključ prekrivajoče, ko so prisotne razlike v vsebini (koncepti, definicije), kar meni tudi Jakus (2016, str. 47). Težave nastopijo tudi zaradi razlike v časovnem

obdobju zbiranja virov, kar pomeni, da jih med seboj ne moremo primerjati (Man, Wei, Gang & Juntao, 2010).

Dimenzije kakovosti so medsebojno odvisne in jih lahko razvrstimo po skupinah. Wang in Strong (1996) sta postavila okvir za hierarhično delitev dimenzij, in sicer sta prepoznala petnajst ključnih dimenzij, pomembnih z vidika uporabnikov, in jih razvrstila v štiri skupine (slika 7).

Slika 7: Okvir za delitev dimenzij kakovosti podatkov



Prirejeno po Wang in Strong (1996).

Skupino notranje dimenzije predstavljajo dimenzije, ki se ukvarjajo s kakovostjo, ki je odvisna izključno od samih podatkov in sicer s točnostjo ocen, objektivnostjo, verodostojnostjo in ugledom podatkov. Skupino kontekstualne dimenzije predstavljajo dimenzije, kjer je kakovost odvisna od podatkov v povezavi s celotnim procesom, torej da podatki predstavljajo dodano vrednost za uporabnika, da so podatki ustrezni, popolni, pravočasno na voljo z ustrežno količino. Skupino predstavitvenih dimenzije predstavljajo dimenzije, ki so odvisne od informacijske komunikacijske infrastrukture za prikaz podatkov. Podatki morajo biti razumljivi, dosledni in jedrnat za uporabnika. Skupino dimenzij dostopnosti predstavljajo dimenzije, ki so odvisne od informacijske komunikacijske infrastrukture z vidika dostopnosti in varnosti za uporabnika. Lahko povzamemo, da morajo biti kakovostni podatki sami po sebi dobri, kontekstualno primerni za uporabo, jasno predstavljeni in dostopni za uporabnike (Wang & Strong, 1996).

Dimenzije kakovosti podatkov delimo tudi na dimenzije zunanega in notranjega pogleda. Določene dimenzije se nahajajo tako v eni kot drugi skupini. Dimenzije zunanega pogleda so tiste, ki so odvisne od uporabe informacijskega sistema in podatkov z vidika uporabnikov, notranji pogled pa so dimenzije, ki so odvisne od oblikovanja in funkcionalnosti samega informacijskega sistema. Dimenzije zunanega pogleda povezane s podatki so lahko pravočasnost, ustreznost in jasnost. Dimenziji zunanega pogleda povezane z informacijskim sistemom sta lahko pravočasnost in prilagodljivost. Dimenzije notranjega pogleda v povezavi s

podatki so lahko točnost ocen, pravočasnost, popolnost, dimenzija notranjega pogleda v povezavi z informacijskim sistemom pa je lahko zanesljivost (Wand & Wang, 1996).

Pomen dimenzij, ki jih uporabljajo statistični uradi, so predstavljeni v naslednjem razdelku 1.5.2.

Oceno dimenzij kakovosti se določi na podlagi kombiniranja subjektivnih in objektivnih ocen, kjer je to mogoče. Subjektivne ocene temeljijo na podlagi znanj in izkušenj ključnih deležnikov podjetja. Objektivne ocene lahko temeljijo izključno na podlagi podatkov, brez poznavanja postopkov obdelav, ali na podlagi ocen postopkov obdelav, kjer so vključena poslovna pravila, regulative ali omejitve skozi proces obdelave ali v podatkovnem okolju. Pri objektivnih ocenah se pogosto uporabi razmerje, minimalna oz. maksimalna vrednost ali povprečje. Prednost pri kombiniranju subjektivnih in objektivnih ocen je v tem, da pridobimo oceno iz dveh perspektiv in v primeru neskladja, le-te poskušamo odkriti in odpraviti. Tako je dosežena maksimalna kakovost dimenzije (Pipino, Lee & Wang, 2002).

Postopek za izboljšanje kakovosti podatkov je lahko usmerjen na podatke ali procese. Podatkovno izboljšanje kakovosti dosežemo tako, da spremenimo vrednosti podatkov, procesno izboljšanje pa, da preverimo in prilagodimo proces, ki nadzira podatke ali pa proces nadomestimo z novimi postopki, kar izboljša kakovost podatkov (Sidi in drugi, 2012).

1.5.2 Dimenzije kakovosti podatkov statističnih uradov

Brackstone je eden izmed prvih avtorjev, ki se je ukvarjal s kakovostjo podatkov statističnih. Avtor pravi, da je zaupanje v informacije ključnega pomena za obstoj urada in seveda zaupanja strank v njihovo delovanje in podatke (Brackstone, 1999).

Kot je omenjeno v uvodu, urad meri kakovost po dimenzijah, ki jih določa zaveza ESS. Članice zaveze so Eurostat, nacionalni statistični organi in drugi nacionalni organi, ki proizvajajo statistične podatke (Eurostat, 2021, str. 13). Podatki so tako mednarodno primerljivi med članicami ESS zaveze. Dimenzije kakovosti, ki tako veljajo za vse članice ESS zaveze so ustreznost, točnost ocen, pravočasnost in točnost objave, dostopnost in jasnost, primerljivost in skladnost. Podrobno si pogledjmo definicijo in posebnosti za vsako izmed dimenzij.

Ustreznost podatkov (angl. Relevance) določa ali podatki in uporabljeni koncepti zadovoljujejo potrebe strank (Eurostat, 2003). Torej uporabnikom nudimo tiste podatke, ki jih potrebujejo. Ključno je, da imamo stik z uporabniki, da sledimo njihovim željam in zahtevam glede podatkov in jih poskušamo v čim večji meri zagotoviti. Podatki, ki nimajo končnih uporabnikov so brez vrednosti in po nepotrebnem bremenijo poročevalske enote in celoten proces zbiranja podatkov.

Točnost ocen (angl. Accuracy) določa, da podatki čim boljje odražajo realno stanje. Pogosto točne ocene ne moremo pridobiti, saj bi bilo zagotavljanje vhodnih podatkov predrago ali pa bi bilo nemogoče, da bi dejansko zbrali oz. pridobili vhodne podatke od vseh ciljnih enot.

Dimenzija določa razliko izračunane vrednosti od prave vrednosti, kar imenujemo napako ocene. Drugače povedano to pomeni, da nismo prepričani o njeni točnosti, kar pa ni napačna vrednost (Eurostat, 2021, str. 98). Točnost ocen tako pomeni, da med podatki ni izstopajočih, ekstremnih vrednosti. Najpogostejše matematične funkcije, ki so v pomoč pri merjenju točnosti ocen so varianca, standardna napaka, koeficient variacije, interval zaupanja. Na točnost ocen pri podatkih, pridobljenih z vzorčno raziskavo, vpliva vrsta vzorčenja, pokritost vzorca (razlika med vzorčno in ciljno populacijo), neodgovor, točnost zbranih podatkov, postopki obdelave podatkov, kot so urejanje, vstavljanje in modeliranje podatkov. Enako velja pri administrativnih virih. Na točnost ocen vpliva pokritost vira, manjkajoči podatki v viru in napake v okviru meritev (Eurostat, 2021, str. 99–100).

Možnost za napačni podatek je lahko v neveljavni ali manjkajoči vrednosti, lahko pa je veljavna vrednost napačna ali pa samo napačno predstavljena, kot prikazuje slika 8.

Slika 8: Meja med točno oceno podatka in napačnim podatkom

Veljavna vrednost		Neveljavna vrednost	Manjkajoča vrednost
Pravilna vrednost		Napačna vrednost	
Pravilno predstavljena	Napačno predstavljena		
Točna ocena podatka		Napačen podatek	

Prيرهjeno po Olson (2003, str. 33).

Pravočasnost in točnost objavljanja (angl. Timeliness and Punctuality), kjer pravočasnost določa, da so podatki dostopni ob pravem času, takrat ko jih uporabniki potrebujejo. Ukvarja se z razliko med časovnim obdobjem podatkov, na katerega se nanašajo, in dejanskim datumom prve ali končne objave, ko so ti podatki objavljeni in dostopni uporabnikom. Pravočasnost podatkov je ključna pri gospodarskih odločitvah, sicer podatki nimajo koristi (Brackstone, 1999). Točnost objavljanja se nanaša na časovni zamik objavljenih podatkov od napovedane objave, ki je določena zakonsko ali javno dostopna v uradnem koledarju objav ali pa točnosti objave ni mogoče napovedati, kar je potrebno utemeljiti. Pomembno je, da ne prihaja do zamud pri objavljanju končnih podatkov. S tem zagotovimo, da so podatki objavljeni točno (Eurostat, 2021, str. 141–144; Statistični urad Republike Slovenije, 2017).

Dostopnost in jasnost (angl. Accessibility and clarity), kjer dostopnost opredeljuje kje in kako so dostopni podatki za uporabnike, v kakšni obliki, časovno kdaj so podatki lahko na voljo, kakšna je cenovna politika, kako je z avtorskimi pravicami, kako je z dostopnostjo mikropodatkov in metapodatkov. Jasnost opredeljuje informacijsko okolje, kjer se nahajajo podatki, torej metapodatke in dodatne vsebine, kot so metodološka pojasnila in slikovni prikazi, ki pripomorejo k razumevanju in pravilni interpretaciji pridobljenih podatkov (Eurostat, 2003).

Pomembno je, da so podatki enostavno dostopni uporabnikom in obenem varovani pred nezaželenim dostopanjem oz. zlorabi podatkov. Pomembno je tudi, da so podatki jasno opredeljeni, k čimer pripomorejo metapodatki in ostale informacije. Posledično uporabniki razumejo pridobljene podatke pravilno, torej poznajo njihove koncepte, spremenljivke, definicije in poznajo proces pridobivanja podatkov od zbiranja, obdelave podatkov do objave (Eurostat, 2021, str. 74–75).

Skladnost in primerljivost (angl. Coherence and Comparability), kjer skladnost predstavlja stopnjo primerljivosti podatkov, torej so uporabljeni primerljivi koncepti, spremenljivke in metode za pridobitev podatkov, ki jih uporabimo pri načrtovanju raziskovanja (Brackstone, 1999). Skladnost se osredotoča na različne vire, ki jih lahko med sabo povezujemo in analiziramo. Koncepti iz različnih virov so med seboj primerljivi in jih lahko med kombiniramo. Primerljivost podatkov lahko obravnavamo kot podkategorijo skladnosti. Pristop se osredotoča na primerljivost podatkov: časovno, geografsko ali na drugo domeno kot so spol, dejavnost podjetja, poklic, tip gospodinjstva, itd. (Eurostat, 2003). Govorimo o enakem viru, ki je primerljiv v času, kraju ali drugi domeni. V okviru skladnosti in primerljivosti lahko govorimo tudi o logični in numerični doslednosti, kar pomeni, da ima vsak koncept točno določeno enotno definicijo in da je ista vsebina vedno enako uporabljena. Podatki članic ESS morajo biti skladni in primerljivi tako v času kot v prostoru, da jih lahko med seboj združujemo in analiziramo (Eurostat, 2021, str. 147–150).

1.5.3 Vpliv in izboljšanje kakovosti podatkov

Organizacije, ki se zavedajo pomembnosti kakovosti, si že v politiki organizacije zadajo različne cilje, ki pripomorejo k izboljšanju kakovosti. Eden izmed načinov, ki je sprejet v okviru ESS zaveze, je kodeks ravnanja evropske statistike. V njem so opredeljena načela za vse dimenzije kakovosti, kjer so za vsako načelo določeni ključni kazalnike, ki so merila kakovosti. Cilj organizacije je, da v čim večji meri upošteva kazalnike, ki so opredeljeni v okviru kodeksa. Seljak (2017) v metodološkem priročniku opredeli ključne kazalnike za urad, ki so pokazatelji kakovosti podatkov.

Za doseg ciljev kazalnikov kakovosti je potrebno sistematično in usmerjeno upravljati s kakovostjo. V pomoč so smernice za zagotavljanje kakovosti, ki za vsako fazo statističnega procesa, kot je razvidno iz priloge 1, raziskovanja poda usmeritve, ki jim moramo slediti, da so statistični procesi kakovostni in posledično tudi rezultati obdelav, izhodni podatki (Verbič & Steenvoorden, 2017). Faze so med seboj prepletene, saj se tudi smernice med seboj prepletajo in lahko ima ena smernica vpliv na več dimenzij. V nadaljevanju za vsako dimenzijo kakovosti pogledamo kazalnike in vzrok vpliva na izboljšanje kazalnikov.

Za zagotavljanje **ustreznosti** je ključno, redno posvetovanje z uporabniki, spremljanje in spodbujanje inovacij za izboljšanje statističnih podatkov. Na podlagi analiz potrebe in zahtev uporabnikov se odloči, katere področja in statistike je potrebno spremeniti ali dodati. Preuči se možnost uporabe sekundarnih virov, saj na ta način prispevamo k racionalizaciji pridobivanja

vhodnih podatkov. Redno se spremlja uporabljeno metodologijo področij, torej teoretičnega dela, kot je ciljna populacija, definicije, metoda vzorčenja, obdelava podatkov in praktičnega dela kot je oblika vprašalnika, način zbiranja, postopek obdelave podatkov. Aktivno se vključimo v mednarodne delovne skupine in tako sledimo spremembam, delimo mnenje in izkušnje z ostalimi člani in soustvarjamo metodologijo področja (Statistični urad Republike Slovenije, 2017; Verbič & Steenvoorden, 2017).

V okviru **točnosti ocen** je potrebno podatke skozi celoten proces redno ocenjevati in potrjevati, vzorčne in nevezorčne napake meriti po veljavnih standardih, jih opremiti z metapodatki in stalno analizirati revizije raziskav. Na kakovost točnosti ocen vplivamo že pri izboru enote opazovanja, med pripravo in samim izvajanjem zbiranja in obdelave podatkov. Za izbor enote opazovanja, ko pridobimo podatke iz primarnega vira, je pomembno, da izberemo prave enote z dovolj velikim vzorcem. Izberemo pravi način zbiranja oz. kombinacijo različnih načinov zbiranja prilagojeno ciljni populaciji in metodologiji zbiranja. Vsebina vprašalnika je brez nepotrebnih vsebin, testirana, z ravno pravšnje mero logičnih kontrol, da so odgovori poročevalskih enot čim bolj popolni, obenem pa jih ne odvrnejo od izpolnjevanja vprašalnika. Pri zbiranju podatkov je pomembnem dejavnik tudi sodelovanje s poročevalskimi enotami in ustrezno usposabljanje anketarjev. Posledično lahko pridobimo popolnejše podatke in zmanjšamo neodgovor. Izvedejo se smiselne logične kontrole pri prevzemu podatkov in se po potrebi izvede urejanje podatkov, da se popravijo napačne vrednosti. Pri neodgovoru je pomembno, da izberemo ustrezno metodo uteževanja ali vstavljanja manjkajočih vrednosti podatkov, s čimer popravimo pristranskost rezultatov in izboljšamo ocene podatkov. Vstavljene vrednosti morajo biti čim bolj podobne pravi, neznanim vrednostim, ki ohranjajo osnovno porazdelitev podatkov (Seljak, 2017; Verbič & Steenvoorden, 2017).

V primeru uporabe sekundarnega vira, je pomembno, da preučimo vir, ki mora biti dovolj kakovosten s potrebnimi informacijami. To pomeni, da kljub anomalijam v podatkih, kot so manjkajoče, neveljavne, nepopolne, podvojene vrednosti, pridobimo dovolj ustrezne podatke za izračun statistik. V primeru nepopolnih, napačnih vrednosti jih uredimo. V fazi agregacije, ko iz končnih mikropodatkov izračunamo statistične ocene oz. statistike, je pomembno, da sočasno računamo tudi vzorčno napako, da jo lahko redno analiziramo in v primeru prevelike vzorčne napake pravočasno ukrepamo (povečamo vzorec, poskusimo pridobiti dodatne vire, spremenimo porazdelitev vzorca ali cenilko – funkcija, ki preračuna podatke iz vzorca na statistike). Skozi celoten proces zbiranja in obdelave podatkov je pomembno, da so koraki v čim večji meri avtomatizirani, kar poenoti in poenostavi postopke (Statistični urad Republike Slovenije, 2017; Seljak, 2017; Verbič & Steenvoorden, 2017).

Pravočasnost podatkov mora biti usklajena mednarodno. Zato je potrebno slediti koledarju objav in potrebam uporabnikov, ugotoviti je potrebno, kako pogosto naj bodo sploh objave in če so podatki sprejemljivi in zanesljivi glede točnosti ocen, se lahko objavijo tudi začasni podatki. Vsako zamudo objave glede na določen rok, je potrebno najaviti, razložiti in napovedati nov rok objave. Natančno načrtovanje vseh potrebnih faz statističnega procesa je ključno, da so podatki lahko pravočasno na voljo. Le-to je odvisno od razpoložljivih virov, kot so kadri, finance,

informacijska tehnologija in koledarja objav. Za natančno načrtovanje so zelo pomembne tudi izkušnje. Na pravočasnost ima vpliv tudi pravočasno izvajanje posameznih korakov procesa. To dosežemo s pravočasnim testiranjem vseh faz procesa pred izvedbo. V primeru težav napake odpravimo. Pomembno je, da je vključenih čim več avtomatiziranih postopkov po vseh fazah, ki poenotijo in poenostavijo postopke. Pri objavljanju statistik na podlagi sekundarnih virov moramo imeti sklenjen dogovor o dostopnosti podatkov s točno določenimi roki pridobivanja podatkov, vsebino in morebitnimi spremembami metodologije. Na ta način si zagotovimo, da imamo vir pravočasno na voljo in smo pravočasno obveščeni tudi o morebitnih spremembah (Statistični urad Republike Slovenije, 2017; Verbič & Steenvoorden, 2017).

V okviru **dostopnosti in jasnosti** so statistike in metapodatki predstavljeni in shranjeni v okolju, ki uporablja sodobno informacijsko komunikacijsko tehnologijo, ki je dostopna uporabnikom in omogoča pravilno razlago ter primerljivost podatkov. Na zahtevo uporabnika se pripravijo analize, če je seveda to mogoče, ki so javno dostopne. V raziskovalne namene se lahko uredi dostop do mikropodatkov. Uporabnikom je dostopna metodologija statističnih procesov, uporaba in združevanje administrativnih virov, poročilo o kakovosti rezultatov glede na merila kakovosti (Statistični urad Republike Slovenije, 2017). Informacijska komunikacijska infrastruktura mora omogočati hranjenje podatkov v zakonsko določenih rokih. Pomembno je, da je vključenih čim več avtomatiziranih postopkov po vseh fazah procesa do objave podatkov, ki poenoti in poenostavi postopke, dokumentacijo in shranjevanje metapodatkov. Dokumentacija in metapodatki so dostopni tudi za uporabnika. Sočasno je pomembno, da nam sam proces omogoča avtomatizirano računanje kazalnikov kakovosti, ki služijo v pomoč pri pripravi poročila kakovosti, ki je dostopno tudi za uporabnike. Pri analizi podatkov pazimo, da izluščimo zakonitosti, ki so značilne za opazovani pojav in pravilno analiziramo in interpretiramo rezultate, ki so dostopni tudi za uporabnike. Interpretacija mora biti jasna, razumljiva, točna, objektivna, nepristranska in prilagojena ciljni populaciji (Statistični urad Republike Slovenije, 2017; Verbič & Steenvoorden, 2017).

Statistike so medsebojno **povezane in skladne**, pripravljene na podlagi skupnih konceptov, definicijah, primerljive v razumnem časovnem obdobju, v primeru uporabe različnih virov usklajene in primerljive. V okviru evropskih statističnih sistemov so podatki primerljivi in se redno izmenjujejo. Metodologijo raziskovanj se postavlja v tesnem mednarodnem sodelovanju. V čim večji meri uporabimo standardne koncepte, definicije in tudi klasifikacije. Klasifikacije so hierarhične, če je le mogoče, in objavljene na dovolj podobni ravni, kjer so podatki dovolj natančni in zaupni. V primeru, ko imamo isto ciljno populacijo za isto referenčno obdobje v več raziskavah, uporabimo iste postopke za določitev vzorčnega okvirja in tako pripomoremo k večji skladnosti med podatki. Statistike raziskovanja so primerljive v nekem časovnem obdobju, dokler ne nastopi prelom v časovni seriji. Prelom se zgodi, ko pride do spremembe definicije ali metodologije področja (Eurostat, 2021, str. 285). Cilj je, da so definicije in metodologije čim dlje časa nespremenjene (Statistični urad Republike Slovenije, 2017; Verbič & Steenvoorden, 2017).

1.6 Sinteza spoznanj literature

Temelj informacijske rešitve so poslovna pravila. Le-ta so direktiva, ki vplivajo ali usmerjajo delovanje poslovnih procesov na željeni način in jih ščitijo pred nevarnostmi in grožnjami (Ross, 2003a). Osnove za določitev poslovnih pravil so termini, dejstva in povezave med njimi. Za boljšo vizualno predstavo lahko uporabimo model dejstev, kjer so gradniki ravno termini in povezave med njimi. Odgovor na vprašanje kaj je pravilo, se skriva v vsebini poslovnega pravila. Le-ta morajo biti čim bolj enostavna, točno določena, enako razumljiva za vse deležnike, brez logike korakov procesnega izvajanja in brez sistemskih ukazov CRUD (Ross, 2009). Poznamo več vrst razvrščanja pravil, ki so v pomoč pri pravilnem določanju pravil, kot so glavne skupine omejitev, uveljavitev dejanj in izpeljava. V primeru, ko se na isti termin nanaša več pravil, za predstavitev uporabimo tabelarni način, tabele ali odločitvena drevesa.

Poslovni proces je zbirka povezanih aktivnosti, dogodkov in odločitvenih točk. Aktivnost je dlje časa trajajoče eno ali več opravil skupaj. Dogodek se zgodi v trenutku in odločitvena točka določa nadaljnjo pot izvajanja. Modeliranje procesov je pomembno za boljše razumevanje in vizualno predstavo. Če je osrednji konstrukt proces, pravimo, da je modeliranje procesno usmerjeno (kot je BPMN). Če je osrednji konstrukt objekt, pa je modeliranje objektno usmerjeno.

Razlika med poslovnimi pravili in procesi je v tem, da pravila določajo, kaj naj se zgodi, procesi pa določa, kako naj se zgodi. Pravila se lahko integrirajo že v sam potek izvajanja poslovnega procesa, če gre za enostavna pravila, ki jih je enostavno vključiti, če se ne spreminjajo pogosto ter imajo uporabniki dovolj znanj za vzdrževanje in uporabo ustrezne tehnologije z minimalnimi vložki, in v kolikor isto pravilo ni uporabljeno večkrat (Wang, Sadiq & Indulska, 2014; zur Muehlen, Indulska & Kittel, 2008).

Podatkovna baza je okolje za shranjevanje relacijskih podatkov predstavljenih v obliki dvodimenzionalne tabele. Podatkovno modeliranje se deli na konceptualno, logično in fizično. Vsak korak je pomemben del načrta shranjevanja podatkov. ER model je osnovni koncept entitetnih tipov ter povezav med njimi, atributov in glavnega ključa entitetnega tipa. Vsak termin iz modela dejstev se transformira v entitetni tip. Po potrebi se dodajo entitetni tipi, da se zagotovijo pravilne kardinalnosti povezav med njimi in normalizacija podatkovne baze vsaj do tretje normalne forme.

ER model se transformira v relacijski logični model. Entitetni tipi se transformirajo v relacije, vrstice entitetnega tipa v n-terice relacije, atributi v attribute relacije, povezave v povezave relacije, in glavni ključ je kandidat za primarni ključ relacije. Relacijski model je logična postavitev objektov v bazi. Vsaka relacija ima točno določene attribute, tipe, domene in obveznosti atributov. Relacijam zagotovimo enoličnost podatkov, imenovano primarni ključ. Relaciji, ki sta v povezavi, določimo glavni objekt (imenovan starš), odvisni objekt (otrok), in zagotovimo referenčno integriteto. Odvisna relacija je povezana preko tujega ključa. Z vključitvijo takšnih poslovnih pravil določimo omejitve relacij.

DDL izhod logičnega modela generira fizični model v podatkovni bazi. Relacije se transformirajo v dvodimenzionalne tabele, kjer so atributi predstavljeni kot polja tabele. Če si vzamemo dovolj časa za natančno modeliranje, fizičnih tabel ni potrebno spreminjati. Kreirajo se indeksi za pospešitev izvajanj transakcij nad tabelami, pogledi za vizualno prilagoditev podatkov različnim uporabnikom z isto vsebino in uredijo se pravice nad objekti. S procedurami zagotovimo logično izvajanje korakov, s katerimi kreiramo poslovne procese, ki jih vključimo kot del aplikativne rešitve za avtomatizirano izvajanje. Pomemben objekt relacijskih baz je tudi prožilec, ki se izvede samo ob sprožitvi nekega dogodka. Primer prožilca je izvedba kompleksnejšega pravila nad podatki.

Kakovost je odvisna od okolja in namena podatkov, zato so v uporabi različne dimenzije kakovosti. Statistični uradi uporabljajo usklajene dimenzije zaradi medsebojne primerljivosti podatkov. Zavezani so k spoštovanju načel kodeksa, kjer ima vsako načelo kazalnik za spremljanje kakovosti. Urad ima usmeritve, kako izboljšati kakovost podatkov. Urad prispeva h kakovosti tako, da so podatki ustrezni, točno ocenjeni, pravočasni, točni, dostopni, jasni, skladni in primerljivi. Na kakovost podatkov vplivajo poslovna pravila, ki so umeščena v upravljanje in v informacijski sistem (Man, Wei, Gang & Juntao, 2010). Torej je izrednega pomena, kakšna so poslovna pravila, poslovni procesi in podatkovna baza, gledano skozi proces zbiranja in obdelave podatkov.

Kakovost podatkov je odvisna tudi od oblikovanja informacijskega sistema, torej od oblikovanja procesov, katerih rezultati so podatki za objavo. Možne težave s podatki v informacijskem sistemu so, da pride do izgube informacij, več informacij realnega sveta se preslika v isto stanje, nepopolni podatki, napačni podatki. Za odpravo teh težav moramo ustrezno prilagoditi informacijski sistem (Wand & Wang, 1996).

Na kakovost podatkovne baze vpliva pravilno načrtovanje, ki se začne z določitvijo pravih poslovnih pravil, konceptualnega in logičnega modela. Faze načrtovanja se med sabo prepletajo. Vsaka prispeva svoj delež h kakovostni rešitvi in posledično h kakovosti podatkov.

V poslovni proces že na začetku vključimo poslovna pravila, na podlagi terminov, dejstev, omejitev, uveljavitev dejanj, izpeljav, in ga tako izboljšamo. Z modeliranjem poslovnih procesov odpravimo anomalije že v začetni fazi. Prispevamo tudi k pravilnemu razumevanju procesa vsem ključnim osebam. Priključimo sekundarne vire, če je le mogoče, odpravimo neodgovor in napačne podatke. V fazi podatkovnega modeliranja vključimo čim več omejitev. Tako stalno spremljamo in izboljšujemo dimenziji točnosti ocen in ustreznosti. Z natančnim načrtovanjem in avtomatizacijo procesov prispevamo k dimenziji pravočasnosti in točnosti. Z uporabo sekundarnih virov, ki zagotavljajo potrebne podatke za uporabnika, prispevamo k dimenziji ustreznosti. Z optimizacijo procesa, varovanjem podatkov in vključitvijo metapodatkov prispevamo k dimenziji dostopnosti in jasnosti. Primerljivi koncepti, spremenljivke, metode, primerljivost v času in prostoru prispevajo k dimenziji skladnosti in primerljivosti. Vzpostavimo avtomatizirano merjenje kakovosti podatkov in jih opremimo z metapodatki. Tako stalno merimo

vsako dimenzijo kakovosti, jih izboljšujemo in prispevamo k višji kakovosti podatkov (Seljak, 2017; Verbič & Steenvoorden, 2017).

Torej z izboljšanjem informacijske rešitve, z vidika procesov in podatkov, kjer so ključni vidiki poslovna pravila, procesi in podatkovno modeliranje, prispevamo h kakovosti podatkov (Sidi in drugi, 2012).

2 OVREDNOTENJE OBSTOJEČE REŠITVE ANP

Za ovrednotenje obstoječe rešitve je pomembno, da se seznanimo s celotno metodologijo področja ANP. Poznati moramo vidike in izkušnje vseh vpletenih v obstoječi proces. Proučimo obstoječo informacijsko rešitev od priprave na zbiranje, zbiranje in obdelava podatkov do objave končnih podatkov za uporabnike. Osredotočimo se na področje poslovnih pravil, procesov in podatkovne baze. Poglobljena analiza, znanje in izkušnje s področja informacijskih rešitev ter vključitev ključnih strokovnjakov obstoječe rešitve so temelji za izboljšanje rešitve.

2.1 Metodologija zbiranja

Raziskava ANP je določena v Letnem programu statističnih raziskovanj s strani Eurostata. Zbiranje podatkov za raziskavo ANP poteka četrletno, kot je že omenjeno v uvodu. Izbor enot v vzorec so osebe, ki živijo v zasebnem gospodinjstvu in so stare med 15 in 74 let, ter pripadajoči gospodinjški člani izbrane osebe. Osnova za pripravo vzorčnega okvira je Centralni register prebivalstva (v nadaljevanju CRP). Vzorčni načrt je stratificiran, enostaven, slučajen, enote so izbrane v vzorec glede na tip naselja in statistično regijo. Vzorec je panelni rotacijski, torej so izbrane osebe v vzorcu eno leto in pol oz. šest valov, kjer pomeni vsako četrletje en val. Osebe v izbranem obdobju odgovarjajo na anketo petkrat in sicer v prvih treh četrletjih oz. valovih, potem en val počivajo in odgovarjajo še v zadnjem valu (Rutar & Tomažič, 2020).

Zbrani podatki so kombinacija podatkov pridobljenih iz terenskega anketiranja, telefonskega anketiranja in iz sekundarnih virov, kot so administrativne zbirke, ter podatkov drugih statističnih področij urada. Izbrane osebe se povabi k sodelovanju z obvestilnim pismom, kjer je vsebina pisma prilagojena glede na način anketiranja in vrsto anketiranja za prvo, ponovljeno ali zadnje anketiranje. Prvo anketiranje osebe poteka vedno na terenu. Pri ponovljenem anketiranju se osebe, od katerih smo pridobili telefonske številke, anketira po telefonu, ostale osebe ponovno obišče terenski anketar. Pri ponovljenem anketiranju se določeni odgovori prevzamejo iz predhodnega anketiranja. Obdobje opazovanja, na katerega se nanašajo odgovori anketiranja, je referenčni teden. Vsako četrletje ima 13 referenčnih tednov. Izbrane enote so znotraj četrletja enakomerno porazdeljene po referenčnih tednih.

Vprašalnik zajema vsebinske sklope o demografskih podatkih izbrane osebe in o vseh članih gospodinjstva izbrane osebe. To so podatki o priselitvah glede na referenčni teden v zvezi z opravljanjem dela, o delovnem času, o dodatnih opravljenih delih, o preteklih delovnih izkušnjah,

o iskanju del, o izobrazbi in izobraževanju, o samo-opredelitvenem statusu in o zdravstvenem stanju osebe. Vprašalnik vsebuje poseben sklop, ki se vsako leto spremeni in se lahko ponovi vsakih nekaj let, ni pa nujno (Tomažič & Zaletelj, 2021).

Zasebno gospodinjstvo je vsaka družinska ali druga skupnost oseb, ki stanujejo skupaj, se skupaj prehranjujejo in skupaj porabljajo dohodke za kritje življenjskih potreb. Prav tako zajema osebe, ki so odsotne med tednom ali imajo daljšo odsotnost, vendar se vrnejo k družini in imajo namen živeti na naslovu vsaj eno leto. Brezposelna oseba je oseba stara od 15 do 74 let, ki v referenčnem tednu ni delala, torej ni opravila nobene ure za plačilo. Sem spadajo tudi osebe, ki aktivno iščejo delo vsaj v obdobju štirih tednov do referenčnega tedna in so pripravljene sprejeti delo vsaj v dveh tednih po koncu referenčnega obdobja. Brezposelna oseba je tudi tista, ki je že našla delo, vendar ga bo začela opravljati v treh mesecih po koncu referenčnega tedna (Tomažič & Zaletelj, 2021).

Objava podatkov je mesečna, četrtna in letna. Objavlja se v podatkovni bazi SiStat ali na spletni strani Eurostata in mednarodne Organizacije za gospodarsko sodelovanje in razvoj (angl. Organisation for Economic Co-operation and Development – OECD). Četrtno in letno se objavi število oz. delež aktivnih in neaktivnih prebivalcev po različnih dimenzijah, mesečno pa stopnja anketne brezposelnosti kot eksperimentalna statistika (Tomažič & Zaletelj, 2021).

2.2 Analiza stanja z vidika poslovnih pravil, procesov in podatkovnega modeliranja

Za analizo stanja smo preučili obstoječe metodološko gradivo, tehnično in uporabniško dokumentacijo. Opravimo pogovore z vsemi vpletenimi v obstoječi proces na uradu, ki so vsebinski metodologi raziskovanja, razvijalci informacijskih rešitev in infrastrukture, strokovnjaki za standarde, strokovnjaki za metodologijo, strokovnjaki za nadzor nad zbiranjem podatkov oseb in gospodinjstev in anketarji na terenu ter telefonu. Na podlagi pogovorov ugotovimo, da se procesno tehnično področje ni spremenilo že več kot deset let. Tehnično se je vsako leto prilagodil vprašalnik le v sklopu posebnega dela vprašanj. Ostalih postopkov od zbiranja do obdelave in objave niso spreminjali. Obstoječe dokumentacije je izredno malo, obstaja popis metodologije in podatkovni tok obstoječega procesa. Dokumentacija s popisi poslovnih pravil ali prikaz pravil v obliki modela ne obstaja. Pogrešam tudi vizualni prikaz poslovnih procesov v modelu, kot je BPMN konceptualnega ali logičnega modela.

2.2.1 Analiza z vidika poslovnih pravil

Analiza je podana izključno na podlagi pogovorov z vsemi vpletenimi. Glede na dejstva prepoznamo ključne termine, kot so izbrana oseba, anketar, kontrolor, vprašalnik, obvestilno pismo, referenčni teden, status izbrane osebe, odgovor, itd. Katalog terminov z opisi konceptov ne obstaja. Ugotovimo, da vsi zaposleni ne razumejo točnega pomena terminov. Npr. termin »odgovor« razvijalec razume kot zbrani, surovi podatki, med katere spadajo tudi občutljivi podatki, za izbrano osebo in člane njegovega gospodinjstva. Vsebinski metodolog razume

omenjeni termin kot zbrani, surovi podatki, brez občutljivih. Kot smo že omenili, model dejstev ne obstaja.

Poslovna pravila, s katerimi se seznanimo na podlagi pogovorov, povežemo z določenimi termini. Za nekatere termine je potrebno ponoviti pogovore, da lahko prepoznamo pravilo, ki naj bi bilo v veljavi. Za primer si pogledjmo primer poslovnega pravila za izbrano osebo in anketarja.

Za izbrano osebo mora veljati vse od naštetega:

- oseba je v registru CRP,
- oseba je stara nad 15 do 74 let,
- oseba je izbrana v vzorec prvič,
- izbrana oseba iz 2., 3., 5. ali 6. vala je ustrezno odgovarjala v zadnjem preteklem anketiranju.

Terenski anketar anketira izbrano osebo, za katero mora veljati nekaj od naštetega:

- izbrana oseba je izbrana v vzorec prvič,
- izbrana oseba iz 2., 3., 5. ali 6. vala je ustrezno odgovarjala v zadnjem preteklem anketiranju in ni posredovala telefonske številke.

Telefonski anketar anketira izbrano osebo, za katero mora veljati vse od naštetega:

- izbrana oseba iz 2., 3., 5. ali 6. vala je ustrezno odgovarjala v zadnjem preteklem anketiranju,
- izbrana oseba je posredovala telefonsko številko.

Pravila za termin referenčni teden določajo začetek in konec referenčnega tedna ter vsebino poslanega pisma. Začetek in konec anketiranja ni točno določen. To pomeni, da ni jasno definirano pravilo, v katerem obdobju naj bo anketiranje osebe zaključeno.

Pri vrsti obvestilnega pisma je točno določena vrsta pisma za prvo anketiranje in zadnje anketiranje. Za ponovljeno anketiranje ne moremo določiti pravil, saj se le ta sproti prilagajajo.

Poglejmo si še primer pravila za odgovor in občutljive podatke. Odgovor mora vsebovati izključno neobčutljive podatke. Vsakemu gospodinjstvu mora pripadati toliko odgovorov, kolikor je članov tega gospodinjstva. Vsakemu odgovoru mora pripadati status izbrane osebe. Za občutljive podatke je določeno pravilo, da se morajo občutljivi podatki odstraniti v varno okolje in pravilo, da mora vsak občutljivi podatek pripadati enoličnemu osebnemu identifikatorju.

2.2.2 Analiza z vidika procesov

Poglejmo si, kako procesno poteka zbiranje podatkov za področje ANP. Vprašalnik za zbiranje se ob prehodu na novo leto pregleda s strani vsebine in standardov in prilagodi vsaj poseben sklop vprašanj, ki se zaradi uredbe spreminja vsako leto. Vprašalnik je razvit v okolju Blaise, ki je namensko orodje za razvoj le-teh. Anketiranje se izvaja četrtletno. V okviru enega leta se

vprašalnik med četrtletji ne spreminja. Strokovnjak za metodologijo pripravi vzorec za prvi val. To so nove izbrane osebe s podatki o naslovih, ki se jih odloži na datotečni strežnik. Razvijalec vprašalnika pripravi skupni adresar. To je seznam oseb za anketiranje z naslovi. Vsebuje tudi telefonske številke, v kolikor so bile le-te pridobljene pri preteklih zbiranj. Razvijalec s pomočjo arhivskih datotek adresarja, zbranih podatkov preteklih zbiranj in seveda novega vzorca, sestavi adresar za zbiranje novega četrtletja. Če povzamemo, v adresarju so osebe iz novega vzorca in osebe iz preteklih zbiranj, katere so ustrezne za anketiranje, kar pomeni, da so v ustreznih predhodnih valovih odgovarjale na anketna vprašanja. Postopek priprave adresarja poteka ročno, s povezovanjem teh datotek in naknadnim izločanjem oseb za ponovljeno anketiranje, ki v tekočem četrtletju počivajo. Osebam za ponovljeno anketiranje se ne osvežuje naslov prebivališča. Adresar se odloži na datotečni strežnik. Pripravijo se sezname, ki so ločeni glede na vrsto obvestilnega pisma, ki služi za pripravo pisem, tiskanje in pošiljanje. Pripravijo se sezname oseb glede na anketarje, ki bodo anketirali na terenu. Vsi sezname se odložijo na datotečni strežnik. Osebam za ponovljeno anketiranje se pripravi določen nabor vprašanj in odgovore iz zadnjega poročanega vprašalnika. Ti olajšajo in pohitrijo ponovljeno anketiranje, saj omenjenih vprašanj ne postavljamo več oz. samo preverimo, če so odgovori še vedno ustrezni. Upravljanje anket na terenu in telefonu krmilijo s sistemom v Blaisu. Po zbranih podatkih za prvih osem tednov referenčnega tedna v četrtletju se pripravijo izhodi iz surovih podatkov iz Blaisa. Izhodi so v tekstovni obliki in se jih odloži na datotečni strežnik. Ločeno se pripravijo izhodi iz terenskega in telefonskega anketiranja.

Strokovnjak za metodologijo združi surove podatke in pri tem izloči občutljive podatke. Pripravi dva izhoda. V enem izhodu so podatki o gospodinjstvih in v drugem podatki o osebah. Izhodi surovih podatkov so pripravljene v okolju SAS. Pripravi se seznam za ročno šifriranje poklicev in dejavnosti glede na standardno klasifikacijo na podlagi pridobljenih opisov iz surovih podatkov. Po končanem ročnem šifriranju se ti podatki pripojijo podatkom o osebah in se pripravi končni izhod surovih podatkov v SAS okolju. Prevzamejo se še podatki ostalih tednov do trinajstega referenčnega tedna in ponovijo se postopki prevzema surovih podatkov. Podatki se morajo prevzeti v dveh delih zaradi ponovljenih anketirancev. Adresar za naslednje četrtletje moramo imeti namreč na voljo že preden zaključimo z zbiranjem podatkov za tekoče četrtletje.

Sledi imputacija podatkov tudi v okolju SAS. Postopek je prilagojen na področje. Koda je avtomatizirana. Minimalne spremembe so potrebne le ob prehodu na novo leto. Proženje imputacije izvaja razvijalec v SAS okolju. Rezultat imputacije so končni mikropodatki v SAS okolju, pripravljene za agregacijo in tabelacijo. Na podlagi končnih mikropodatkov strokovnjak za metodologijo izračuna uteži, ki jih odloži na datotečni strežnik. Agregacija in tabelacija se izvaja avtomatsko na podlagi shranjenih procesov na strežniku. To proži vsebinski metodolog raziskovanja preko shranjene procedure, ki se izvaja na strežniku. Pomoč in prilagoditev programov ob prehodu na novo leto izvaja SAS programer. Urejene so tudi pravice, kdo lahko izvaja, kdo lahko spreminja proceduro. Objava podatkov v podatkovni bazi SiStat in na spletni strani se izvaja po pravilih in postopkih, ki veljajo na uradu. Mikropodatki se posredujejo tudi na Eurostat po sistemu za izmenjavo podatkov.

Enkrat letno se zbranim podatkom na ravni oseb, na podlagi enoličnih osebnih identifikatorjev oseb, pripojijo tudi dohodninski podatki. Ker imamo ta podatek samo osebe izbrane v vzorcu, za ostale družinske člane pa tega podatka nimamo, ga pripeljemo na podlagi drugih zbranih osebnih podatkov. Uporablja se splošna rešitev za iskanje enoličnih identifikatorjev s prilagojenimi vhodnimi podatki. Postopek izvaja SAS razvijalec.

2.2.3 Analiza z vidika podatkovne baze

Shranjevanje podatkov poteka izključno na datotečnem strežniku na dveh lokacijah z omejenim dostopom. Seznam oseb, ki imajo dodeljene določene vrste dostopov do obeh lokacij datotečnega strežnika, je viden v interni spletni aplikaciji. Vse osebe imajo pravice branja in pisanja. Lokacija, kjer so shranjeni vzorci, adresar, sezname za obvestilna pisma, pomožni sezname za lažje komuniciranje z anketarji, podatki za predhodnike in surovi podatki, ima posebno označbo, da gre za občutljive podatke. Do te lokacije ima dostop majhno število ljudi. Na drugem strežniku so shranjeni ostali podatki, kot so surovi podatki brez občutljivih podatkov, podatki pred in po iskanju skd, skp, končni mikropodatki in uteži. Na isti lokaciji so shranjeni tudi izhodi tabelacij. Končni mikropodatki so shranjeni v formatu, ki je prilagojen za branje v okolju SAS. Izhodne datoteke se vsebinsko lahko tudi podvajajo. Vsebina datotečnih strežnikov se arhivira.

2.2.4 Ugotovitve na podlagi analize

Analiza z vidika poslovnih pravil je pokazala, da manjka dokumentacija in model dejstev, na katerih so določena poslovna pravila in so osnova za poslovne procese in podatkovno modeliranje. Analiza je opravljena izključno na podlagi pogovorov s ključnimi osebami. Brez kataloga terminov so termini lahko napačno razumljeni. Ugotovimo, da za ključne termine pravila sicer imamo, vendar so pogosto nepopolna.

Z vidika poslovnih procesov lahko podamo kar nekaj ugotovitev. Prednosti orodja Blaise so, da pokriva razvoj vprašalnika za različne načine anketiranja, ga vizualno prilagaja napravi s katero se izvaja anketiranje, omogoča veliko funkcionalnosti glede oblikovanja, povezuje vprašalnik z relacijskimi bazami in podpira izvajanje anketiranja na različnih operacijskih sistemih. V proces so vključene ključne osebe. Upoštevajo se standardi in usmeritve urada. Ob prehodu na novo leto se ob večjih spremembah vprašalnik tudi kognitivno testira in po potrebi nadgradi, tako da je vprašalnik čim bolj kakovosten.

Pri postopku priprave adresarja je veliko ročnega dela in možnosti napak. Podatki so shranjeni na datotečnih strežnik, kjer imajo vsi pravico pisanja, kar ni ustrezno. Lahko se zgodi, da se vrednosti pokvarijo, podatki zbrisajo itd. Postopek priprave adresarja je odvisen od ene osebe, kar predstavlja veliko tveganje. Težava je, da se za osebe za ponovno anketiranje ne osvežuje naslova prebivališča. Ker so osebe v vzorcu eno leto in pol, lahko v tem obdobju pride do selitve izbrane osebe ali nastopi kakšen drug dogodek, ki lahko vpliva na to, da osebe ni več na istem naslovu ali da ni več ustrezna za anketiranje. S podatki o predhodnikih se izvajanje ankete pohitri,

kar pozitivno vpliva na odgovor anketirancev. Slabost pa je, da se predhodniki pripravijo ročno iz datotek, pri čemer je velika možnost nastanka napake. Slabost je, da se sezname za obvestilna pisma in anketirance pripravijo ročno, podobno se ročno odložijo surovi podatki brez osebnih podatkov.

Zbrani surovi podatki se iz okolja Blaise zapišejo v tekstovno datoteko na datotečni strežnik. Dostop do podatkov je omejen, vendar za vse osebe z enakimi pravicami. Slabost v tem je, da lahko osebe nenadzorovano pokvarijo podatke. Prav tako ni urejena sledljivost dostopanja do podatkov. Obdelava podatkov poteka v okolju SAS. Pri izvedbi imputacij je prednost v tem, da se izvaja s splošnimi programi, le da jih proži razvijalec. Težava je v tem, da so postopki odvisni od različnih oseb. Prednost pri agregaciji in tabelaciji je, da so postopki avtomatizirani in jo proži vsebinski metodolog preko shranjene procedure. Z vidika varnosti je to ustrezno, saj je dostop do procedure, katero se proži na strežniku, urejen s pravicami za proženje oz. spreminjanje procedure. Slabost skozi celoten proces zbiranja in obdelave podatkov je, da se večinoma izvaja ročno. Postopki so odvisni od posameznikov, kar zahteva veliko medsebojnega usklajevanja. V primeru nepredvidenih odsotnosti sodelujočih, je osebo težko nadomestiti, še zlasti pri pripravi adresarja.

Z vidika podatkovne baze ugotavljamo, da se podatki shranjujejo izključno na datotečnih strežnik. Podatkovne baze torej ni. To je z vidika nadzora, hitrosti pridobivanja podatkov, sočasnega dostopa do podatkov, integracije, redundance, slabost obstoječe rešitve. Varovanje podatkov je sicer delno urejeno, slabost pri tem pa je, da imajo vsi uporabniki enake pravice ne glede na njihovo dejansko potrebo. Razvijalec podatkovne baze bi moral imeti pravice branja, pisanja in kreiranja objektov, medtem ko bi moral vsebinski metodolog dostopati samo do nekaterih objektov za branje in pisanje, do nekaterih pa zadostuje zgolj pravica branja. Na ta način bi si dodatno zagotovili varovanje podatkov.

3 PREDLOG NOVE REŠITVE ANP

Na podlagi pridobljenega znanja teorije, metodologije področja, analize stanja z vidika poslovnih pravil, procesov, podatkovnega modeliranja in baze, znanja in izkušenj strokovnjakov, podamo predlog nove rešitve, kjer v čim večji meri upoštevamo teorijo s ciljem, da prispevamo h kakovosti rešitve in podatkov. Temelj dobre rešitve so poslovna pravila in zato je ključno, da poznamo teorijo poslovnih pravil. Le-te vključimo v poslovne procese, podatkovno modeliranje in okolje za shranjevanje podatkov.

3.1 Določitev poslovnih pravil

Na podlagi dogodkov in stvari določimo poslovna pravila, ki se izključno nanašajo na pripravo in izvedbo procesov za pridobivanje podatkov raziskave ANP. Za določitev poslovnih pravil vključimo ključne osebe, kot so vsebinski metodolog, ki pokriva vsebino ankete, strokovnjak za nadzor nad zbiranjem, ki skrbi za fizično pripravo in nadzor nad zbiranjem, strokovnjak za

standarde, ki nadzoruje in vpeljuje standardne postopke in strokovnjak za metodologijo, katerega naloga je, da skrbi in nudi pomoč pri vzpostavitvi metodologije, od priprave na zbiranje, pridobivanja podatkov in skozi celoten proces obdelave podatkov do končnih rezultatov.

Poslovna pravila so temelj dobre rešitve. Glede na spoznanja teorije, moramo najprej določiti termine in dejstva, ki jih povezujejo med sabo. Za področje ANP opredelimo ključne termine, ki so izbrana oseba, anketar, kontrolor, vprašalnik, obvestilno pismo, referenčni teden, odgovor, status izbrane osebe, osebni odgovor, neosebni odgovor, itd.. Za pravilno razumevanje terminov vseh deležnikov, jih zapišemo v katalog skupaj z njihovimi koncepti.

Poglejmo definicije nekaterih terminov. Pomen termina »izbrana oseba« pomeni osebo, ki je v centralnem registru prebivalcev, ki ustreza pogoju ustreznosti za anketiranje in je izbrana v vzorec. Termin »vprašalnik« je razvit vprašalnik s seznama vprašanj za telefonsko ali terensko anketiranje v računalniško podprtem Blaise okolju, na katerega odgovori izbrana oseba. Termin »anketar« je oseba, ki izvaja anketiranje izbrane osebe in odgovore zapisuje v vprašalnik. Termin »anketar« se deli na terenskega, ki anketira izbrane osebe na terenu ali telefonskega, ki dela v telefonskem studiu, in anketira izbrane osebe preko telefona. Termin »kontrolor« je strokovnjak za pripravo in nadzor nad delom telefonskih kot terenskih anketarjev. Termin »odgovor« so zbrani surovi podatki, na katere odgovori izbrana oseba o vseh članih gospodinjstva, kjer ta oseba živi. Termin »osebni odgovor« pomeni občutljive surove podatke, do katerih mora biti urejen omejen dostop in sledljivost podatkov. Termin »urejeni odgovor stalnih vprašanj« predstavlja po potrebi urejene in vstavljene zbrane podatke, ki so v anketi stalno na voljo in niso občutljivi ter so primerni za izračun statistike.

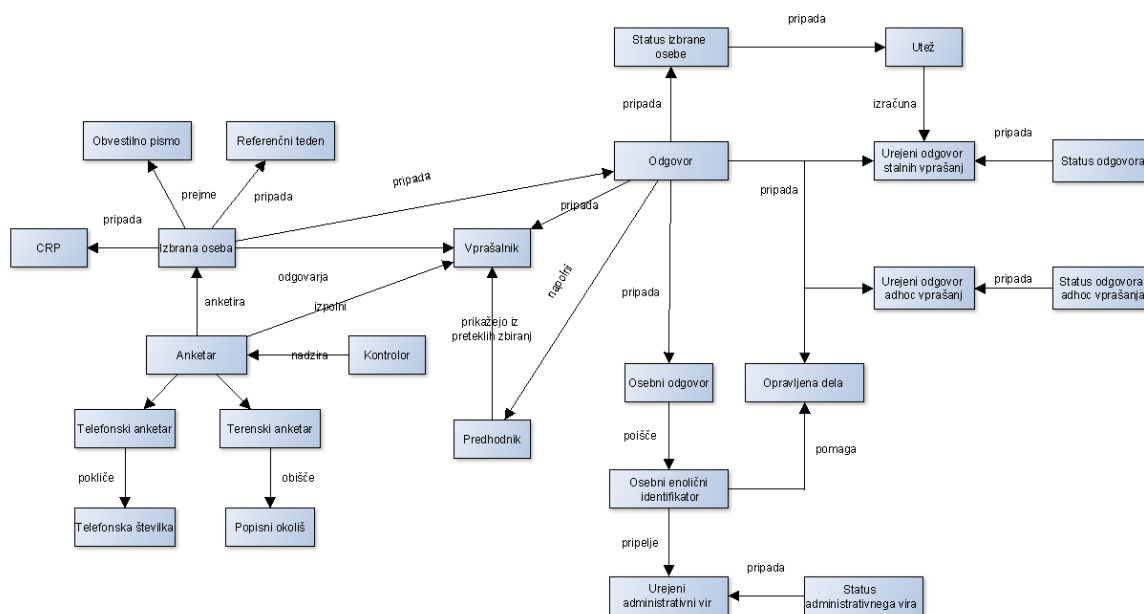
Za lažje prepoznavanje povezav med termini in v izogib podvajanju dejstev, izrišemo model dejstev, kar prikazuje slika 9.

Primeri dejstev so lahko:

- Izbrana oseba prejme obvestilno pismo.
- Izbrana oseba pripada registru CRP.
- Izbrana oseba odgovarja na vprašalnik.
- Kontrolor nadzira anketarja.
- Anketar anketira izbrano osebo.

Na podlagi modela dejstev lažje določimo poslovna pravila, ki odgovorijo na vprašanje, kaj je dano pravilo. Poglejmo si nekaj primerov pravil skozi celoten proces, od priprave na zbiranje do končnih urejenih podatkov, pripravljenih za agregacijo in tabelacijo. Pravila razvrstimo v kategorije, s čimer prispevamo, da so natančno določena in pravilno razumljiva. Za boljšo vizualno predstavo izberemo čim bolj primerno obliko, ki je odvisna od lastnosti pravil. Osredotočimo se na oblike, kot so enostavna poved, tabelarna oblika, tabele ali drevesa.

Slika 9: Model dejstev nove rešitve



Vir: lastno delo.

Primer pravil, ki veljajo za kontrolorja. Kontrolor mora določiti popisni okoliš terenskim anketarjem. Kontrolor mora nadzorovati anketarje pri anketiranju. Omenjeni pravili sta predstavljena v enostavni povedi s poudarkom na zahtevi z besedo »mora«, kot to določajo smernice za določanje pravil. Pravili spadata v skupino omejitev oz. zavračevalcev. V skupino omejitve spadajo tudi pravila za izbor izbranih oseb in terenskih oz. telefonskih anketarjev. Pravila so prikazana v tabelarni obliki, saj se več pravil nanaša na isti termin. Na ta način zagotovimo, da so pravila enostavna in pravilno razumljiva.

Poglejmo si pravila za izbor izbranih oseb in anketarjev. Za izbrano osebo mora veljati vse od naštetega:

- oseba je v registru CRP,
- oseba je stara od 15 do 74 let,
- oseba živi v zasebnem gospodinjstvu,
- oseba je izbrana v vzorec prvič,
- izbrana oseba iz 2., 5., ali 6. vala je ustrezno odgovarjala v zadnjem preteklem anketiranju.

Terenski anketar anketira izbrano osebo, za katero mora veljati nekaj od naštetega:

- izbrana oseba je izbrana v vzorec prvič,
- izbrana oseba iz 2., 5., ali 6. vala je ustrezno odgovarjala v zadnjem preteklem anketiranju in ni posredovala telefonske številke.

Telefonski anketar anketira izbrano osebo, za katero mora veljati vse od naštetega:

- izbrana oseba iz 2., 5., ali 6. vala je ustrezno odgovarjala v zadnjem preteklem anketiranju,
- izbrana oseba je posredovala telefonsko številko.

Pravila, ki se nanašajo na entitetni tip referenčni teden, spadajo v skupino proizvajalcev oz. izpeljave, saj se datumi pošiljanja pisem, obdobje začetka in konca anketiranja, določi glede na začetek referenčnega tedna po predpisani formuli. Pravilo pravi, da se pisma pošljejo na isti dan, kot je začetek referenčnega tedna, anketiranje se začne čez 7 dni, konec anektiranja je 27 dni od začetka referenčnega tedna.

Za določitev vrste obvestilnega pisma si pomagamo z odločitveno tabelo z enim pravilom v eni vrstici, kjer vrstni red pravil ni pomemben in je stil uporaben za razvijalce (tabela 4). Ker imamo samo dva dejavnika, prvi z dvema primerkoma, drugi pa z največ šestimi primerki, hitro opazimo podvajanje pravil. Vsaka vrstica predstavlja svoje pravilo, s katerim določimo vsebino obvestilnega pisma. Pravilo v prvi vrstici se glasi: vsebino obvestilnega pisma za novi teren, dobijo izbrane osebe, ki so iz prvega vala in nimajo telefonske številke.

Tabela 4: Odločitvena tabela eno pravilo v eni vrstici

Ali obstaja telefonska številka?	Val izbrane osebe	Vrsta obvestilnega pisma
ne	prvi	novi terensko
ne	drugi	ponovljeni terensko
ne	tretji	ni obvestilnega pisma
ne	četrti	ni obvestilnega pisma
ne	peti	ponovljeni terensko
ne	šesti	ponovljeni zadnjič terensko
da	drugi	ponovljeni telefonsko
da	tretji	ni obvestilnega pisma
da	četrti	ni obvestilnega pisma
da	peti	ponovljeni telefonsko
da	šesti	ponovljeni zadnjič telefonsko

Vir: lastno delo.

Za določitev pravil nad pridobljenim neobčutljivim odgovorom izbrane osebe, kjer se več pravil nanaša na isti termin, si odgovorimo na vprašanje »Katere postopke naj izvedemo nad odgovorom z neobčutljivimi podatki?«. V tem primeru si pomagamo z matričnim stilom križišč, kjer razmislimo, ali je odgovor izbrane osebe glede lokacije in vrsto opravljenega dela, ali je odgovor na adhoc vprašanje, kot prikazuje tabela 5. Vsaka celica predstavlja odločitev glede na podana dejstva.

Poglejmo še primer prožilca, ki pripada skupini uveljavitev dejanj oz. projektorja. Pravilo za termin »status odgovora« pravi, če izbrišemo urejen podatek odgovora stalnih vprašanj, moramo zbrisati tudi pripadajoče statute odgovora, ki se med seboj povezujejo glede na časovno točko, šifro osebe gospodinjstva in status zapisa.

Za vsak termin imamo določena ključna pravila skozi celoten proces od priprave, zbiranja do obdelave podatkov. To je osnova za vzpostavitev poslovnih procesov in konceptualnega modela.

Tabela 5: Odločitvena tabela stil križišč

		Ali je odgovor na adhoc vprašanje?	
		da	ne
Ali je odgovor glede lokacije in vrsti opravljenega dela?	da	razdeli odgovor na odgovor stalnih, odgovor adhoc ali opravljena dela	klasificiraj opise na standardno klasifikacijo
	ne	izvedi urejanje adhoc vprašanja	Izvedi urejanje stalnega vprašanja

Vir: lastno delo.

3.2 Določitev procesov

Temeljni proces urada je zagotavljanje statističnih podatkov zunanjim deležnikom. Vsebinska področja aktivno spremljajo dogajanje drugih statističnih uradov na trgu in so odgovorna za zagotavljanje podatkov zunanjim deležnikom. Podporo temeljnim procesom zagotavljajo infrastrukturna področja, kot so metodologija in standardi, zbiranje podatkov, objavljane in komuniciranje in razvoj informacijskih rešitev ter infrastrukture.

Želimo si pravilno razumljivost procesov vseh deležnikov, od uporabnikov do razvijalcev, zato procese modeliramo. Delovanje urada je procesno usmerjeno, zato je za prikaz aktivnosti, vloge akterjev in ciljev najbolj primeren standard BPMN. Procesni model je sestavljen iz dveh plavalnih bazenov urada in respondenta, ki predstavlja izbrane osebe za anketiranje. Povezavo aktivnosti urada in respondentov uvrščamo med kolaboracije, saj gre za povezovanje urada z zunanjimi deležniki, ki so dajalci podatkov. Urad je razdeljen na plavalne steze, ki predstavljajo notranja področja organizacijske strukture. Ključni akterji so vsebinski metodolog, ki je nosilec vsebinskega področja, metodologija in standardi, IT, kontrolor in anketar.

Procesni model je sestavljen iz glavnega modela, kot prikazuje slika 10, in dveh podprocesov. Prvi je za pripravo adresarja in podatkov za predhodno izpolnjen vprašalnik anketirancev, ki so v preteklosti že odgovarjali na vprašanja vprašalnika. Drugi podproces je sestavljen iz aktivnosti za obdelavo zbranih podatkov. Pri postavitvi procesov upoštevamo poslovna pravila, ki smo jih prepoznali in jih lahko upoštevamo. Le-ta določajo sprožitveni dogodek, aktivnosti akterjev in razvejišča. V nadaljevanju sledi opis procesa nove rešitve in nekaj primerkov povezav med poslovnimi pravili in procesi, kar prikazuje tabela 6.

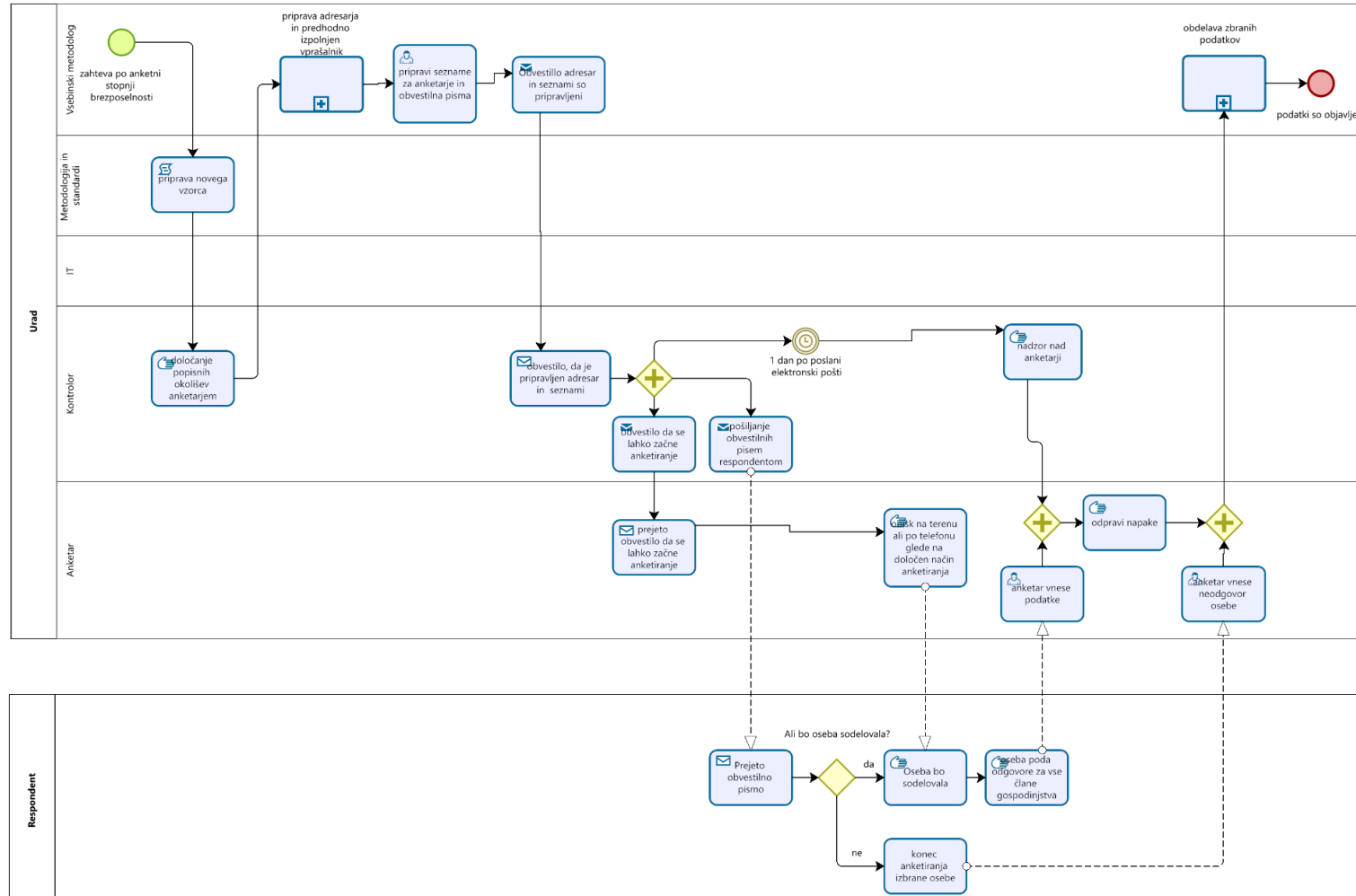
Sprožitveni dogodek procesa je zahteva po podatku o anketni stopnji brezposelnosti. Vsebinski metodolog poda zahtevo na metodologijo in standarde. Strokovnjak za metodologijo določi vzorčni okvir in pripravi vzorec oseb za novi val adresarja. Kontrolor preveri obstoječe popisne okoliše glede na anketarje in po potrebi naredi menjave. Sledi podproces za pripravo adresarja in podatkov za predhodno izpolnjen vprašalnik. Vsebinski metodolog sproži postopke za pripravo adresarja, kjer prevzame podatke novega vzorca, podatke oseb iz adresarja preteklega četrletja, ki imajo ustrezen status odgovora in so še vedno aktivni, ter zadnje pridobljene telefonske številke. Dobljeni adresar za novo četrletje se preveri in v primeru odpovedi anketarjev, jih kontrolor zamenja z novimi. Vsebinski metodolog sproži pripravo predhodnikov iz preteklih zbiranj. V primeru težav le-te odpravi razvijalec procesov. Sledi pregled vprašalnika in v primeru potrebnih sprememb le-tega prilagodi razvijalec vprašalnika. Vsebinski metodolog uskladi šifrant pomembnih datumov: datum pošiljanja pisem, referenčni teden, obdobje anketiranja, in odpravi odstopanja.

Po zaključku podprocesa za pripravo adresarja in predhodnikov, vsebinski metodolog sproži postopek za pripravo seznamov za terensko anketiranje in obvestilnih pisem ter o tem obvesti kontrolorja. Le-ta sproži pošiljanje obvestilnih pisem respondentom glede na predviden termin pošiljanja pisem. O tem obvesti anketarje in naslednji dan prične z nadzorom nad anketarji. Respondent prejme obvestilno pismo in se odloči, ali bo sodeloval v anketi. Respondenta glede na načrt zbiranja pokliče telefonski anketar ali obišče terenski anketar na naslovu. V primeru zavrnite je anketiranje zaključeno in anketar vnese neodgovor za izbrano osebo. Če respondent privoli v anketiranje, se le-to prične. Anketar vnaša odgovore v računalniško podprt vprašalnik. Delo anketarjev nadzorujejo kontrolorji. V primeru napak le-te odpravijo. Surovi podatki se shranjujejo v podatkovno bazo.

Po zbranih surovih podatkih sledi obdelava podatkov, prikazana v drugem podprocesu. Surovim podatkom se določi status gospodinjstva, ki določi ustreznost enote. Surovi podatki se delijo na osebne oz. občutljive podatke in neosebne podatke. Izbrana oseba poroča podatke za vse člane gospodinjstva, katerim je potrebno določiti osebni enolični identifikator. Postopek avtomatskega iskanja na podlagi osebnih podatkov sproži vsebinski metodolog. V primeru, da osebnega enoličnega identifikatorja ne moremo določiti avtomatsko, se le-te poišče ročno. Pretvornik med številko osebe v gospodinjstvu in osebnim enoličnim identifikatorjem se redno osvežuje in je osnova za povezovanje pridobljenih podatkov z administrativnimi viri na uradu. Neosebni podatki se delijo na podatke o opravljenih delih, adhoc odgovore in vse ostale stalne podatke.

Glede na opise opravljenih del članov oseb v gospodinjstvu se sproži avtomatsko določanje šifre standardne klasifikacije dejavnosti in poklica. V primeru težav se avtomatsko pošlje obvestilo razvijalcu rešitve, ki le-te odpravi. Po uspešni zaključitvi avtomatskega iskanja šifer se preveri, ali imajo vsi opisi določeno šifro. V primeru manjkajočih šifer, se te določijo ročno in se napolnijo v podatkovno bazo. Metodolog sproži postopke za prevzem administrativnih virov in urejanje podatkov enot, ki so ustrezne. Na urejenih podatkih strokovnjak za metodologijo sproži postopke za izračun uteži. Vsebinski metodolog sproži agregacijo in tabelacijo.

Slika 10: Del glavnega procesnega modela BPMN nove rešitve



Vir: lastno delo.

Tabela 6: Primerki povezav med poslovnimi pravili in procesi

Poslovno pravilo	Proces v BPMN
Vsebinski metodolog mora objaviti anketno stopnjo brezposelnosti.	Sprožitveni dogodek vsebinskega metodologa: zahteva po anketni stopnji brezposelnosti.
Strokovnjak za metodologijo mora pripraviti novi del vzorca za anketiranje.	Aktivnost klic skripte strokovnjaka za metodologijo: priprava novega vzorca.
Kontrolor mora določiti popisni okoliš terenskim anketarjem.	Aktivnost ročno opravilo kontrolorja: določanje popisnih okolišev anketarjem.
Za izbrano osebo mora veljati vse od naštetega (del pravila): izbrana oseba iz 2., 5., ali 6. vala je ustrezno odgovarjala v zadnjem preteklem anketiranju.	Aktivnost klic skripte vsebinskega metodologa: osebe iz preteklih zbiranj, ki so ustrezno odgovarjale.
Izbrana oseba mora prejeti obvestilno pismo določeno z vsebino v tabeli 4.	Aktivnost prejetje sporočila respondenta: prejeto obvestilno pismo.
Anketiranje izbrane osebe se mora začeti 7 dni po začetku referenčnega tedna kateremu le-ta pripada.	Aktivnost ročno opravilo anketarja: obisk na terenu ali po telefonu glede na določen način anketiranja.
Za izbrano osebo mora veljati nekaj od naštetega: <ul style="list-style-type: none"> – sprejme sodelovanje, – zavrne sodelovanje. 	Ekskluzivno razvejišče v območju respondenta: Ali bo oseba sodelovala?
Vsaka izbrana oseba in njeni družinski člani morajo imeti določen osebni enolični identifikator.	Aktivnosti in razvejišča vsebinskega metodologa od sprožitve aplikacije: iskanje enoličnega osebnega identifikatorja, do sprožitve aplikacije: osvežitve pretvornika številka osebe z enoličnim identifikatorjem.

Vir: lastno delo.

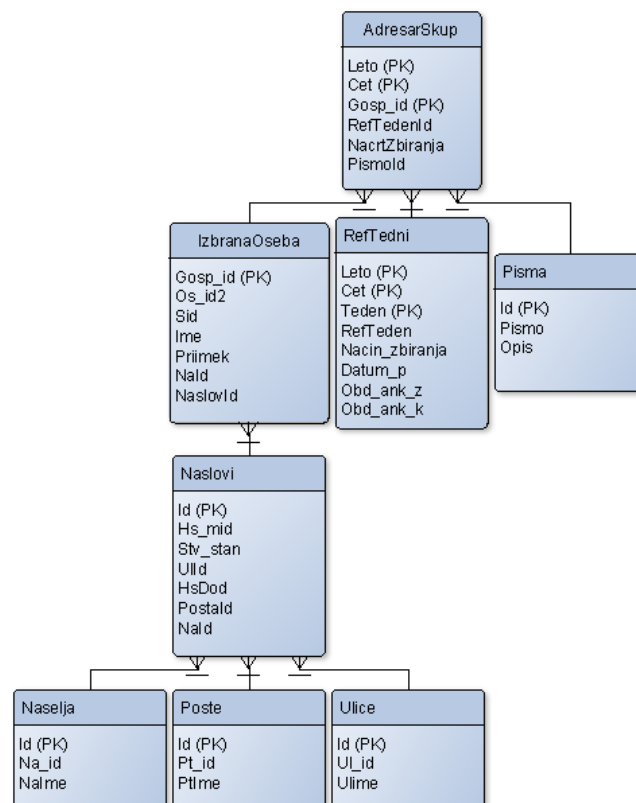
V primeru težav se obrne na strokovnjaka za metodologijo ali razvijalca rešitve. Po odpravljenih težavah imamo končno tabelacijo pripravljeno za objavo. Le-ta vključuje tudi stopnjo brezposelnosti.

Za integracijo poslovnih pravil v procese ne uporabimo orodja za eksplicitno urejanje le-teh. Pravila so kompleksna in večinoma uporabna zgolj za področje ANP. Poleg tega vsebinski metodologi nimajo ustreznega znanja za upravljanje pravil. Na uradu tudi ni ustaljene prakse za uporabo orodja, ki bi omogočalo to funkcionalnost. Posledično pazimo, da imamo za vsak ključni del procesa ustrezno pravilo, ker skladnost med pravili in procesi ni avtomatizirana.

3.3 Podatkovno modeliranje

Na izbiro podatkovnega modela vpliva fizično okolje, kjer bodo podatki dejansko shranjeni. Na uradu uporabljamo relacijsko podatkovno bazo Microsoft SQL Server. Za konceptualni model izberemo ER, za logični pa relacijski model. Kot je omenjeno v poglavju 1.3.3 sta ER in relacijski model povezana, saj se ER lahko transformira v relacijski model. Del ER modela prikazuje slika 11, celoten ER je viden v prilogi 4.

Slika 11: Primer ER modela za pripravo adresarja



Vir: lastno delo.

Osnove za postavitev ER modela so model dejstev in poslovna pravila. Na podlagi modela dejstev, ki je sestavljen iz terminov in povezav med njimi, se določijo poslovna pravila. Subjekti pravil se preslikajo v entitetne tipe ER modela in glagoli v povezave med njimi. Modeliramo postopoma. Vsakemu entitetnemu tipu določimo attribute in glavni ključ. Povežemo jih med seboj na podlagi glavnega ključa. Če entitetni tipi niso povezljivi ali imamo povezavo mnogo proti mnogo, dodamo entitetne tipe oz. razmislimo o dodatnih atributih. Postopek ponavljamo, dokler ER model ne ustreza zahtevam. V fazi konceptualnega modeliranja poskušamo v čim večji meri upoštevati tudi normalizacijo baze. Torej zmanjšamo funkcionalne odvisnosti atributov in zmanjšamo redundanco podatkov. Podatkovno bazo normaliziramo do tretje normalne forme. To pomeni, da vsi atributi vsebujejo atomične vrednosti, je določljiv glavni ključ, so vsi neključni atributi odvisni od

celotnega glavnega ključa in neključni atributi niso funkcionalno odvisni od ostalih neključnih atributov.

ER model (priloga 4) lahko vsebinsko razdelimo na tri dele: zbiranje podatkov, prevzem surovih podatkov in obdelava. Seznam entitetnih tipov in opisov se nahaja v tabeli 7, tabeli 8 in tabeli 9, kjer je vidna povezava med poslovnimi pravili in ER modelom.

V fazi zbiranja podatkov (tabela 7) je entitetni tip AdresarSkup osnova za začetek zbiranja. Njegovi glavni ključni so: leto, četrletje in številka gospodinjstva. Adresar je v povezavi mnogo proti ena z entitetnimi tipi IzbranaOseba, RefTedni in Pisma. IzbranaOseba je v povezavi mnogo proti ena z entitetnim tipom TelStevilke in Naslovi. Glavni ključ slednjemu določimo kot enolični identifikator, ki se samodejno povečuje za ena in omogoča lažjo povezljivost s podatki o izbrani osebi, saj je glavni ključ sestavljen iz treh atributov. Entitetni tip Naslovi so v povezavi mnogo proti ena z entitetnimi tipi Naselja, Ulice in Poste. Za lažje vzdrževanje ločimo podatke o tipu naselja in občini kot dva entitetna tipa, med seboj pa jih povežemo preko entitetnega tipa z glavnim ključem NaId. Podobno tudi ločimo podatke o občini in regiji in jih med seboj povežemo z entitetnim tipom ObcineRegije z glavnim ključem ObcId. Tako povežemo adresar za izbrane osebe s točnim naslovom prebivališča.

Zbiranje podatkov ne more potekati brez anketarjev. Entitetni tip Anketarji se deli na entitetna podtipa AnketarjiTel in AnketarjiTer. Entitetna tipa Anketarji in Kontrolorji sta v povezavi mnogo proti mnogo, saj lahko en kontrolor nadzira več anketarjev in obratno. Posledično dodamo vmesni entitetni tip AnketKontrolorji s povezavami na številko anketarja in številko kontrolorja, ki povezavo deli na ena proti mnogo. Vsakemu terenskemu anketarju pripada eden ali več popisnih okolišev, zato sta entitetna tipa AnketarjiTel in PopOkolis v povezavi mnogo proti ena. Zaradi pogoste menjave anketarjev obstaja ločeni entitetni tip UpdateAnk, kjer določimo novega anketarja za popisni okoliš in se sprememba vnese tudi v entitetni tip PopOkolis. Entitetni tip PopOkolis_naslov je v povezavi mnogo proti ena s PopOkolis in Naslovi. Tako povežemo naslove z anketarjem.

Seznam entitetnih tipov za prevzem surovih podatkov prikazuje tabela 8. Entitetni tip Anp_data je v povezavi s šifrantom Anp_id. Za enolično povezljivost je še dodan vmesni entitetni tip Anp_Data_An_p_Id. V Anp_data pomeni ena entiteta eno vrednost iz vprašalnika. Za zapis podatkov v produkcijsko okolje, kjer se izvaja obdelava podatkov, potrebujemo še entitetni tip BlaisePret, ki določa povezljivost surovih podatkov s podatki v produkcijskem okolju.

Seznam entitetnih tipov za obdelavo podatkov prikazuje tabela 9. Entitetni tip Odgovor je v povezavi ena proti ena z Osebni. Entitetni tip Gosp_skupaj je v povezavi z Odgovor ena proti ena samo z izbrano osebo. Osebni enolični identifikator se določi na podlagi podatkov entitetnega tipa Osebni. OsebeSid omogoča povezljivost z administrativnimi viri na uradu, zato je v povezavi z OsebeAdmin_skupaj. Entiten tip Odgovor je v povezavi ena proti mnogo z OsebeSkdSkp, Osebe_skupaj in OsebeAdhoc_skupaj.

Tabela 7: Seznam entitetnih tipov za zbiranje podatkov in njihov pomen

Ime entitetnega tipa	Pomen
AdresarSkup	Seznam gospodinjstev, kjer izbrana oseba živi, z določenim načrtom zbiranja in vrsto obvestilnega pisma.
IzbranaOseba	Seznam izbranih oseb, ki so primerne za anketiranje, tako novi del vzorca kot osebe iz preteklih zbiranja, ki so ustrezne.
RefTedni	Seznam referenčnih tednov v četrletju in povezanih datumov, kot so pošiljanje pisem, obdobje anketiranja.
Pisma	Seznam po vrstah pisem, ki so določene glede na načrt zbiranja in vala, v katerem se izbrana oseba nahaja.
TelStevilke	Seznam pridobljenih telefonskih številke za poljubno osebo znotraj gospodinjstva.
Naslovi	Seznam točnih naslovov, kjer izbrane osebe živijo.
Naselje	Seznam naselji.
Ulice	Seznam ulic in hišnih števil.
Poste	Seznam poštnih števil in nazivov.
TipNaselja	Seznam tipov naselji.
Obcine	Seznam občin.
NaseljeObcineTipNaselja	Seznam katerim občinam pripada katero naselje in kakšnega tipa je vsako naselje. Eni občini pripada več naselji, enemu tipu naselji pripada več naselji.
Regije	Seznam regij.
ObcineRegije	Seznam katerim regijam pripadajo katere občine. V eni regiji je lociranih več občin.
Kontrolorji	Seznam kontrolorjev.
Anketarji	Seznam anketarjev z določeno vrsto anketiranja.
AnketKontrolorji	Seznam katere anketarje nadzirajo kateri kontrolorji. En anketar je lahko pod nadzorom več kontrolorjev in en kontrolor lahko nadzira več anketarjev.
AnketarjiTel	Seznam šifre anketarjev s pripadajočimi telefonskimi številkami, na katere pokliče.
AnketarjiTer	Seznam terenskih anketarjev z določenimi popisnimi okoliši, kateri mu pripadajo.
PopOkolis	Seznam popisnih okolišev.
PopOkolis_Naslovi	Seznam katerim popisnim okolišem pripadajo kateri naslovi, ki jih obiše anketar. V enem popisnem okolišu je lociranih več naslovov.
UpdateAnk	Menjave šifre anketarjev glede na popisne okoliše.

Vir: lastno delo.

Tabela 8: Seznam entitetnih tipov za prevzem surovih podatkov in njihov pomen

Ime entitetnega tipa	Pomen
Anp_Data	Surovi podatki zbrani s terenskim in telefonskim anketiranjem izbranih oseb in pripadajočih družinskih članov.
Anp_Id	Seznam poimenovanj vseh spremenljivk vprašalnika glede na tip vprašanja, ki jih avtomatsko generira Blaise.
Anp_dat_An_p_id	Seznam, katero vprašanje glede na tip ustreza surovemu podatku v produkcijskem okolju za obdelavo.
BlaisePret	Seznam, katere spremenljivke iz vprašalnika se povežejo na kateri atribut v produkcijskem okolju za obdelavo.
Anp_id_BlaisePret	Seznam, katere spremenljivke iz vprašalnika se uporabijo iz šifranta vseh spremenljivk.

Vir: lastno delo.

Tabela 9: Seznam entitetnih tipov za obdelavo podatkov in njihov pomen

Ime entitetnega tipa	Pomen
Odgovor	Seznam vseh odgovorov na vprašalnik za izbrane osebe in pripadajoče družinske člane.
Osebni	Seznam občutljivih odgovorov na vprašalnik za izbrane osebe in pripadajoče družinske člane.
Gosp_skupaj	Seznam statusov, ki določajo ustreznost izbranih oseb v gospodinjstvu oz. ustreznost gospodinjstva za statistično obdelavo.
OsebeSid	Pretvornik med številko osebe in enoličnim osebnim identifikatorjem za vse izbrane osebe in pripadajočih članov.
OsebeSkdSkp	Seznam odgovorov o opravljenem delu za osebe, ki so delale, z določenimi šiframi standardne klasifikacije poklica oz. dejavnosti.
Osebe_skupaj	Seznam odgovorov in po potrebi urejeni in imputirani odgovori na stalna vprašanja za izbrane osebe in pripadajoče družinske člane.
OsebeAdhoc_skupaj	Seznam odgovorov in po potrebi urejeni in imputirani odgovori na adhoc vprašanja za izbrane osebe in pripadajoče družinske člane.
Skd	Seznam šifer in opisov standardne klasifikacije dejavnosti.
Skp	Seznam šifer in opisov standardne klasifikacije poklicev.
OsebeAdmin_skupaj	Seznam podatkov in po potrebi urejeni in imputirani podatki administrativnih virov za izbrane osebe in pripadajoče družinske člane.
Utez	Seznam uteži za izbrane osebe.

Vir: lastno delo.

Vsak član gospodinjstva lahko opravlja več različnih del, zato je entitetni tip `OsebeSkdSkp` v povezavi mnogo proti ena z `Odgovor`. V pomoč pri določanju standardne klasifikacije poklica in dejavnosti na podlagi opisa so tudi administrativni viri, ki jih pripeljemo preko entitetnega tipa `OsebeSid`. V povezavi z opravljenimi deli sta tudi šifranta standardna klasifikacija poklica (SKP) in dejavnosti (SKD).

Podatki o odgovorih in administrativnih virih, pomembnih za področje ANP, se urejajo in imputirajo, zato je vsak entitetni tip v povezavi z istoimenskim tipom in enako poimenovanimi atributi, le da je dodana pripona »s«. To pomeni, da vodimo statuse za vsak atribut, ki služijo kot informacija o spremembi podatka skozi faze urejanja. Za končno agregacijo potrebujemo še uteži določene na ravni gospodinjstva. Entitetni tip o utežeh je v povezavi ena proti ena z entitetnim tipom `Gosp_skupaj`.

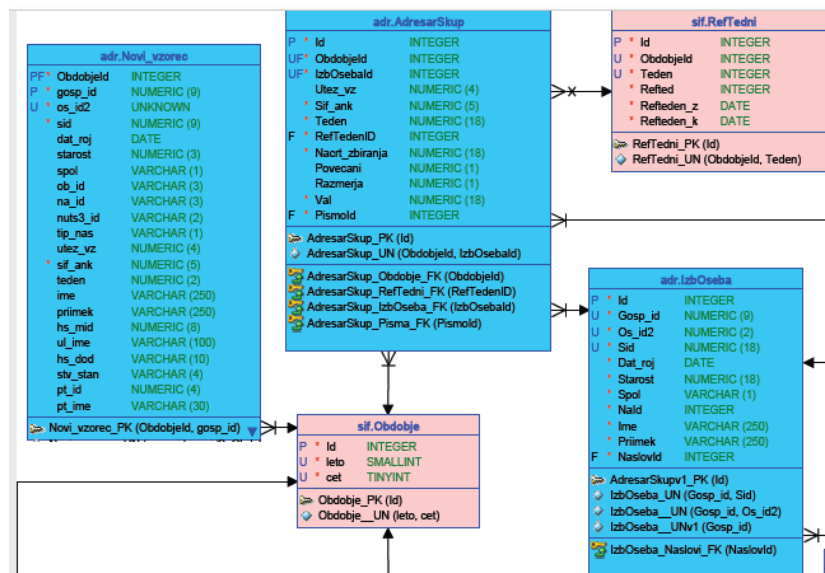
V fazi logičnega modeliranja zrišemo relacijski model v okolju SQL Developer Data Modeler. Entitetne tipe transformiramo v relacije, entitete v vrstice, attribute v stolpce in obdržimo oz. dodamo povezave. Na podlagi poslovnih pravil določimo točen nabor stolpcev, tipe, domene in ostale lastnosti, kot so obvezno izpolnjeni stolpci, primarni ključ, unikatni ključ, tuji ključ, povezave, prožilec, za vsako relacijo. Z označbo, kateri stolpci so obvezno izpolnjeni, se izognemo nepotrebni praznim vrednostim. Domene čim bolj omejimo, kar dodatno prispeva k pravilnemu podatku. Če npr. vemo, da je šifra anketarja 5 mestna numerična vrednost, potem je domena »numeric (5)«. Vsaki relaciji določimo primarni ključ in unikatni ključ. Relacije, ki so povezane, povežemo preko tujega ključa, da imamo povezavo starš-otrok. Povezavi mnogo proti mnogo smo že v večini primerov odpravili v fazi konceptualnega modeliranja.

Atributa leto in četrtletje sta pogosto uporabljena kot del primarnega ključa, zato dodamo novo relacijo, kjer je id-ju umetno dodan surogatni primarni ključ. Zaradi lažje povezljivosti relacij med sabo pogosto uporabimo surogatni ključ, določimo pa še unikatni ključ, ki določa enoličnost relacij na pravih vrednostih. Relacije, kjer se nahajata leto in četrtletje, povežemo z novo relacijo, pri čemer id nastopa kot tuj ključ imenovan `ObdobjeId`.

Slika 12 prikazuje točni seznam atributov, tipe polj in domene, primarni ključ za relacijo `AdresarSkup` in relacije, ki so v povezavi. Zaradi lažje povezljivosti relacij med seboj pogosto uporabimo surogatni ključ. Določimo tudi unikatni ključ, ki določa enoličnost zapisov relacij na izbranih vrednostih. `Adresar` ima unikatni ključ, sestavljen iz tujih ključev `ObdobjeId` in `IzbOsebaId`, torej je `adresar` v odvisnosti od relacij o izbrani osebi in obdobju. Relacije s tujim ključem lahko določimo različne scenarije aktivnosti, ki naj se zgodijo ob spreminjanju primarnega ključa. `Obdobje` je v povezavi z `adresarjem`. Če bi v fizičnem okolju želeli pobrisati obdobje, ki se že nahaja v tabeli `adresar`, potem bi dobili opozorilo, da obdobja ni mogoče pobrisati, ker se vrednost nahaja v povezani tabeli kot tuj ključ. Tabela `RefTeden`, ki predstavlja referenčne datume, je v povezavi z `adresarjem`. Če pobrišemo id v `RefTeden`, se avtomatsko pobrišejo vsi zapisi v `AdresarSkup`. To pomeni, da imamo fizično na tabeli `RefTeden` prožilec, ki avtomatsko sproži brisanje podatkov druge tabele. Na ta

način lahko vnesemo pravila v relacijskem modelu, ki varujejo podatke pred nenadzorovanimi akcijami. Za boljšo vizualno predstavo relacije obarvamo. Šifranti so roza, relacije potrebne za pripravo na zbiranje so modre, relacije za zbrane surove podatke so zelene in ostale relacije so rumene barve. Del relacijskega modela prikazuje slika 12, celoten je viden v prilogi 5 in prilogi 6.

Slika 12: Primer relacijskega modela za pripravo adresarja



Vir: lastno delo.

Relacijski model izvozimo v strukturo zapisa DDL, ki ga uporabimo za kreiranje fizičnega podatkovnega okolja s tabelami. Relacijska baza je Microsoft SQL Server. Dodamo indekse na tabele, kjer je to še potrebno. Kreiramo poglede, s katerimi zmanjšamo število fizičnih objektov v bazi. Primeri pogledov so predhodniki za vprašalnike, sezname za anketarje in obvestilna pisma. Uredimo dostope do objektov z ustreznimi pravicami samo za branje ali tudi branje, pisanje in brisanje. Primer: na tabeli adresar mora imeti vsebinski metodolg, ki proži postopke za pripravo adresarja, pravice za branje, pisanje in brisanje, kontrolorju zadostujejo pravice za branje. Z omejitvijo dostopov dodatno prispevamo k upoštevanju poslovnih pravil. Uredimo tudi particioniranje tabel o odgovorih, saj gre količinsko za večje tabele v bazi, in prispevamo k hitrejši dostopnosti do podatkov.

Poslovni procesi so osnova za razvoj postopkov za polnjenje podatkov. Le tako lahko testiramo podatkovno bazo. Splošne rešitve, ki že obstajajo in jih ni potrebno razviti, so iskanje enoličnega identifikatorja, urejanje, izračun uteži, agregacija in tabelacija. Razvijemo procedure za:

- polnjenje in osveževanje šifrantov, kot so obdobja, referenčni tedni, standardna klasifikacija dejavnosti in poklicev, naselja, popisni okoliši itd.,
- pripravo adresarja za tekoče četrletje,

- menjavo anketarjev glede na popisni okoliš,
- polnjenja surovih podatkov,
- integracija in transformacija surovih podatkih in delitev na osebne, podatke o delu, adhoc in stalne odgovore,
- iskanje enoličnega osebnega identifikatorja,
- iskanje šifer standardne klasifikacije dela in poklica za opravljena dela,
- polnjenje administrativnih podatkov.

S procedurami testno napolnimo podatke in preverimo delovanje podatkovne baze. Preverimo ustreznost objektov, pravic, odzivnost.

3.4 Ovrednotenje nove rešitve

Za ovrednotenje obstoječe rešitve opravimo intervju z zaposlenimi na uradu, ki so aktivno vključeni v sam proces oz. imajo širši pogled nad delovanjem urada. Osredotočimo se na predlog nove rešitve, in povprašamo po njihovem mnenju o prednostih in slabostih le-te. Zanima nas ali intervjuvanci prepoznajo medsebojno povezanost poslovnih pravil, procesov in podatkovnega modeliranja in če se zavedajo vpliva teh dejavnikov na kakovost podatkov.

3.4.1 Metodologija

Intervju opravimo po testiranju treh časovnih serij nove rešitve. Opravimo ga z vsebinskim metodologom, vodjem področja vsebine, strokovnjakom za metodologijo, razvijalcema informacijskih rešitev s področja razvoja vprašalnikov in procesov ter informacijskim pooblaščenecem urada. Skupaj tako opravimo 6 intervjujev na sami lokaciji urada. Oblika vprašalnika je polstrukturirana. To pomeni, da so vprašanja predhodno pripravljena in jih glede na potek pogovora sproti prilagajamo. Vprašanja so v kvalitativni obliki. Intervju poteka od 30 do 60 minut. Vse pogovore opravimo v enem tednu. Intervjuvanci podajo odgovore oz. mnenja na sledeča vprašanja:

- kakšna je vloga zaposlenega v obstoječem procesu področja ANP,
- kakšna je vloga zaposlenega v novem procesu,
- kakšne so prednosti nove rešitve,
- kako pomembni so vidiki poslovna pravila, procesi in podatkovno modeliranje v novi rešitvi,
- kakšno kakovost podatkov smo dosegli glede na dimenzije kakovosti urada,
- ali vidijo možnosti izboljšave nove rešitve,
- ali poslovna pravila, procesi in podatkovno modeliranja vplivajo na kakovost podatkov,
- ali lahko rečemo za novo rešitev, da je bolj kakovostna.

Pri opredelitvi kakovosti podatkov so intervjuvancem v pomoč smernice za zagotavljanje kakovosti podatkov (poglavje 1.5.3). Odgovore zabeležimo in podamo rezultate na podlagi skupne analize.

3.4.2 Ovrednotenje z vidika prednosti in slabosti nove rešitve

Vloge deležnikov so se z novo informacijsko rešitvijo bistveno spremenile. Večino procesov, ki pokrivajo pripravo na zbiranje in obdelavo podatkov, proži in nadzoruje preko aplikacije vsebinski metodolog. Izjema je priprava vzorca za novi val, izračun uteži, anketiranje in nadzor nad izvedbo, kar je v domeni strokovnjaka za metodologijo, anketarjev in kontrolorjev. Razvijalca informacijskih rešitev sta potrebna samo ob spremembah, torej ob prehodu na novo leto, ali za nudenje tehnične pomoči.

Vsi intervjuvanci so si enotni glede prednosti nove rešitve v shranjevanju podatkov v podatkovno bazo, kjer ni podvajanj podatkov. Za shranjene podatke je točno določen podatkovni tip, domena, z določeno obveznost podatkov, kar dodatno zmanjša shranjevanje slabih, nepopolnih podatkov in prispeva k večji kakovosti le-teh. Postopek od priprave na zbiranje, zbiranje in obdelave podatkov poteka po točno določenem procesu, kar minimalizira možnosti človeških napak. V uporabi so standardne rešitve urada za postopke vstavljanja, urejanja, agregacije in tabelacije. Celoten proces obdelave poteka bistveno hitreje, saj so postopki skoraj v celoti avtomatizirani in optimizirani. Podatki so skrbno varovani in urejeni z dostopi, do katerih lahko dostopajo samo tisti zaposleni, ki jih dejansko potrebujejo.

Vsebinski metodolog in njegov vodja sta si enotna, da se z rednim iskanjem enoličnega identifikatorja, lahko za izračun statistike uporabljajo tudi administrativni viri. Le-ti izboljšajo objavljene statistike in razbremenijo poročevalske enote. Z avtomatskim določanjem šifer standardne klasifikacije dejavnosti in poklica so podatki bolj kakovostni. Obenem se zmanjša ročno določanje šifer in pohitri postopke. Prednost je tudi v tem, da je urejena sledljivost nad spremembami skozi proces vstavljanja in urejanja podatkov.

Vodja oddelka predvideva, da so zaradi izboljšane informacijske rešitve tudi rezultati obdelave bistveno bolj kakovostni.

Strokovnjak za metodologijo dodatno izpostavi, da odpade ročno prevzemanje surovih podatkov in določanje statusa gospodinjstva. Izračun uteži je bistveno bolj enostaven, saj so vsi podatki na voljo v enem okolju. Postopki so dosledni, ponovljivi in hitrejši. Adresarski podatki so posledično kakovostnejši.

Razvijalec vprašalnika izpostavi prednost v avtomatiziranem postopku priprave adresarja in obvestilnih pisem. Surovi podatki se shranjujejo direktno v podatkovno bazo, odpadejo dodatni izpisi in manj verjetno pride do nenadzorovanih akcij, kot je brisanje podatkov. Razvijalec podatkovne baze in procesov izpostavi, da je za razvoj nove rešitve potrebno več

časa, informiranj, usklajevanj, da je rešitev čim bolj ustrezna. Dodana vrednost pravilno načrtovane nove rešitve je, da se s produkcijo procesi bolj pravilno, nadzorovano, brez podvajanj in hitreje izvedejo. Podatki so kakovostnejši, saj so procesi skrbno načrtovani z vključitvijo ključnih deležnikov in upoštevanjem poslovnih pravil. Podatkovna baza je načrtovana z upoštevanjem poslovnih pravil in normalizirana, kar pomeni, da so shranjeni podatki bolj kakovostni.

Informacijski pooblaščenec izpostavi, da smo zaposleni dolžni spoštovati pravilnik o varstvu osebnih podatkov, ki določa omejeno dostopanje, skrbno ravnanje in varno okolje za shranjevanje osebnih podatkov (Statistični urad Republike Slovenije, 2022b). Urad s politiko varovanja informacij določa celostni sistem varovanja informacij in podatkov, ki temelji na ISO standardu 27001 (Statistični urad Republike Slovenije, 2022a). Glede na politiko je pravilni pristop shranjevanje podatkov v podatkovni bazi z omejenimi dostopi in varovanje občutljivih podatkov. Do občutljivih podatkov ima dostop izredno malo oseb, in še te imajo kar najbolj omejene pravice. Urejena je sledljivost dostopanja do občutljivih podatkov z dnevniki obdelav, kot določa zakon o varovanju informacij.

Če povzamemo ključne prednosti nove rešitve:

- shranjevanje podatkov v podatkovno bazo z upoštevanjem pravil in nadzorovano dostopanje oseb z omejenimi pravicami,
- urejena sledljivost dostopanja do občutljivih podatkov,
- celoten proces od priprave na zbiranje, zbiranje, prevzem surovih podatkov in obdelava so sistematični, skoraj v celoti avtomatizirani in se tako zmanjša možnost napak,
- z avtomatiziranim postopkom določanja enoličnega identifikatorja so redno na voljo administrativni podatki,
- poveča se število izračunanih statistik,
- z avtomatskim določanjem šifer standardne klasifikacije dejavnosti in poklica na podlagi opisov, se zmanjša ročno določanje šifer in zmanjša možnost napačnih podatkov
- celoten proces poteka bistveno hitreje,
- odpravljena je redundanca v podatkih in anomalije, kot so napačne, prazne vrednosti, nepovezljivost podatkov,
- podatkovna baza je normalizirana, kar omogoča enostavnejše polnjenje in posodabljanje podatkov, z večanjem števila podatkov v bazi pa prispevamo k čim manjši porabi prostora,
- vsak del prispeva k večji kakovosti podatkov.

Z vidika pomembnosti poslovnih pravil, procesov in podatkovnega modeliranja so si vsi intervjuvanci enotni, da so poslovna pravila nujno pomembna za doseganje kakovostnejše informacijske rešitve. Pomembno je, da se ob prenovi vključeni vsi ključni deležniki in s svojim znanjem in izkušnjami prispevajo k izboljšanju rešitve. Poslovna pravila so vključena tako v fazi procesa kot tudi v fazi podatkovnega modeliranja.

Glede možnosti izboljšave, so izpostavili področje priprave novega vzorca in izračun uteži, kjer sta postopka še vedno v domeni strokovnjaka za metodologijo. Predlagajo, da se izboljšajo postopki. Posledično se lahko ta del procesa avtomatizira in preideta v domeno vsebinskega metodologa. Predlagajo tudi možnost spletnega anketiranja. Le-to je zaradi obsežnega vprašalnika, anketiranja izbrane osebe in družinskih članov izredno kompleksno in tvegano glede kakovosti zbranih podatkov.

3.4.3 Ovrednotenje glede na dimenzije kakovosti podatkov

Zanima nas vpliv poslovnih pravil, procesov in podatkovnega modeliranja na kakovost podatkov za vsako izmed dimenzij. Podamo skupne rezultate analize za vse intervjuvance.

Ustreznost podatkov je izboljšana z uporabo administrativnih podatkov. Viri so povezljivi s pridobljenimi podatki preko osebne enoličnega identifikatorja. Le-ti zagotavljajo večji nabor izračunanih statistik, ki jih stranke potrebujejo.

Potek vzorčenja ostaja nespremenjen in na točnost ocen ne vpliva. Izboljšanje točnosti zbranih podatkov dosežemo s postavitvijo podatkovne baze, procesov in avtomatizacijo rešitve. Na kakovost podatkovne baze vpliva načrtovanje, ki se prične s poslovnimi pravili, procesi in podatkovnim modeliranjem, ki se med seboj prepletajo. S pravilnim pristopom na podlagi terminov in dejstev določimo poslovna pravila. Le-ta vključimo v poslovne procese ali v podatkovno bazo. Vsak korak v načrtovanju prispeva k točnosti ocen od procesnega, konceptualnega, logičnega ali fizičnega podatkovnega modela. Pri postavitvi tabel v podatkovni bazi so upoštevana pravila, kot so omejitve enoličnosti, referenčna celovitost, pravilni tipi, omejitve vrednosti polj, določitev obveznih podatkov, vsebinska in časovna ustreznost, ki prispevajo h kakovosti podatkov. V uporabi je splošna rešitev urada za urejanje in vstavljanje podatkov, kar dodatno izboljša točnost ocen.

Z določitvijo poslovnih pravil le-te vključimo v modeliranje procesov in podatkovno modeliranje. Izvajanje korakov procesa in podatkovnega modeliranja je na ta način bolj transparentno in ponovljivo. Z opredelitvijo pravil, procesov in podatkovnega modeliranja dosežemo, da so postopki pravilni, s tem zmanjšamo možnost napak v samem izvajanju in rezultati obdelav so bolj kakovostni. Postopke obdelav lahko avtomatiziramo in posledično pohitrimo celotno obdelavo podatkov. Tako zmanjšamo tveganje nepravočasne razpoložljivosti podatkov glede na instanco zbiranja. Podobno zmanjšamo tveganje zamude objave podatkov glede na napovedano objavo. Po njihovem mnenju nova rešitev izboljša dimenzijo pravočasnosti in točnosti objav.

Na dimenzijo dostopnosti in jasnosti nova rešitev nima direktnega vpliva. Način objavljanja končnih podatkov za uporabnike ostaja nespremenjen. Agregirani podatki so dostopni na spletni strani kot objave in v sistemu podatkovne baze. Podatki so opremljeni z metapodatki. Nova rešitev posredno vpliva na objavo dodatnih statistik, posodobitvi metapodatkov in pravilni razlagi dodatno objavljenih statistik.

Urad stalno mednarodno sodeluje in usklajuje definicije ter koncepte področja. Podatki se stalno izmenjujejo v okviru statističnega sistema, vendar kljub temu nova rešitev vpliva na skladnost in primerljivost. Z vključitvijo pravil v poslovne procese in podatkovno modeliranje zagotovimo, da so podatki bolj primerljivi in skladni, saj zmanjšamo možnost napak v samem procesu oz. v postopku shranjevanja podatkov v podatkovno bazo.

SKLEP

V magistrskem delu nas zanima kako so medsebojno povezani vidiki poslovnih pravil, procesov in podatkovnega modeliranja ter kako ti vidiki vplivajo na kakovost podatkov. Omejili smo se na procesno usmerjene rešitve v relacijski podatkovni bazi in proučili kakšen je vpliv procesno usmerjene rešitve na dimenzije kakovosti, ki so v veljavi na statističnih uradih.

Iz pregleda literature izhaja, da so pravila temelj vsake poslovne rešitve, zato je pravilni pristop, da se le tega lotimo sistematično, premišljeno z vključitvijo oseb, ki imajo o poslovnem procesu največ znanj in izkušenj. Rezultat so ključna poslovna pravila. Poslovni procesi prikažejo potek izvajanja aktivnosti in dogodkov, ki ga z modeliranjem vizualno približamo vsem deležnikom. Pomembno je, da poslovna pravila vključimo v potek izvajanja procesa. Poslovna pravila in procesi so osnova za podatkovno modeliranje, torej za načrtovanje shranjevanja podatkov. Le-to mora biti skrbno načrtovano v sledečih fazah konceptualnega, logičnega in fizičnega modeliranja. Podatkovni modeli so med sabo odvisni in vsaka faza je temelj za naslednjo. Tako pravila podajo odgovor, kaj naj se zgodi; procesi, kako naj se zgodi; podatkovno modeliranje, kako naj bodo podatki shranjeni v podatkovni bazi. Vsi ti dejavniki modeliranja pripomorejo, da so shranjeni podatki čim bolj kakovostni.

Analiza obstoječe rešitve je pokazala vrsto pomanjkljivosti, kot so slabo definirani koncepti terminov, pomanjkljivo definirani procesi, veliko oseb vključenih v izvajanje procesa, shranjevanje podatkov na datotečni strežnik, podvajanje podatkov, manjkajoči podatki, napačni podatki, slabo nadzorovani podatki, kar vse skupaj vpliva na kakovost shranjenih podatkov. Posledično vse skupaj vpliva na kakovost objavljenih končnih podatkov za uporabnike.

Nove rešitve smo se lotili po pravilih ugotovljenih v teoriji, z omejitvami glede na informacijsko komunikacijsko infrastrukturo, ki je na voljo. Prepoznali smo ključna poslovna pravila. Le-ta so vključena na različnih delih informacijske rešitve. S poslovnimi procesi pripeljemo podatke v podatkovno bazo. Z vključitvijo poslovnih pravil v procesih so le-ti transparentni in kakovostni. Posledično so podatki skozi procese bolj kakovostni in dodatno pripomorejo h kakovosti končnih podatkov. Za integracijo pravil v procese mora oseba imeti napredna IT znanja in ne le poznati osnov. Kljub različnim dejavnikom, ki jih navajajo avtorji Wang, Sadiq in Indulska (2014), je v praksi bolj uporabno, da so pravila integrirana v procesih kot uporaba namenskega orodja za implementacijo le-teh.

Poslovna pravila so vključena v podatkovno modeliranje kot prepoznavanje ključnih entitetnih tipov, povezav med njimi, glavnih ključev, določanje enoličnosti, referenčne integritete, povezljivosti med tabelami, omejitve na atributih, dostopanje do fizičnih podatkov, itd. Poleg tega so le ta tudi osnova, da je podatkovna baza normalizirana in tako prispevamo k minimalni porabi prostora in lažjemu vzdrževanju podatkov v podatkovni bazi.

Na podlagi omenjenega trdimo, da so poslovna pravila temelj, na katerih slonijo poslovni procesi in podatkovno modeliranje. Vsak od teh vidikov prispeva k učinkovitejši in uspešnejši informacijski rešitvi ter kakovosti podatkov. Posledično trdimo, da poslovna pravila, procesi in podatkovno modeliranje vplivajo na kakovost podatkov.

Kakovost podatkov je v ospredju že vrsto let in je še vedno aktualna tematika. Z zavedanjem pomena in stalnem izboljševanju kakovosti, le-to prispeva k uspešnosti poslovanja podjetja. Za opredelitev kakovosti je potrebno prepoznati njene dimenzije in dejavnike, ki vplivajo na to ter jih stalno izboljševati. V magistrski nalogi nas zanima vpliv poslovnih pravil, procesov in podatkovnega modeliranja na kakovost podatkov.

Prikazali smo značilnosti poslovnih pravil, procesov, podatkovnega modeliranja in kakovosti podatkov. Omejili smo se na relacijske podatkovne baze. Z vidika kakovosti podatkov obstaja veliko različnih teorij in zato smo se tudi tu omejili na dimenzije, ki so v veljavi na statističnih uradih. Teorija je podana na podlagi literature za vsako posamezno področje, torej za poslovna pravila, procese, podatkovno modeliranje in kakovost podatkov. Literature, kjer bi bile omenjena vsa področja skupaj, ni zaslediti. Obstajajo le gradiva o povezavi poslovna pravila s procesi in poslovna pravila s podatkovnim modeliranjem.

V teoretičnem delu smo prepoznali poslovna pravila kot temelj za poslovne procese in podatkovno modeliranje ter kako vsak izmed njih lahko prispeva h kakovosti rešitve in kakovosti podatkov. Z modeliranjem poslovnih procesov in podatkovnim modeliranjem smo dosegli, da podatke shranimo v podatkovno bazo. Kot smo dokazali, je to primerno okolje za shranjevanje podatkov. Pomembno je, da modeliramo v fazah konceptualnega, logičnega in fizičnega modela. Na ta način gradimo kakovostno podatkovno bazo. Z upoštevanjem poslovnih pravil in pravil podatkovnega modeliranja vplivamo na kakovost podatkovne baze in posledično tudi na kakovost podatkov. Spoznanja v teoretičnem delu smo potrdili in podkrepili na primeru raziskave ANP. Področje smo analizirali na podlagi razpoložljive dokumentacije in znanj ter izkušenj zaposlenih, ki so vključeni v fazah zbiranja in obdelave podatkov. Izpostavili smo pomanjkljivosti obstoječe rešitve. Podali smo predlog nove rešitve z določitvijo poslovnih pravil, procesov, konceptualnega, logičnega in fizičnega modeliranja. Razvili smo manjkajoče procedure in testno napolnili podatke. Preverili smo delovanje podatkovne baze in kakovost podatkov. Na podlagi intervjujev smo prikazali, da je temelj nove rešitve določitev poslovnih pravil. Le-ta vplivajo na nadaljnje faze razvoja

informatijske rešitve in imajo vpliv na dimenzije kakovosti in posledično na kakovost podatkov. Če poslovna pravila podajajo odgovore na vsa pomembna vprašanja, odpravimo težave že v začetni fazi razvoja in je postopek razvoja nove rešitve enostavnejši ter učinkovitejši.

Omejitve dela so, da temelji izključno na procesno urejeni organizaciji, glavna dejavnost le-te je zagotavljanje podatkov, ki so shranjeni v relacijski podatkovni bazi. Nova rešitev je ovrednotena na podlagi intervjujev. Le-to bi lahko izboljšali in podkrepili še s kvantitativnimi metodami. Izračunali bi kazalnike kakovosti, kot so popolnost statističnih rezultatov, standardna napaka, stopnja neodgovora enote, delež vstavljenih podatkov, pravočasnost prve objave, itd., za vsako dimenzijo kakovosti in statistično dokazali vpliv na kakovost. Delo bi lahko izboljšali tudi s poglobljeno analizo dodatnih dimenzij kakovosti, kot je npr. varnost. Le to smo bežno omenili, je pa dimenzija izrednega pomena v dobi varovanja informacij, številnih podatkov, digitalnega poslovanja ter podjetij, kjer je glavna dejavnost pridobivanje in objavljanje podatkov za poslovno odločanje.

Želimo si, da bi delo pripomoglo k zavesti o povezanosti poslovnih pravil, procesov in podatkovnega modeliranja v fazi razvoja informatijske rešitve. Le-ti pripomorejo h kakovosti informatijske rešitve in posledično h kakovosti podatkov.

LITERATURA IN VIRI

1. Abdel-Fattah, M. A., Khedr, A. E. & Aldeen, Y. N. (2017). An evaluation framework for business process modeling languages in healthcare. *Journal of Theoretical and Applied Electronic Commerce Research*, 15(5), 382–392.
2. Alizamini, F. G., Pedram, M. M., Alishahi, M. & Badie, K. (2010). Data quality improvement using fuzzy association rules. *2010 International Conference on Electronics and Information Engineering (ICEIE 2010) Data*, 1(Iceie), 468–472.
3. Bajec, M. & Krisper, M. (2005). A methodology and tool support for managing business rules in organisations. *Information Systems*, 30(6), 423–443.
4. Ben-Gan, I. (2012). *Microsoft Sql server 2012 T-SQL fundamentals*. Pearson Education.
5. Brackstone, G. (1999). Managing Data Quality in a Statistical Agency. *Survey Methodology*, 25(2), 139–150.
6. Cheng, R., Sadiq, S. & Indulska, M. (2011). Framework for business process and rule integration: A case of BPMN and SBVR. *Lecture Notes in Business Information Processing*, 87, 13–24.
7. Codd, E. F. (1983). A relational model of data for large shared data banks. *Communications of the ACM*, 26(1), 64–69.
8. Coronel, C., Morris, S. & Rob, P. (2013). *Database Systems: Design, Implementation and Management* (10. izd). Cengage Learning.
9. Davidson, L. & Moss, J. M. (2012). *Pro SQL Server 2012 relational database design and implementation*. Apress.

10. Dumas, M., La Rosa, M., Mendling, J. & Reijers, H. A. (2013). *Fundamentals of Business Process Management* (2.izd.). Springer.
11. Elmasri, R. & Navathe, S. B. (2016). *Fundamentals of Database Systems* (7. izd.). Pearson.
12. Eurostat. (2003). *Methodological documents - Definition of quality in statistics*. Pridobljeno 9. novembra 2022 iz <https://ec.europa.eu/eurostat/documents/64157/4373735/02-ESS-quality-definition.pdf/f0fdc8d8-6a9b-48e8-a636-9a34d073410f>
13. Eurostat. (2021). *European Statistical System handbook for quality and metadata reports*. Publications office of the European Union.
14. Fiorina, C. (2022). *What Is Data Modeling – Conceptual, Logical, Physical Models*. Pridobljeno 9. novembra 2022 iz <https://www.softwaretestinghelp.com/data-modeling-tutorial/>
15. Flowers, R. & Edeki, C. (2013). Business process modeling notation. *International Journal of Computer Science and Mobile Computing*, 2(3), 35–40.
16. Fredrick, F. T. (brez datuma). *General Knowledge of Sql Database*. https://www.academia.edu/33291724/GENERAL_KNOWLEDGE_OF_SQL_DATABASE
17. Godnov, U. (2012). *Opredelitev kakovosti podatkov in njeno zagotavljanje v relacijskem podatkovnem modelu poslovno informacijskega sistema*. Koper: Društvo za akademske in aplikativne raziskave.
18. Hoberman, S. (2009). *Data Modeling Made Simple: A Practical Guide for Business and IT Professionals* (2. izd.). Technics Publications, LLC.
19. Jakus, D. (2016). *Zagotavljanje Kakovosti Podatkov V Sistemih Poslovne Inteligence Na Primeru Študentskega Informacijskega Sistema Ekonomske Fakultete V Ljubljani* (str. 1–88). Pridobljeno 20. marca 2022 iz <http://www.cek.ef.uni-lj.si/magister/jakus4990.pdf>
20. Juan, Y.-C. & Yuan, K.-Y. (2013). Control Flow Pattern Recognition for Bpmn Process Models. *International Journal of Electronic Business Management*, 11(2), 133–143.
21. Kent, W. (1983). A simple guide to five normal forms in relational database theory. *Communications of the ACM*, 26(2), 120–125.
22. Kovačič, A. & Bosilj Vukšič, V. (2005). *Management poslovnih procesov*. GV Založba.
23. Kroenke, D. M. & Auer, D. J. (2010). *Database Concepts* (7. izd.). Upper Saddle River, NJ: Prentice Hall.
24. Krogstie, J. (2013). Perspectives to process modeling. In *Studies in Computational Intelligence*, 444, 1–39. Springer, Berlin, Heidelberg.
25. Kumar, V. & Reinartz, W. (2012). *Customer relationship management* (2. izd.). Berlin, Heidelberg: Springer.
26. Lam, G. S. W. (2020). Business Rules 101: What's a Business Rule? *Business Rules Journal*. Pridobljeno 13. junija 2022 iz <https://www.brcommunity.com/articles.php?id=c018>
27. Lee, R. G. & Dale, B. G. (1998). Business process management: A review and evaluation. *Business Process Management Journal*, 4(3), 214–225.

28. Li, Q. & Chen, Y.-L. (2009). *Modeling and Analysis of Enterprise and Information Systems*. Springer.
29. Lightstone, S. S., Teorey, T. & Nadeau, T. (2007). *Physical Database Design: The Database Professional's Guide to Exploiting Indexes, Views, Storage, and More*. Morgan Kaufmann Publishers.
30. Lindsay, A., Downs, D. & Lunn, K. (2003). Business processes - Attempts to find a definition. *Information and Software Technology*, 45(15), 1015–1019.
31. Man, Y., Wei, L., Gang, H. & Juntao, G. (2010). A novel data quality controlling and assessing model based on rules. *2010 Third International Symposium on Electronic Commerce and Security*, 29–32.
32. Masood-Al-Farooq, B. A. (2014). SQL Server 2014 Development Essentials. In *Packt Publishing*. Packt Publishing Ltd.
33. Navathe, B. S. (1992). Evolution of data modeling for Databases. *Communications of the ACM*, 35(9), 112–123.
34. Oliveira, P., Rodrigues, F. & Henriques, P. (2005). A formal definition of data quality problems. *Proceedings of the 2005 International Conference on Information Quality, ICIQ 2005*.
35. Olson, J. E. (2003). *Data Quality: The Accuracy Dimension*. Morgan Kaufmann Publishers.
36. OMG. (2011). *Business process model and notation (BPMN)*. Pridobljeno 17. septembra 2022 iz <https://www.omg.org/spec/BPMN/2.0/PDF>
37. Pipino, L. L., Lee, Y. W. & Wang, R. Y. (2002). Data Quality Assessment. *Communications of the ACM*, 45(4), 211–218.
38. Ross, R. G. (2003a). Business rules come first. *Computerworld*, 37(35), 30–31.
39. Ross, R. G. (2003b). *Principles of the Business Rule Approach*. Business Rule Solutions, LLC.
40. Ross, R. G. (2009). *Basic RuleSpeak Guidelines. Do's and Don'ts in Expressing Natural-Language Business Rules in English*. Business Rule Solutions LLC.
41. Ross, R. G. (2013a). *Decision Tables: A Primer - How to Use Table Speak*. Business Rule Solutions LLC.
42. Ross, R. G. (2013b). *Tabulation of lists in RuleSpeak: A primer using "The following" clause*. Business Rule Solutions LLC.
43. Rutar, K. & Tomažič, M. (2020). *Aktivno in neaktivno prebivalstvo. Metodološko pojasnilo* (interno gradivo). Ljubljana: Statistični urad Republike Slovenije.
44. Sebastian-Coleman, L. (2013). *Measuring data quality for ongoing improvement: a data quality assessment framework*. Morgan Kaufmann Publishers.
45. Seljak, R. (2017). *Metodološki priročnik: Kazalniki kakovosti 2017*. Ljubljana: Statistični urad Republike Slovenije.
46. Shamim, A., Hussain, H. & Shaikh, M. U. (2010). A framework for generation of rules from decision tree and decision table. *2010 International Conference on Information and Emerging Technologies, ICIET 2010*.

47. Sidi, F., Panahy, P. H. S., Affendey, L. S., Jabar, M. A., Ibrahim, H. & Mustapha, A. (2012). Data quality: A survey of data quality dimensions. *Proceedings - 2012 International Conference on Information Retrieval and Knowledge Management, CAMP'12*, 300–304.
48. Simsion, G. & Witt, G. (2005). *Data Modeling Essentials* (3. izd.). Amsterdam: Morgan Kaufmann.
49. Sörensen, K. & Janssens, G. K. (2003). Data mining with genetic algorithms on binary trees. *European Journal of Operational Research*, 151(2), 253–264.
50. Statistični urad Republike Slovenije (2017). *Kodeks ravnanja evropske statistike. Za nacionalne statistične organe in Eurostat (statistični organ EU)*. Pridobljeno 20. marca 2022 iz <https://www.stat.si/StatWeb/File/DocSysFile/10190/KODEKS.pdf>
51. Statistični urad Republike Slovenije. (2022a). *Politika varovanja informacij Statističnega urada Republike Slovenije* (interno gradivo). Ljubljana: Statistični urad Republike Slovenije.
52. Statistični urad Republike Slovenije. (2022b). *Pravilnik o varstvu osebnih podatkov zaposlenih na Statističnem uradu Republike Slovenije*. 1–13 (interno gradivo). Ljubljana: Statistični urad Republike Slovenije.
53. Statistični urad Republike Slovenije. (brez datuma). *Klasifikacije, vprašalniki in metode*. Pridobljeno 10. marca 2023 iz <https://www.stat.si/StatWeb/Methods/ClassificationsQuestionnairesMethods>
54. Teorey, T. J., Lightstone, S. S., Nadeau, T. & Jagadish, H. V. (2011). *Database modeling and design: Logical design* (5. izd.). Amsterdam: Morgan Kaufmann.
55. Teorey, T. J., Yang, D. & Fry, J. P. (1986). A logical design methodology for relational databases using the extended entity-relationship model. *ACM Computing Surveys (CSUR)*, 18(2), 197–222.
56. Tomažič, M. & Zaletelj, T. O. (2021). Statistični urad Republike Slovenije. *Aktivno in neaktivno prebivalstvo. Metodološko Pojasnilo*. Pridobljeno 20. marca 2022 iz <https://www.stat.si/StatWeb/File/DocSysFile/7788/07-008-MP.pdf>
57. Verbič, A. B. & Steenvoorden, T. (2017). *Smernice za zagotavljanje kakovosti*. Ljubljana: Statistični urad Republike Slovenije.
58. Wan-Kadir, W. M. N. & Loucopoulos, P. (2004). Relating evolving business rules to software design. *Journal of Systems Architecture*, 50(7), 367–382.
59. Wand, Y. & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86–95.
60. Wang, R. Y. & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33.
61. Wang, T. J., Du, H. & Lehmann, C. M. (2010). Accounting For The Benefits Of Database Normalization. *American Journal of Business Education (AJBE)*, 3(1), 41–52.
62. Wang, W., Sadiq, S. & Indulska, M. (2014). Factors affecting business process and business rule integration. *Proceedings of the 25th Australasian Conference on Information Systems, ACIS 2014*.

63. Zairi, M. (1997). Business process management : a boundaryless approach to modern competitiveness. *Business Process Management Journal*, 3(1), 64–80.
64. zur Muehlen, M. & Indulska, M. (2010). Modeling languages for business processes and business rules: A representational analysis. *Information Systems*, 35(4), 379–390.
65. zur Muehlen, M., Indulska, M & Kittel, K. (2008). Towards integrated modeling of business processes and business rules. *ACIS 2008 Proceedings - 19th Australasian Conference on Information Systems*, 690–699.

PRILOGE

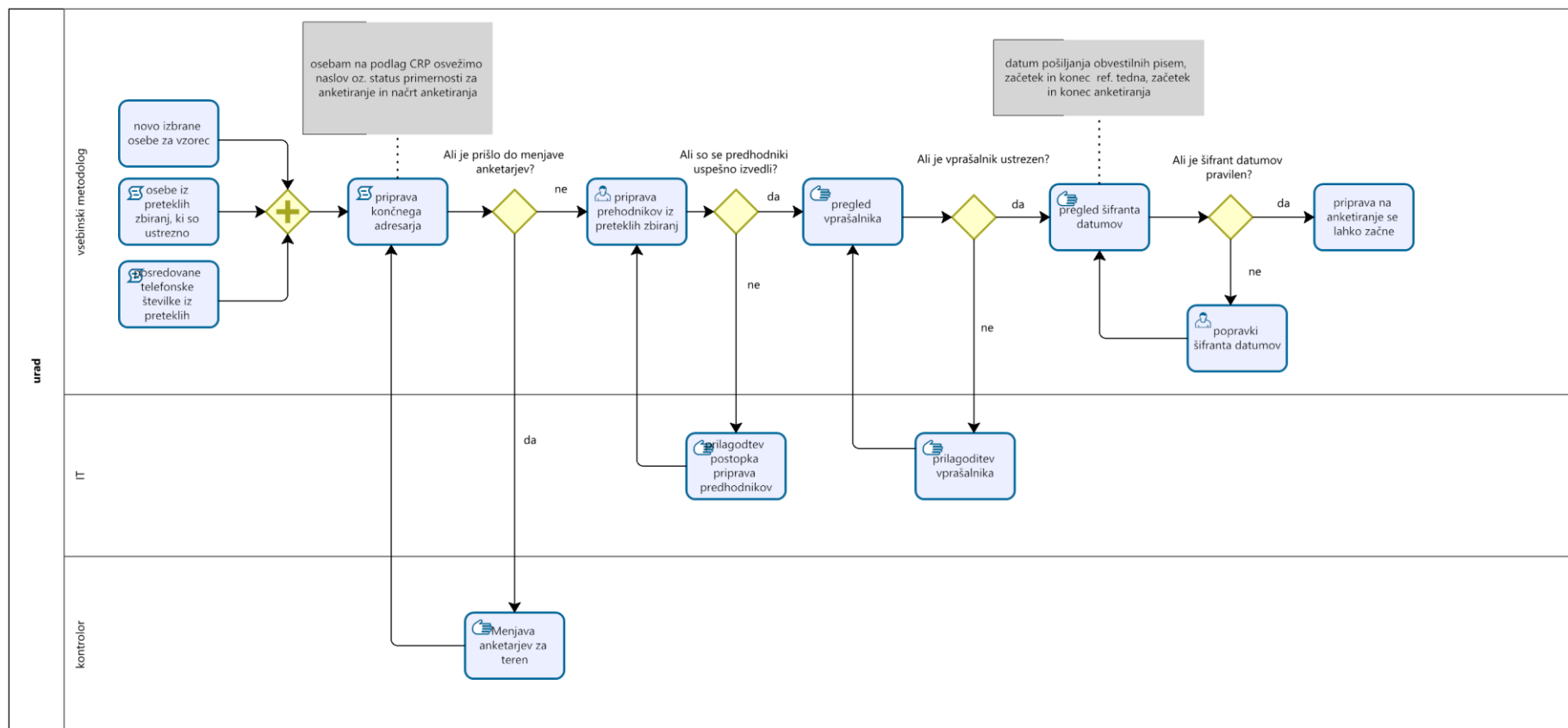
Priloga 1: Procesni model urada

PROCESNI MODEL

1	2	3	4	5	6	7	8
ANALIZA POTREB IN ZAHTEV	NAČRTOVANJE IN PRIPRAVA RAZISKOVANJA	IZBOR ENOT OPAZOVANJA	ZBIRANJE PODATKOV	OBDELAVA PODATKOV	ANALIZA PODATKOV	IZKAZOVANJE PODATKOV	DOKUMENTIRANJE IN EVALVACIJA RAZISKOVANJA
Ugotavljanje potrebe po podatkih	Načrtovanje sredstev in določitev seznama aktivnosti z roki	Prilava podatkovnih virov za izgradnjo vzorčnega okvira	Prilava na zbiranje	Urejanje administrativnih in drugih zbirk podatkov	Analiza časovnih vrst	Posodabljanje izhodnih podatkov	Izdelava dokumentacije o raziskovanju
Preučitev virov	Definiranje rezultatov raziskovanja	Prilava vzorčnega okvira	Prevzem administrativnih in drugih zbirk podatkov	Integracija različnih podatkovnih virov	Analiza ustreznosti ter potrditev rezultatov	Predstavitvev rezultatov	Zbiranje informacij za oceno kakovosti
Preverjanje metodologije	Prilava metodologije zbiranje podatkov	Izbor enot opazovanja	Zbiranje podatkov in komuniciranje s poročevalskimi enotami	Urejanje podatkov na mikroravni	Interpretacija rezultatov	Objavljanje	Izvedba ocene postopkov in procesov
	Prilava metodologije izbora enot opazovanja	Izdelava adresarja	Zajem podatkov	Vstavljanje podatkov (imputacije)		Podpora uporabnikom	
	Prilava metodologije obdelave podatkov			Uteževanje		Hramba statističnih mikropodatkov	
	Načrtovanje in testiranje izvedbe raziskovanja			Izračun statističnih ocen (agregacija)		Hramba agregiranih podatkov	
				Deflacija			
				Urejanje podatkov na makroravni			
				Tabelacija			
				Statistična zaščita podatkov			

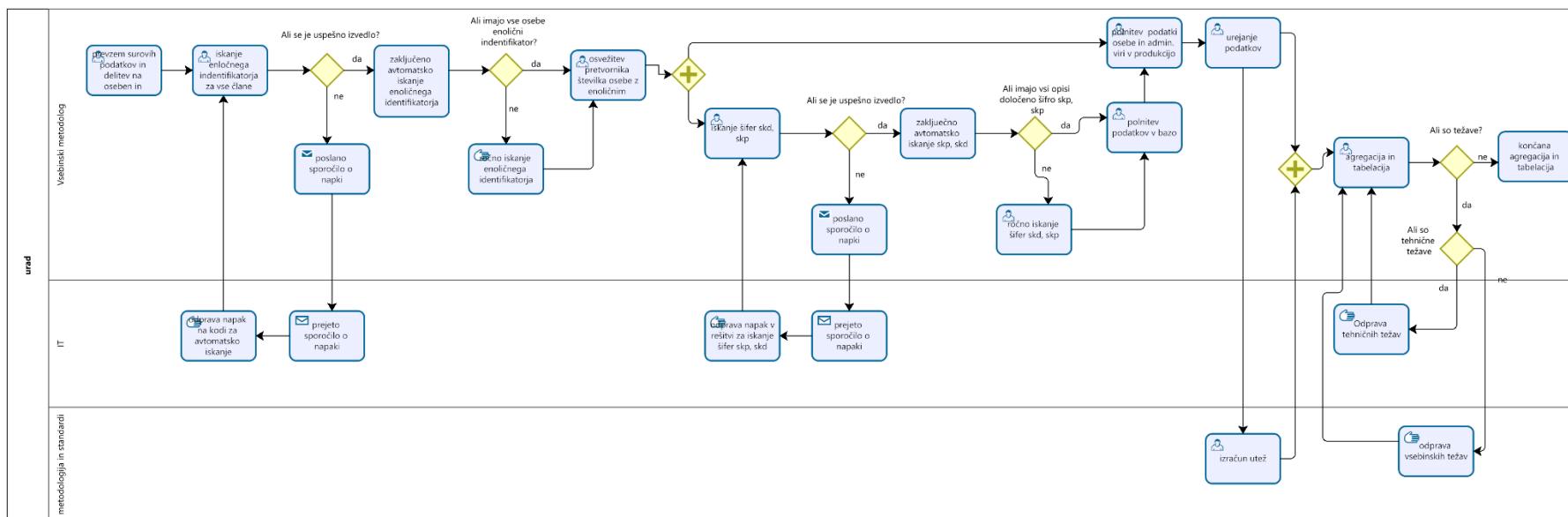
Vir: Statistični urad Republike Slovenije (brez datuma).

Priloga 2: Model podprocesov nove rešitve za pripravo na zbiranje



Vir: lastno delo.

Priloga 3: Model podprocesov nove rešitve za obdelavo podatkov



Vir: lastno delo.

