

Informacijos integravimas: poreikiai ir iššūkiai

Rimvydas Skyrius

Vilniaus universiteto Ekonomikos fakulteto profesorius,
Ekonominės informatikos katedros vedėjas
Vilnius University, Faculty of Economics,
Head of Economics Informatics Department, Professor
Saulėtekio al. 9, II rūmai, 309 kab.
El. paštas: Rimvydas.skyrius@ef.vu.lt

Saulius Šimkonis

Vilniaus universiteto Ekonomikos fakulteto
Ekonominės informatikos katedros doktorantas,
Vilnius University, Faculty of Economics,
Economics Informatics Department, Doctoral student
Saulėtekio al. 9, II rūmai, 309 kab.
El. paštas: simkonis@gmail.com

Indrius Sirtautas

Vilniaus universiteto Ekonomikos fakulteto Vadybos
informacinių sistemų bakalauro programos studentas,
Vilnius University, Faculty of Economics, Information
Systems Management program bachelor student
Saulėtekio al. 9, II rūmai, 309 kab.
El. paštas: indrius.sirtautas@gmail.com

Informacijos integravimas yra vienas iš svarbesnių iššūkių, su kuriais susiduria šiuolaikinė verslo informatika. Vieningos prasmės gavimą iš įvairialypės informacijos, esančios įvairiuose šaltiniuose ir formatuose, komplikuoja nemažai technologinių, sisteminių ir organizacinių iššūkių. Šio straipsnio tikslas yra išnagrinėti informacijos integravimo srityje egzistuojančias problemas ir pasiūlyti praktines jų sprendimo rekomendacijas. Esama nemažai informacijos integravimo požiūrių ir technologinių priemonių, kurias sieja siekiai valdyti ir sujungti informacijoje esančią prasmę, todėl akivaizdu, kad efektyvus informacijos integravimas turi būti vykdomas jungtinėmis vartotojų ir technologijos pastangomis. Sprendžiant informacijos integravimo problemas pravartu pasinaudoti gimininga patirties valdymo arba išmokyti pamokų srities problematika ir patirtimi.

Pagrindiniai žodžiai: informacijos integracija, informacijos poreikiai, išmokyti pamokos.

Įvadas

Informacijos integravimą galime laikyti vienu svarbesnių iššūkių, su kuriais susiduria šiuolaikinė verslo informatika

(Ananthanarayanan ir kt., 2013). Kai kurių šaltinių teigimu, stambiose organizacijose informacijos integravimo funkcijai tenka apie 40 proc. informacinių technologijų ir

sistemų biudžeto (Bernstein ir Haas, 2008). Informacijos integravimo poreikį sukuria dvi pagrindinės varomosios jėgos: duomenų ir neaktualios informacijos perteklius ir tuo pat metu reikalingos ir prasmingos informacijos trūkumas. Esama daugybė situacijų, kai reikia integruoti informaciją, dažniausiai iš skirtingose vietose esančių ir nesuderintų šaltinių. Visas šias situacijas vienija prasmės paieškos – siekis integruotos informacijos pagrindu gauti naujos, iki tol nežinotos ir naudingos informacijos. Tokios situacijos pavyzdžiais gali būti kliento informacijos analizė, kuri turi susieti įvairius su klientu susijusius duomenis – pardavimo, apmokėjimo, reakcijų į skatinimo akcijas, demografinius ir t. t., siekiant įvertinti, kokiai vertingumo kategorijai klientas priklauso. Tai gali būti ir turimos informacijos apie konkurentą analizė, kur iš atskirų fragmentuotų duomenų siekiama suvokti, kokių konkurento veiksmų galima tikėtis.

Kaip teigiama (A Single View..., 2002), bet kurioje veikloje informacijos perteklius atsiranda dėl prieinamo struktūzuotos ir nestruktūruotos informacijos kiekio. Norint išspręsti šią problemą, vartotojams reikia galimybės filtruoti ir personalizuoti informaciją, kuri reikšminga būtent jiems. Žinių trūkumas atsiranda dėl to, kad įvairių duomenų yra labai daug ir jie dažniausiai pateikiami be konteksto, yra išsisklaidę ir neturi lengvai suprantamos prasmės. Todėl vartotojai yra priversti rinkti šią informaciją iš kelių šaltinių ir neretai skirtingais formatais. Dažnai viename organizacijos sektoriuje yra informacijos, susijusios su kitame sektoriuje išskylančiomis problemomis, tačiau nėra tinkamos struktūros ta informacija dalintis ir ji reikiamu momentu nepasiekia reikiamo asmens. Jei vartotojas neturi informacijos, kaip integruoti esamus duomenis,

tuomet organizacijose bendradarbiavimas, darbo eiga, ypač tarp komandų, skyrių ar padalinių, yra toli gražu ne tokie, kokie galėtų būti. Kadangi informacija tarp padalinių nesidalijama, galima nepastebėti sąryšio tarp problemų, kylančių skirtingose organizacijos veiklose, taip atitolinant problemų sprendimą. Šio straipsnio tikslas – remiantis akademinių ir praktinių literatūros šaltinių analize išnagrinėti informacijos integravimo srities problemas ir pasiūlyti praktinių jų sprendimo rekomendacijų.

Informacijos integravimo samprata ir tikslai

Visose organizacijose naudojamus informacijos šaltinius galima suskirstyti į dvi grupes: struktūruotų duomenų šaltinius – pavyzdžiui, pardavimo eigos duomenis, ir nestruktūruotų duomenų saugyklas – teksto dokumentus, elektroninius laiškus, klientų atsiliepimus ir t. t. Tam tikriems veikloje kylantiems klausimams užtenka paieškos, vykdomos atskirai šiose duomenų saugyklose, tačiau sudėtingesni klausimai gali pareikalausiti įvairialypių duomenų apie dominančią problemą ar objektą iš struktūruotų ir nestruktūruotų šaltinių. Jei galima rasti būdų integruoti paiešką abiejų tipų šaltiniuose – struktūruotuose ir nestruktūruotuose, vertingesnės išsamios informacijos gavimo potencialas labai išauga (Pferd, 2010).

Struktūruotos informacijos integravimo priemonės žinomos jau seniai, nes didžioji šios informacijos dalis saugoma reliacinėse duomenų bazėse (DB), kurių informacijos radimo, derinimo ir pateikimo funkcijos abejonių nekelia. Kalbant apie struktūruotą informaciją negalima nepaminėti ir kito struktūruotos informacijos pavidalo – elek-

troninių lentelių. Gana daug informacijos saugoma darbo vietose *Microsoft Excel* lentelių pavidalu, tačiau šiose lentelėse esanti informacija dažniausiai nėra jokios informacinės sistemos dalis ir todėl integruojama gana sudėtingai. Kalbant apie nestruktūruotą informaciją dažniausiai turimi omenyje teksto formato dokumentai, nors čia taip pat priskiriama ir garso bei vaizdo informacija. Nestruktūruotos informacijos integravimas vykdomas tekstų apdorojimo priemonėmis, pavyzdžiui, tekstų analitikos įranga, kurios galimybės integruoti informaciją tokiu pat mastu, kaip tai atlieka reliacinių DB priemonės, yra ribotos. Literatūroje pastaruoju metu daugiausia akcentuojamas struktūruotos ir nestruktūruotos informacijos integravimas, pabrėžiant šių dviejų informacijos formatų paplitimą ir lūkesčius gauti naudą iš šio integravimo. Nemažai šaltinių teigia (Improve ..., 2014; Barrenechea, 2013; Pritchard, 2012), kad nestruktūruota informacija užima apie 80 proc. organizacijose saugomos informacijos, nors teisybės dėlei reikia pasakyti, kad nė viename iš jų neminimas bent vienas atliktas tyrimas, patvirtinantis šią dalį. Rinkoje pasirodė specifinės programinės platformos ir standartai nestruktūruotai informacijai valdyti; tarp jų galima paminėti Apache UIMA (angl. *Unstructured Information Management Architecture*), GATE (angl. *General Architecture for Text Engineering*), NLTK (angl. *Natural Language Toolkit*) ir kiti.

Kalbant apie informacijos integravimą jos poreikių požiūriu, integravimas reikalingas tenkinant tam tikrą specialų (dažniausiai, tačiau gali būti integruojama ir bendrajam stebėjimui) informacijos poreikį, gilinantis į tam tikrą situaciją. Informacijos integravimo tikslai šiuo atžvilgiu praktiškai

nesiskiria nuo kitų informacijos apdorojimo metodų ir technologijų – siekiama suvaldyti augančią duomenų gausą ir jos pagrindu vartotojams pateikti prasmingą ir naudingą informaciją. Remdamiesi (Skyrius 2008) pateikta informacijos poreikių klasifikacine struktūra galime pateikti kelis pavyzdžius, kai tam tikras informacijos poreikių tipas sukelia ir atitinkamą informacijos integravimo poreikį.

1. *Paprastojo* integravimo pavyzdys: kliento pageidavimas gauti nuolaidą susiejamas su kliento pirkimų ir mokėjimų istorija siekiant įvertinti, ar klientas tikrai yra tiek vertingas. Šiuo atveju informacija yra nuosava, prieinama ir lengvai suderinama.
2. *Sudėtingojo* integravimo pavyzdys: rinkoje stebimos perspektyvios technologinės inovacijos – naujos produktų versijos, galimi jų pakaitalai, šakos ir susijusių šakų inovacijos bei tyrimų rezultatai, kita galbūt susijusi informacija. Tokia informacija nepriklauso subjektui organizacijai, yra išsklaidyta ir įvairių formatų, jos šaltiniai gali būti iš anksto nežinomi.
3. *Bendrojo* integravimo pavyzdys: stebima visų konkurentų veikla ir elgsena rinkoje, derinant pardavimo, finansinę ir kitokią prieinamą informaciją.
4. *Specialiojo* integravimo pavyzdys: įdėmiai stebima konkretaus konkurento veikla, siekiant gerai suprasti, ką konkurentas ruošiasi daryti, ir prireikus parengti atsakomuosius veiksmus.

Integravimo tikslai gali būti gana paprasti, pavyzdžiui, susieti kelis susijusius informacijos fragmentus, tačiau jis gali leisti atsakyti ir į sudėtingesnius klausimus, reikalaujančius įvairialypės ir iš įvairių šal-

tinių gautos informacijos. Dažniausiai esti integravimo ašis – tema, kurios pagrindu informacija jungiama. Informacijos integravimo pavyzdžių gali pasitaikyti bet kur, o pats poreikis kilo gana seniai, yra sukurta ir IT pagrįstų priemonių – pavyzdžiui, duomenų bazėse esami duomenų integravimo mechanizmai. Tačiau didesnio dėmesio informacijos integravimas sulaukė būtent pastaruju metu. Priežastys yra kelios:

- organizacijoje „paslėpta“ potenciali įvairios vidinės informacijos vertė;
- „didžiųjų duomenų“ (angl. *Big Data*) problemos atsiradimas – plečiantis interneto informacinei erdvei, daugėja prieinamos ir potencialiai naudingos išorinės informacijos;
- išaugo nestruktūruotos informacijos vaidmuo;
- daug dėmesio ir intelekto reikalaujantis informacijos integravimo procedūras komplikuoja šaltinių, formatų, sutartinių terminų ir ženklų įvairovė;
- daugėja siekių automatiškai vykdyti informacijos integravimo procedūras.

Informacijos integravimo sukelti lūkesčiai ir iššūkiai

Šiuo metu dalis IT industrijoje su „didžiais duomenimis“ dirbančių įmonių nurodo dideles sąnaudas, susijusias su rankinio proceso dalimi, prieš paruošiant duomenis prasmingai analizei. Teigiama, kad tai „mitas, jog galima pateikti algoritmui nesutvarkytus duomenis ir gauti prasmingas išvagas“ (Lohr, 2014), ir nurodoma, kad pateiktų duomenų paruošimas analizei šiuo metu užima nuo 50 proc. iki 80 proc. jų darbo laiko. Tačiau tai atspindi dabartinę situaciją ir labiau parodo problemos dydį ir tolesnių tyrimų šioje srityje aktualumą. Skirtingi tyrėjai jau dabar siūlo tyrinėji-

mų kryptis, kuriomis dirbant būtų galima priartėti prie šios problemos sprendimo. Siūloma (Loshin, 2013) nustatyti verslo poreikių atvejus kartu su tikslais, motyvais ir sėkmės kriterijais, taip išlaikant vartotojų lūkesčius gauti naudą iš informacijos integracijos neperžengiant realistiškų rėmų. Šio straipsnio autorių nuomone, pagrindinės čia galinčios kilti problemos yra visų pirma prasmės sujungimas, kuris žmonėms vartotojams pavyksta gana nesunkiai, o technologijų srityje efektyvių prasmės valdymo mechanizmų nėra sukurta. Antra, raktinius terminus ar požymius integruojamoje informacijoje susiejančios schemos gali būti ir gana efektyvios, tačiau jei dalykinė sritis yra nestabili ir didelė jos dalis atsinaujina, tokios schemos gali sunkiai derintis prie šių pokyčių.

Informacijos integravimas paprastai siejamas su lūkesčiais, kad pagerės organizacijos darbuotojų informuotumo, o kartu veiklos ir ypač priimamų sprendimų kokybė. Autoriai siūlo išskirti šią galimą informacijos integravimo naudą:

- Sumažėja informacijos perteklius darbuotojams; mažėja „informacinio triukšmo“, daugėja kryptingos ir susijusios informacijos;
- Informacija pateikiama su kontekstu ir todėl labiau visavertė;
- Galima lengviau daryti išvadas įvertinus daugiau situacijos aplinkybių;
- Lengviau rasti sąryšius tarp problemų ir jas spręsti;
- Sutaupoma laiko; darbuotojai gali dirbti produktyviau.

Šių lūkesčių tenkinimas atrodo viliojama, tačiau reikia nepamiršti, kad informacijos integravimo atveju, kaip ir įgyvendinant daugelį kitų principų ir metodų, technologijos diegimas turi būti derinamas su reikiamomis organizacinėmis ir vady-

binėmis priemonėmis. Todėl įgyvendinant informacijos integravimo priemones laukia nemažai iššūkių, iš kurių galime išskirti dvi svarbiausias grupes – techninius ir organizacinius iššūkius.

Tarp labiausiai tikėtinų techninių iššūkių galime pateikti šiuos (A Single View..., 2002):

1. Reikiami duomenys neretai yra gaunami iš išorinių šaltinių (žinių, analitikų, įvairių internetinių svetainių ir t. t.). Tai sukelia sunkumų, nes informacija iš išorinių šaltinių dažniausiai yra nestruktūrizuota, jos formatas gali nesutapti su kompanijos turimos duomenų bazės formatu, taip pat, gali kilti pavadinimų dubliavimosi problema (nors dažniausiai duomenų bazės yra sutvarkytos taip, kad ši problema išsispresėtų automatiškai, be žmogaus įsikišimo).
2. Tik maža nestruktūrizuotų duomenų dalis jau yra integruota ir klasifikuota. Dauge lyje organizacijų struktūrizuoti ir nestruktūrizuoti duomenys būna integruojami tik tada, kai to prireikia konkrečiam departamentui, konkrečiai problemai spręsti. Taip tikriausiai įvyksta dėl to, kad įmonių vadovai nesupranta, kokią naudą gali atnešti kuo didesnės įmonei reikalingos informacijos dalies integravimas.
3. Gali nesutapti duomenų loginės struktūros. Ši problema kyla, nes dažniausiai nestruktūruoti duomenys būna pateikiami taksonomijose (angl. *Taxonomy*), kitaip sakant, klasifikuotų aplankų ir poaplančių sistemos pavidalu. Keblumas kyla siekiant taksonomiją suderinti su struktūruotos informacijos duomenų bazės schema (t. y. sudaryti loginius sąryšius).
4. Dar viena problema yra metaduomenų trūkumas. Dažnai nestruktūruota informacija neturi lengvai identifikuojamų metaduomenų, todėl ją yra daug sunkiau

klasifikuoti. Pavyzdžiui, jeigu kokia nors konkreti prekė neturi metaduomenų, o tik pavadinimą, gali būti, kad nebus aišku, prie ko ją priskirti ką tik atgabenus į sandėlį. O jei ta pati prekė turi metaduomenis (prekės rūšis, pvz., maisto prekė, porūšis, pvz., šviesi duona ir t. t.), ją klasifikuoti daug lengviau.

5. Kartais įmonėms yra reikalingi visiškai skirtingų formatų duomenys. Labai sunku susieti, pavyzdžiui, vaizdinį failą su tekstiniu ar garsiniu dokumentu, nes dažnai neužtenka juos tiesiog panašiai ar taip pat pavadinti dėl to, kad tokiu būdu bus susieti tik keli failai, o reikia, kad būtų susieta kiek įmanoma daugiau.
6. Bloga duomenų kokybė yra viena iš svarbiausių problemų taikant bet kurią duomenų integravimo strategiją. Duomenys turi būti tvarkingi prieš juos konvertuojant ir integruojant, kitaip juos integruojant tikrai kils sunkumų. Jei duomenys netikslūs, sugadinti ar dėl kokios kitos priežasties nekokybiški, tai jais remiantis padaryti sprendimai taip pat bus neteisingi (Challenges to ..., 2014).

Organizaciniai iššūkiai, su kuriais tenka susidurti integruojant informaciją, literatūroje sulaukia mažiau dėmesio, nors nemažai šaltinių teigia, kad būtent jie labiausiai veikia informacijos integravimo pastangų riziką. Dažniausiai minimi šie organizaciniai iššūkiai:

1. Duomenų nuosavybė organizacijos viduje. Nors formaliai visi organizacijoje esantys duomenys ir informacija priklauso organizacijai, atskiri padaliniai ir tarnybos gali nenorėti dalintis savo sunkiai surinktais duomenimis su kitais padaliniais. Tokiu būdu organizacijoje susiformuoja vertikalūs informaciniai barjerai.

2. Analitinės kultūros skatinimas ir horizontali informacijos sklaida. Būtent analitinė kultūra ir vertikalių informacinių barjerų įveikimas atveria kelius laisvai svarbios informacijos sklaidai, o tai savo ruožtu skatina proaktyvią pavojų ir progų diagnozę. Organizacinės kultūros ir efektyvaus turimos informacijos panaudojimo sąsają nurodė ir Trainor, teigdamas, kad skirtingi organizacinės kultūros tipai skirtingai skatina išminktų pamokų registravimą ir vėlesnį tokios informacijos panaudojimą (Trainor et al., 2008).
3. Standartų diegimas ir sąmoningas jų laikymasis. Nors ne visada lengva iš anksto įvertinti, kurios informacijos grupės bus tarpusavyje derinamos, o standartų diegimui neretai priešinamasi, tačiau nedidelis svarbiausių standartų kiekis galėtų būti lengvai įdiegtas ir apsaugoti nuo daugelio informacijos integravimo problemų ateityje.
4. Kintantis veiklos kontekstas, verčiantis tobulinti integravimo mechanizmus, o juose nustatyti integruojantys ryšiai gali būti sunkiai keičiami ar papildomi. Kaip ir standartų atveju, „kieti“ integruojantys ryšiai sunkiai tobulinami kintančioje veiklos aplinkoje, todėl projektuojant informacijos integravimo aplinkas reikia spręsti kebias tokių aplinkų lankstumo problemas. Vienas galimas sprendimo kelias pateiktas Indijos mokslininkų darbe (Ananthanarayanan ir kt., 2013).
5. Integruojamos informacijos šaltiniai gali turėti skirtingus ir prieštarigus informacijos reikalavimus – pavyzdžiui, konfliktuojantys laikotarpiai, nevienodas tų pačių objektų klasifikavimas ir t. t., o tai komplikuoja integruotos informacijos korektiškumą.
6. Informacijos integravimo sąnaudų įvertinimas. Tai nėra lengva, nes sudėtinga iš anksto tiksliai įvertinti, kiek kainuos darbo jėga, skirta planuoti, programuoti ir papildomiems duomenims rinkti, programinės ir techninės įrangos pirkimas, nenumatyti technologijų pokyčiai, duomenų saugojimas ir veikiančios sistemos palaikymas (Challenges..., 2013). Lengviausia įvertinti programinės bei techninės įrangos kainą, tačiau gana sunku apskaičiuoti, kiek kainuos sistemos palaikymas, darbuotojų mokymai ir kitos išlaidos, tiesiogiai nesusijusios su pačios sistemos pirkimu. Sąnaudų įvertinimą papildomai sunkina kintantis veiklos kontekstas, kai nelingva numatyti, kokius esamos sistemos pokyčius reikės atlikti, kokios bus su tuo susijusios sąnaudos ir kaip kis jos nauda pasikeitus aplinkai.

Informacijos integravimo ir išminktų pamokų sąryšis

Diegiant informacijos integravimo metodus ir technologijas organizacijoje, pravartu atsižvelgti į jau turimą patirtį integruojant informaciją tam tikrose specifinėse veiklos srityse. Viena iš tokių sričių gali būti patirties valdymo, arba kitaip – išminktų pamokų (angl. *lessons learned*) sritis. Nors šio tipo taikymai taip pat yra palyginti naujas dalykas, tačiau jau yra sukaupta tam tikros patirties išsaugant integruotą informaciją apie spęstas problemas, vykdytus projektus ar kitus didelius darbus. Informacijos integravimo ir išminktų pamokų sritys turi nemažai bendrumų. Pateiksime kelis pavyzdžius.

1. Ir vienu, ir kitu atveju organizacijoje panašiais principais vykdoma potencialiai vertingos informacijos ir patirties sklaida. Informacija surenkama ir pagal tam

tikrus principus agreguojama, siekiant išlaikyti svarbiausias prasmes ir jų kontekstą. Taip pat rūpinamasi jos sklaida ir prieiga suinteresuotiems asmenims.

2. Sėkmingas informacijos integravimo ar išmokyti pamokų sistemų diegimas ir naudojimas reikalauja panašios ar tokios pat organizacinės kultūros: reikalingi efektyviai veikiantys horizontalūs informaciniai ryšiai, kertantys atskirų tarnybų ribas.
3. Abiem atvejais panaši siekiamo rezultato „informacinė inžinerija“ – aktualios informacijos fragmentai sujungiami į naują prasmę.
4. Informacijos integravimo srityje galima pasinaudoti jau atliktais išmokyti pamokų srities problematikos tyrimais.

Kalbant apie išmokyti pamokų vaidmenį integruojant informaciją galima remtis LaBrosse siūlomu apibrėžimu, kuris apima ir procesą, ir rezultatą: „Išmokyti pamokų sesija yra komandos narių grupinis darbas, siekiant dokumentuoti, kas pavyko, kas nepavyko ir kas turi būti pakeista vykdant panašius projektus ateityje“ (LaBrosse, 2010). LaBrosse nuomone, šio proceso tikslas – „surinkta informacija ir patirtis, kuri gali būti panaudota kitiems projektams“.

LaBrosse apibrėžimas parodo dar vieną svarbų informacijos integravimo aspektą – nors visa informacija apie projektą žinoma visiems projekte dalyvavusiems asmenims, jiems reikia susirinkti kartu ir atlikti darbą grupėje, siekiant iš kiekvienam žinomos informacijos sukurti naują informaciją – pamoką, į kurią tikslinga atsižvelgti kitame projekte. Sukurta ir užrašyta tokia pamoka tampa informacijos integravimo objektu, kuris gali padėti priimti efektyvesnius sprendimus. Užrašymo, integravimo ir savalaikio tokios informacijos radimo iššūkius gerai apibūdino Barnes, nurodydamas, kad

„problema yra pateikti pamokas pakankamai glaustai, kad patrauktų užsienius žmonių dėmesį, ir kartu pateikti jas pakankamai specifiškai“ (Barnes ir Wearne, 1993).

Anksčiau minėtos išmokyti pamokos gali apimti skirtingo sudėtingumo integravimo atvejus. Jei pamoka jau yra užrašyta ir prieinama organizacijos informacinėje sistemoje, o ją taikyti paprasta – tai būtų paprasto integravimo pavyzdys. Jei organizacija ketina imtis naujo projekto ir planuoja pamokų paieškos sesiją, įtraukdama skirtingų skyrių darbuotojus, kurie gali turėti susijusios patirties, tačiau organizavimo metu neaišku, ar tikrai turėjo ir kokios, – turime sudėtingo integravimo pavyzdį.

Išmokyti pamokų informacija dažnai yra nestruktūruota, lydimą struktūruotų atributų, palengvinančių reikalingos pamokos suradimą. Čia susiduriama ir su „žmogiškojo veiksnio“ problema – turi būti atliktas tokios informacijos paieškos veiksmas. Kai kurie projektų valdymo metodai, pavyzdžiui, PRINCE2, labai aiškiai priskiria atsakomybę už išmokyti pamokų panaudojimą ieškančiajai pusei, teigdami, kad informacija apie išmokyti pamokas turi būti ieškoma, o ne pateikiama (OGC, 2009). Tai susiję ir su organizacijos kultūra, ir su individo supratimu apie savo supratimą (angl. *metacognition*). Tam, kad individas efektyviai pasinaudotų išmokyti pamokomis, jis turi:

- turėti pakankamai žinių, kad galėtų formuluoti korektišką klausimą;
- suprasti, kad šiuo klausimu jam reikia pagalbos, ir jaustis saugus organizacijoje, iškeldamas pagalbos poreikį;
- tikėjimo, kad jo aplinkoje pastangos rasti naudingos informacijos bus mažesnės nei surastos informacijos nauda.

Efektivaus patirties panaudojimo svarba. Informacijos integravimo proble-

matika yra sudėtinga, apimanti ir technines priemones bei procedūras, ir žmonių nuostatas, savęs vertinimą bei organizacinę kultūrą. Kaip jau minėjome, net jei organizacija turi patirties tam tikroje srityje, svarbu, kad jos būtų ieškoma (OGC, 2009). Čia susiduriame su veiksniais, kurie turi įtakos patirties panaudojimo laipsniui. Net jei integruota informacija jau egzistuoja įmonėje arba yra galimybė ją integruoti ir panaudoti, tai neužtikrina, kad ji bus panaudota tinkamu laiku ir tinkamoje situacijoje. Efektyviam panaudojimui būtinas tinkamas technologinių, organizacinių ir asmeninių veiksnių derinys. Technologinius ir organizacinius iššūkius jau aptarėme. Dabar atkreipsime dėmesį į kai kuriuos asmeninius veiksnius ir su jais susijusius iššūkius.

1. Pirmoji esamos patirties panaudojimo kliūtis yra individo supratimas apie savo žinių ribas, t. y. supratimas, kiek jis nesupranta, ir probleminių klausimų suformulavimas apie neapibrėžtas sritis. Tol, kol specialistas mano, kad jis viską žino apie iškeltą uždavinį, jam nėra poreikio skirti laiko informacijos paieškai, net jei tokia informacija yra pasiekiamą visiškai lengvai. Tokioje situacijoje papildomos informacijos paieška dažnai laikoma laiko gaišimu. Tai visiškai pagrįsta realioje aplinkoje, kur ištekliai (tiek laiko, tiek finansiniai) yra riboti, visos įmanomos užduotys niekada nebus atliktos ir reikia susitelkti į bent jau būtiniausių užduočių atlikimą ir svarbiausių sprendimų priėmimą. Taigi, neperėjus pirmojo filtro – atpažinimo, kad kažkas yra nežinoma, – tolesni veiksmai nebus atliekami.
2. Atpažinus, kad yra situacija, kuriai trūksta informacijos, gali prasidėti tolesnių alternatyvų vertinimas: kiek reikšminga gali pasirodyti nežinoma ar nesupranta-

ma sritis, kaip suformuluoti problemą, kiek laiko gali reikėti patikslinimams gauti, kiek tikėtina, kad patikslinimai bus vertingesni, nei jiems gauti įdėtos sąnaudos. Kiekvienas iš šių klausimų gali nutraukti patirties paieškos ir informacijos integravimo pastangas.

3. Neapibrėžtos srities reikšmingumo suvokimas suprantamas kaip galimos grėsmės ar galimybės dydis. Realioje valdymo aplinkoje visada bus detalių, kurios nėra visiškai aiškos, tačiau jei sprendimą priimančias asmuo įvertintų, kad neapibrėžtos srities grėsmė arba galimybė yra pakankamai nedidelė, jis neturėtų skirti dėmesio detaliai nagrinėjimui dėl mūsų jau minėtų ribotų išteklių. Čia svarbu atkreipti dėmesį, kad kalbama ne apie tikrąją grėsmę ar galimybę, o apie jų suvokimą, kuris priklauso nuo sprendimą priimančio asmens kvalifikacijos, patirties bei daugelio visiškai subjektyvių veiksnių. Tad svarbu įdiegti organizacines priemones ir kultūrą, padedančias iškelti bent pačius svarbiausius neapibrėžtumo atvejus, minimizuojant individualių veiksnių įtaką pavojingiausioms ar labiausiai žadančioms situacijoms.
4. Nepakanka identifikuoti, kad turimos informacija ar patirties nepakanka. Reikia suformuluoti integracijos poreikį arba klausimą, leidžiantį ieškoti patirties arba integracijos šaltinių. Siekiant sėkmingai įveikti šį barjerą svarbi jau anksčiau minėta patirtis ir kvalifikacija, tačiau taip pat labai svarbi organizacinė kultūra, kuri skatintų ieškoti ir klausti nebausdama už nežinojimą.
5. Jau suformulavus informacijos integracijos poreikį ar patirties trūkumą, sprendėjas turi įvertinti, kokiais būdais jis galėtų šį poreikį patenkinti. Čia tampa svarbios

technologinės priemonės bei organizacinis jų palaikymas (duomenų nuosavybė organizacijos viduje, standartų diegimas ir laikymasis). Jei numatomos integravimo poreikio ar patirties paieškos sąnaudos yra didelės, tai neigiamai veikia ketinimus tęsti integravimo ar patirties paieškos sąnaudas. Sprendėjas taip pat vertina tokių pastangų galimą naudą dydį, t. y. ar skyrus pastangas klausimas bus visiškai išspręstas, ar neapibrėžtumas bus sumažintas tik nedideliu mastu.

Pirmiau pateikti veiksniai iš dalies įvardija ir informacijos integravimo proceso etapus. Tai reiškia, kad esant neefektyviam ankstesnio veiksnio sprendimui, tolesnių veiksmų stiprinimas yra neefektyvus išteklių švaistymas. Pavyzdžiui, net jei organizacija turi puikių informacijos integravimo ir patirties dalijimosi priemonių, jos neturės jokios įtakos, jei specialistas nematys poreikio jomis naudotis ar manys, kad jo žinios ir taip pakankamos. Peržvelgus šiuos etapus, arba „filtrus“, galima teigti, kad informacijos integravimo bei patirties panaudojimo efektyvumą lemia ne stipriausia, o būtent silpniausia grandis iš organizacinių, techninių ir asmeninių veiksmų. Tolesniuose tyrimuose būtų tikslinga patikrinti šį nuoseklumą įvertinant, ar tikrai kiekviename kitame etape galimas tik mažesnis informacijos integravimo ir patirties panaudojimo poreikis.

Informacijos integravimo technologijos

Yra sukurta nemažai technologijų platiems ir įvairiems informacijos integravimo poreikiams vienaip ar kitaip tenkinti. Svarbia aplinkoje galime laikyti tai, kad bet koks integravimas reikalauja suderinamumo ir

tam tikrų susitarimų ar standartų naudojimo. Susitarimų ar standartų formos gali būti įvairios: duomenų bazėse tai yra duomenų aibių raktai, dokumentų valdymo sistemoje – rakto žodžiai ar specifiniai terminai, kitokiose sistemose – įvairūs metaduomenys, žymos ir t. t.

Trumpai apžvelgiant svarbiausias informacijos integravimo technologijas, pirmiausia reikia paminėti duomenų saugyklas, arba duomenų sandėlius, tarptautiniuose šaltiniuose įvardijamus anglišku terminu *Data Warehouse*. Kaip žinoma, pagrindinė duomenų sandėlio paskirtis yra organizacijos veiklos duomenų kaupimas ir saugojimas integruotu pavidalu, siekiant visų pirma tenkinti analitinius veiklos valdymo poreikius. Šiems poreikiams skirtos analitinės užklausos paprastai būna gana sudėtingos, sudarytos iš nemažo parametų skaičiaus, o rezultatų kokybė priklauso ir nuo aiškaus užklausos formulavimo, ir nuo gero technologinių priemonių išmanymo. Duomenų saugyklos naudoja tą pačią duomenų integravimo technologiją, kaip ir duomenų bazės – duomenys siejami pasitelkiant rakto rekvizitus. Kitaip sakant, šis būdas yra daugiau duomenų integravimo būdas.

Verta paminėti kelis kitus duomenų ir informacijos integravimo būdus.

1. *Rankinis duomenų integravimas* (angl. *Manual data integration*). Rankinio duomenų integravimo metu vartotojai tiesiogiai sąveikauja su duomenų šaltiniais ir programinės įrangos sistemomis, siekdami integruoti duomenis (Tsierkezos, 2010). Taikant šį integravimo būdą nenaudojamos automatizuotos integravimui skirtos programos; svarbiausias vaidmuo tenka ne techninei ar programinei įrangai, o darbuotojui. Konkretūs šio būdo įgyvendinimo va-

riantai gali būti įvairūs: rankinis naujų duomenų ir informacijos skirstymas į katalogus pagal kategorijas (pavyzdžiui, pirkimas, pardavimas, gamyba ir t. t.); metaduomenų panaudojimas susiejant atskiras tą pačią temą ar objektą liečiančias duomenų aibes; rankinis duomenų iš kelių duomenų bazių susiejimas; integravimas naudojant SQL kalbos galimybes. Rankinis integravimas yra nebrangus ir gana gerai tinka ten, kur integruojami palyginus nedideli duomenų kiekiai, tačiau darbas yra lėtas ir neefektyvus, taip pat dėl rankinių operacijų išauga žmonių klaidos tikimybė.

2. *Duomenų integravimas naudojant bendrą vartotojo sąsają* (angl. *Common user interface*). Naudojant bendros vartotojų sąsajos integravimo metodą, vartotojui yra pateikiamas jau pažįstamas pateikimo pavidalas, dažniausiai naršyklės tipo. Jame duomenys iš susijusių informacinių sistemų yra parodomi atskirai, taigi juos suvienodinti ir integruoti paliekama vartotojams (Ziegler ir Dietrich, 2007). Pagrindinis šio būdo skirtumas nuo paprasto mechaninio duomenų integravimo yra tas, kad visus reikiamų duomenų šaltinius sujungia viena programinė paieška; vartotojui nereikia pačiam naršyti po skirtingus aplankus ar duomenų bazes, dėl to reikiama informacija randama greičiau. Kaip ir populiariausiuose paieškos varikliuose (pvz., *Google.com*, *www.bing.com* ar *Yahoo!*), vartotojas gali įvesti paieškos frazę ir pagal ją gauti rezultatus. Šis būdas neblogai siejasi su pasiūlyta (Franklin ir kt., 2005) duomenų erdvės idėja, kai aktualūs tam tikrai problemai duomenys pateikiami viename matomame lauke, o jų prasmės integracija ir sudėtingesni analizės būdai

paliekami vartotojo iniciatyvai. Panašiai siūloma (Holzinger ir kt., 2013) vienoje sistemoje jungti natūralios kalbos apdorojimo, gilios nestruktūrotų duomenų analizės ir žmogaus eksperto aktyvaus dalyvavimo technologijas. Šis būdas yra gana lankstus ir santykinai nebrangus, kadangi nereikalauja papildomos įrangos, tačiau netinka darbui su dideliais informacijos kiekiais ir sudėtingomis užklausomis.

3. *Duomenų integravimas per taikomąją programą* (angl. *Integration by application*). Čia integravimui naudojamos programos, kurios turi prieigą prie skirtingų duomenų šaltinių, naudojančių tarpusavyje suderinamas technologijas ir vartotojui pateikia jau integruotus rezultatus (Ziegler ir Dietrich, 2007). Šis būdas remiasi integruojančio tarpininko metodu (angl. *Wrapper Mediator Method*). Naudojamos programos veikia kaip tarpininkai, surenka reikiamą informaciją pagal užklausą iš kelių šaltinių ir ją iš karto standartizuoja. Kadangi integravimas atliekamas automatiškai, darbas vyksta gana efektyviai, tačiau šis būdas netinka, kai duomenis reikia integruoti iš didelio šaltinių skaičiaus, nes kiekvienam šaltiniui gali reikėti skirtingų algoritmų duomenims konvertuoti į vieningus, o tai labai sunkina programos darbą.
4. *Duomenų integravimas naudojant tarpinę programinę įrangą* (angl. *Mediator arba Middleware*). Tarpinė programinė įranga – tai „programinė įranga, atliekanti tarpininko vaidmenį keičiantis duomenimis tarp skirtingų programų, sistemų, platformų, kai tiesioginis keitimasis tarp jų neįmanomas arba nėra sklandus“ (Dagienė ir kt., 2014). Tarpinės programinės įrangos sistemos

duomenų bazėse gali imti duomenis iš daug skirtingų informacinių sistemų arba duomenų bazių ir iš jų vartotojui pateikti vieną integruotą duomenų schemą. Tačiau, kitaip nei integruojant duomenis per taikomąją programą, yra galimybė rinkti informaciją iš paprastomis programomis nesuderinamų šaltinių – pavyzdžiui, šaltiniai naudoja visiškai skirtingas operacines sistemas. Šis būdas leidžia integruoti duomenis iš šaltinių, kurie yra sunkiai suderinami, ir pateikti sudėtingas užklausas, tačiau tam reikalingi aukštos kvalifikacijos specialistai ir brangi programinė įranga, kurią gali tekti tobulinti susidūrus su naujais duomenų šaltiniais.

5. *Suvienodintos duomenų prieigos metodas* (angl. *Uniform data Access*). Naudojant šį metodą duomenys integruojami įvedimo lygyje (Ziegler ir Dietrich, 2007). Programinė įranga, pasinaudodama tam tikra pasirinkta logika, susiskirsto duomenis pagal tipus. Paskui visos užklausos vykdomos virtualioje aplinkoje, nebetikrinant fizinių duomenų. Kitomis programomis galima keisti, pridėti, trinti ar kaip kitaip modifikuoti duomenis, esančius fiziniuose laikmenose, bet tai nedarys jokios įtakos integruotai informacijai. Šis būdas palaiko praktiškai visus duomenų tipus, taip pat jį naudojant galima labai greitai atlikti pačias sudėtingiausias užklausas, bet jis nepalaiko naujų duomenų įvedimo (Tsierkezos, 2010). Norint, kad būtų integruoti atnaujinti duomenys ar pridėta informacija iš naujo šaltinio, viską reikia daryti iš naujo, t. y. programa iš naujo turės išanalizuoti visus duomenis. Naudojant šį būdą galima integruoti bet kokių tipų duomenis, naudoti sudėtingas užklausas, kurios atliekamos labai greitai. Tačiau

būdas nepalaiko duomenų atnaujinimo ir naujų šaltinių įtraukimo.

Nemažai šaltinių (pavyzdžiui, Loshin, 2013; Doan ir kt., 2009), nagrinėjančių informacijos integravimo problematiką, siekia pasiūlyti vienokią ar kitokią informacijos integravimo schemą, savo esmę artimą duomenų modeliui duomenų bazių valdymo sistemose, ir tokiu būdu reikšmingai automatizuoti integravimo procedūras. Siūloma nestruktūrotų duomenų valdymo sistema, kuri iš nestruktūrotų duomenų išgauna struktūras ar struktūruotus duomenų fragmentus – vardus, adresus, telefono numerius; jų pagrindu kuriamos duomenų bazės, kurios palengvina informacijos paiešką. Kitaip sakant, nestruktūroti duomenys spraudžiami į struktūrotą schemą, kurios veikimas žinomas ir daugiausia gali būti vykdomas automatiškai. Tačiau tokiu būdu informacijos integravimas suprimityvinamas, redukuojant jį į raktinių terminų atitikmenis analogiškai reliacinėms duomenų bazėms. Standartiniai tokiu būdu išgaunami rekvizitai – vardai, adresai, telefono numeriai – neapima didelės dalies norimų integruoti duomenų, kurie gali būti daug laisvesnės formos nei pavieniai rekvizitai (pavyzdžiui, konkurento veiklos paminėjimas įvairiuose kontekstuose).

Įdomus informacijos integravimo principas pateikiamas (Ananthanarayanan ir kt., 2013), kur nestruktūruotoje informacijoje atskleidžiamas temos panašumas sudarant vadinamąjį dokumento atspaudą (angl. *signature*). Pirmiausia nustatomas dokumento turinio tipas – skaičiai, grynas tekstas, tekstas su skaičiais ar įvairūs tarpiniai variantai. Tada išgaunamas dokumento atspaudas skaidant jo turinį į simbolių bigramas, trigramas, žodžių unigramas ar bigramas, posakius, žodžių sekas ir t. t. Tuo remiantis skaičiuojamas

dokumentų panašumo laipsnis, dokumentai grupuojami į kategorijas ar klasterius. Šių procedūrų metu dokumento prasmė nei tirama, nei generuojama, tačiau klasterizavimo požymiai lyginami su etaloniniu duomenų rinkiniu. Rinkinys sudaromas rankiniu būdu, naudoja vartotojų nuožiūra įtraukiamus rakto žodžius arba naudoja sudarytas ontologijas. Šiuo būdu integruota informacija leidžia atsakyti į specifinius klausimus, pavyzdžiui, „parodyti visus kliento X nusiskundimus, gautus iš visų šaltinių“, ir centralizuoti informacijos vaizdą apie konkretų objektą – klientą, bendrovę ar kitą objektą. Taip galima pasiekti gana lankstaus reagavimo į pokyčius dalykinėje srityje.

Kiekvienas duomenų integravimo būdas turi savo pranašumų ir trūkumų. Vienas gali labiau tikti vienai organizacijai ar veiklai, bet visiškai netikti kitai. Jeigu organizacija yra didelė ir jai reikia apdoroti milžiniškus kiekius duomenų, kurie nuolatos turi būti atnaujinami, taip pat reikia daryti daug užklausų ir atskaitų, tikriausiai geriausias integravimo metodas – duomenų sandėliai. Jeigu duomenų nereikia dažnai atnaujinti, o daugiau dirbti su jau turimais, tinkamiausia naudoti suvienodintos duomenų prieigos metodą. Jei įmonė maža ar vidutinio dydžio ir dirba atitinkamai su mažu arba vidutiniu duomenų kiekiu, tikslingiausia būtų atsižvelgiant į pajamas ir poreikius rinktis vieną iš mechaninio integravimo metodų variantų, jei įmonė gali skirti mažai lėšų, arba dirba su labai mažai duomenų, arba naudoti bendros vartotojų sąsajos metodą, jei įmonė gali investuoti daugiau ir nenori, kad darbuotojai gaištų laiką ieškodami informacijos. Taip pat gali tikti ir duomenų integravimo per taikomąją programą metodas arba duomenų integravimo naudojant tarpinę programinę įrangą metodas, jeigu įmonė gali išleisti

daugiau pinigų programinei įrangai bei kvalifikuotiems darbuotojams samdyti.

Viena iš alternatyvų integruojant kompanijos duomenis – įsigyti jau sukurta ir paruoštą naudoti programinę informacijos integravimo įrankį. Duomenų integravimo įrankis (angl. *Data integration tool*) – tai programa arba programų paketas, padedantis vartotojui integruoti informaciją ir apskritai apdoroti įvairių tipų duomenis taip, kad iš jų verslas gautų kuo daugiau naudos. *Gartner* analitikai sudarė, jų nuomone, funkcionaliausius duomenų integravimo įrankius kuriančių gamintojų sąrašą su gana detaliais kiekvieno produkto įvertinimais, apžvelgdami ne tik programų pranašumus, silpnybes, bet ir vartotojų atsiliepimus (Thoo ir kt., 2014). Šiame sąrašė paminimi tokie gamintojai kaip *IBM*, *Informatica*, *SAP*, *Oracle*, *SAS*, *Microsoft* ir kiti. Alternatyva pasirinkti vieną iš siūlomų programinių informacijos integravimo įrankių yra svarstyti pirmiausia dėl to, kad nusipirkti ir įdiegti jau parengtą programą ar jų paketą užtrunka daug trumpiau, nei kurti savo sistemą. Dar vienas pranašumas yra tas, kad tokia programinė įranga yra ekspertų sukurta ir išbandytas produktas, apie kurį galima rasti nemažai informacijos (vartotojų atsiliepimai, ekspertų įvertinimai – pavyzdžiui, minėtasis *Gartner* analitikų darbas), kuri padeda apsispręsti renkantis informacijos integravimo įrankį. Be abejo, pasirinkimas pirkti informacijos integravimo programinę įrangą turi ir trūkumų. Visų pirma, universali įranga tikriausiai nebus idealiai pritaikyta prie organizacijos poreikių; be to, ji gali būti sunkiai suderinama su kita joje naudojama programine įranga ir dažnai kainuoti brangiau, nei pačių kurtas sprendimas. Tačiau jei organizacija neturi reikiamos kvalifikacijos specialistų,

kuriems leistų kurti informacijos integravimo sistemą, arba jei jos veikla ir poreikiai yra gana standartiniai (pavyzdžiui, prekių perpardavimas), tuomet pirkti informacijos integravimo įrankį gali apsimokėti labiau nei kurti jį savo jėgomis.

Išvados ir pasiūlymai

Informacijos integravimo problema neišvengiamai iškyla, jei vartotojai iš savo organizacijoje veikiančios informacinės sistemos ar sistemų siekia gauti platesnio konteksto, o kartu ir vertingesnę informaciją. Tai darant tenka susidurti su keliomis probleminėmis situacijomis, kai reikia priimti konkrečius veiklai labiausiai tinkamus integravimo sprendimus.

Visų pirma reikia nuspręsti, kokio apibrėžtumo integracija reikalinga veiklai, t. y. kokie duomenys ar informacija bus siejami, iš kokių šaltinių ir kokiomis priemonėmis. Jei siekiama paprasto duomenų susiejimo duomenų bazės ar duomenų saugyklos sąlygomis, tada integracijai bus naudojamos paprastos sisteminės priemonės – reliaciniai ryšiai, raktai ir jų variacijos. Jei siekiama susieti nestruktūruotą informaciją arba ją derinti su struktūruotais duomenimis, tuomet bus reikalinga teksto analizė ieškant susijusių terminų. Gali būti siekiama sieti informaciją pagal sudėtingesnes savybes, pavyzdžiui, emocijas – tarkime, reikalinga klientų neigiamų atsiliepimų analizė siekiant nustatyti pagrindines tokios reakcijos priežastis; tuomet reikalinga iš esmės kitokia įranga, kurios nemaža dalis dar yra eksperimentinės plėtros stadijos.

Antra, reikia apsispręsti, ar informacijos integravimas turi būti automatinis, ar automatizuotas; pirmuoju atveju visas ar didžiąją dalį procedūrų atlieka technologija,

o antruoju procedūros paskirstomos tarp vartotojo ir sistemos. Visiškai automatinio atveju išgautą iš informacijos požymių pagrindu sukuriamą reliacinę ar kitokia schema, kuri vėliau naudojama įprastais būdais; šiuo atveju labai tikėtina, kad bus prarasta didžioji informacijos turtingumo dalis. Automatizuotu paramos atveju galutinį sprendimą dėl prasminių sąsajų priima vartotojas. Tokios integruojančios technologijos pavyzdžiu gali būti (Pferd, 2010) pateikta informacijos integravimo prenumeratos idėja, šiek tiek primenanti neseniai dar buvusias populiarias naujienų sindikavimo technologijas RSS ir kitas.

Trečia, renkantis duomenų integravimo būdą apsispręsti gali verstis ir riboti piniginiai ištekliai. Čia svarbu atsižvelgti į įmonės dydį, veiklos sritį, galimą skirti pinigų bei laiko kiekį, apdorojamų duomenų ir jų formatų kiekį. Jeigu kompanija šiuo metu gali skirti daugiau lėšų, tačiau nedaug laiko, tikriausiai jai būtų geriausia rinktis vieną iš jau sukurtų duomenų integravimo įrankių. Perkant jau sukurtą informacijos integravimo įrankį, pradinės išlaidos tikriausiai bus didesnės, tačiau laikui bėgant specialiai sukurta sistema turėtų labiau atsipirkti. Taip pat jau sukurtą įrankį rinktis reikėtų tuo atveju, kai kompanija neturi pakankamai kvalifikuotos darbo jėgos arba galimybės ją pasamdyti kokybiškai integravimo sistemai sukurti. Kalbant apie įmones, kurios dirba su mažais duomenų kiekiais ir mažai skirtingų duomenų tipų, joms būtų geriausia pasirinkti pigiausią variantą, t. y. vieną iš mechaninio integravimo metodo būdų. Pačioms didžiausioms kompanijoms, turinčioms labai daug tiek vidinių, tiek išorinių duomenų šaltinių (savaime aišku, ir daug duomenų) tinkamiausias pasirinkimas dažniausiai yra kurti duomenų sandėlį. Žinoma, tai sudėtingas,

ilgas ir nemažų išlaidų reikalaujantis procesas, tačiau ilguoju laikotarpiu tinkamai naudojamas duomenų sandėlis turėtų atsipirkti ir atnešti naudos. Matyt, sunkiausia tinkamą sprendimą pasirinkti organizacijoms, dirbančioms su vidutiniu duomenų kiekiu, nes labai sunku surasti tą ribą, kada jau geriausias sprendimas yra kurti duomenų sandėlį, o kada galima pasitenkinti ir kuriuo nors iš paprastesnių sprendimų – integravimą per taikomas programas, naudojant tarpinę programinę įrangą, ar netgi vieną iš mechaninio integravimo būdų.

LITERATŪRA

ANDRADE, J. ; ARES, J. ; GARCIA, R. ; PAZOS, J. ; RODRIGUEZ, S. ; RODRIGUEZ-PATON, A. ; SILVA, A. (2006). Towards a lessons learned system for critical software. *Reliability Engineering and System Safety*, no. 92, p. 902–903.

ANANTHANARAYANAN, R.; BALAKRISHNAN, S.; REINWALD, B.; YEE, Y. (2013) *Unstructured information integration through data-driven similarity discovery*. IBM India research lab.

BARNES, N. M.; WEARNE, S. H. (1993). The future for major project management. *International Journal of Project Management*, no. 11(3), p. 135–142.

BARRENECHEA, M. (2013). Big Data: Big Hype? [interaktyvus] [žiūrėta 2014 m. rugsėjo 15 d.] Prieiga per internetą: <<http://www.forbes.com/sites/ciocentral/2013/02/04/big-data-big-hype/>>.

BERNSTEIN, P. A.; HAAS L. (2008). Information Integration in the Enterprise. *Communications of the ACM*, no. 51 (9), p. 72–79.

Challenges to Data Integration (2014). US Department of Transportation [interaktyvus] [žiūrėta 2014 m. rugsėjo 15 d.] Prieiga per internetą: <<http://www.fhwa.dot.gov/asset/dataintegration/if10019/dip06.cfm>>.

Dagienė, V.; Grigas, G.; Jevsikova, T. (2014) Enciklopedinis kompiuterijos žodynas [interaktyvus] [žiūrėta 2014 m. rugsėjo 15 d.]. Prieiga per internetą: <<http://ims.mii.lt/EK%C5%BD/>>.

DOAN, A.; NAUGHTON, J. F.; BAID, A.; CHAI, X.; CHEN, F.; CHEN, T.; CHU, E.; DER-ROSE, P.; GAO, B.; GOKHALE, C.; HUANG, J.; SHEN, W.; VUONG, B. (2009). The Case for a

Apibendrinant galima konstatuoti, kad informacijos integravimo problematika, būdama ne visai nauja, pastaruoju metu sulaukia didėjančio tyrėjų ir praktikų dėmesio, tačiau siekiai automatizuoti integravimo ir prasmės sudarymo procesus kol kas reikšmingesnių rezultatų nedavė. Akivaizdu, kad šiuo metu labiausiai realistiškas požiūris į informacijos integravimą turi remtis efektyviu funkcijų paskirstymu ir sąveika tarp technologijos ir vartotojo, todėl ir vartotojų lūkesčiai greitai sukurti vertę iš informacijos integravimo vertintini atsargiai.

Structured Approach to Managing Unstructured Data. In *CIDR-09. 4th Biennial Conference on Innovative Data Systems Research (CIDR)*. January 4–7, 2009, Asilomar, California, USA.

FRANKLIN, M.; HALEVY, A.; MAIER, D. (2005). From Databases to Dataspaces: A New Abstraction for Information Management. *ACM SIGMOD Record*, vol. 34, no. 4, December 2005, p. 27–33.

HOLZINGER, A.; STOCKER, C.; OFNER, B.; PROHAZKA, G.; BRABENETZ, A.; HOFMANN-WELLENHOF, R. (2013). Combining HCI, Natural Language Processing, and Knowledge Discovery – Potential of IBM Content Analytics as an Assistive Technology in the Biomedical Field. *HCI-KDD 2013*, p. 13–24.

Improve decision making with content analytics (2014). [interaktyvus] [žiūrėta 2014 m. rugsėjo 15 d.]. Prieiga per internetą: <http://www.ibm.com/smarterplanet/us/en/business_analytics/article/it_business_intelligence.html>.

KERZNER, H. (2009). *Project Management. A Systems Approach to Planning, Scheduling and Controlling*. 10th edition. New York: John Wiley & Sons, Inc.

LABROSSE, M. (2010). Documenting the lessons learned for every project. *Employment Relations Today* (Wiley), no. 37(1), p. 93–97.

LOHR, S. (2014). As big data is mined for insight, manual input still creates a bottleneck. *The International New York Times*, August 18, p. 15.

LOSHIN, D. (2013). *Integrating Structured and Unstructured Data*. TDWI Checklist Report. [interaktyvus] [žiūrėta 2014 m. rugsėjo 15 d.]. Prieiga per

interneta: <<http://h20195.www2.hp.com/V2/GetPDF.aspx%2Fc03670334.pdf>>.

OGC (2009). *Managing Successful Projects with PRINCE2*. London: The Office of Government Commerce (OGC).

PFERD, J. (2010). Integrating Structured and Unstructured Data. In *14th PNEC Conference*, Houston, TX, USA.

PRITCHARD, S. (2012). *How to manage unstructured data for business benefit* [interaktyvus] [žiūrėta 2014 m. rugsėjo 15 d.]. Prieiga per internetą: <<http://www.computerweekly.com/feature/How-to-manage-unstructured-data-for-business-benefit>>.

A Single View: Integrating Structured and Unstructured Data/Information within the Enterprise. *UNITAS* [interaktyvus], January 2002 [žiūrėta 2014 m. rugsėjo 15 d.]. Prieiga per internetą: <<http://lsdis.cs.uga.edu/GlobalInfoSys/Structured-and-Unstructured-for-EIPs.pdf>>.

SKYRIUS, R. (2008) The current state of decision support in Lithuanian business. *Information Research* [interaktyvus], 13(2), paper 345 [žiūrėta 2014 m. lapkričio 25 d.]. Prieiga per internetą: <<http://InformationR.net/ir/13-2/paper345.html>>.

THOO, E.; FRIEDMAN, T., BEYER, M. A. (2014). *Magic Quadrant for Data Integration Tools* [interaktyvus]. Gartner Research [žiūrėta 2014 m. rugsėjo 29 d.]. Prieiga per internetą: <<http://www.citicia.co.uk/content/files/magic-quadrant-for-data>

[integration-tools-gartner-july-2013_32583404.pdf](http://www.citicia.co.uk/content/files/magic-quadrant-for-data-integration-tools-gartner-july-2013_32583404.pdf)>.

TRAINOR, T. E.; BRAZIL, D. M.; LINDBERG, T. (2008). Building Knowledge From Organizational Experience: Approaches and Lessons Learned From US Army Base Camp Workshops. *Engineering Management Journal*, no. 20(2), p. 37-45.

TSIERKEZOS, S. (2010) *Comparing Data Integration Algorithms*. Thesis abstract. University of Manchester [interaktyvus] [žiūrėta 2014 m. rugsėjo 15 d.]. Prieiga per internetą: <http://studentnet.cs.manchester.ac.uk/resources/library/thesis_abstracts/BkgdReportsMSc10/Tsierkezos-Sebastian.pdf>.

WEBER, R.; AHA, D.; BECERRA-FERNANDEZ, I. (2000). Categorizing intelligent lessons learned systems. *Intelligent Lessons Learned Systems: Papers from the AAAI 2000 Workshop*, Technical Report WS-00-03, p. 63–67.

WEBER, R.; AHA, D.; BECERRA-FERNANDEZ, I. (2001). Intelligent Lessons Learned Systems. *Expert Systems with Applications*, no. 17, p. 17–34.

ZIEGLER, P.; DIETRICH, K.R. (2007) Data Integration – Problems, Approaches, and Perspectives [interaktyvus]. In: *Conceptual Modelling in Information Systems Engineering*, ed. by Krogstie J., Opdahl A. L., and Brinkkemper S., p. 39–58 [žiūrėta 2014 m. rugsėjo 15 d.]. Prieiga per internetą: <<http://www.uniriotec.br/~tanaka/SAIN/ZieglerSolvberg2007.pdf>>.

INFORMATION INTEGRATION: NEEDS AND CHALLENGES

Rimvydas Skyrius, Saulius Šimkonis, Indrius Sirtautas

S u m m a r y

Information integration is one of the significant challenges for today's business informatics to deal with. The development of a unified meaning out of multifaceted information from assorted sources and formats is complicated by a number of technological, system and organizational challenges. The goal of this paper is to analyze the problems in the area of business information integration and propose practical recommendations for solving them. There are numerous approaches and technology platforms for information integration, unified by the goal to manage and unite the meaning contained in information, and it becomes evident that effective information integration has to be performed by a joint effort of

users and technology. Different approaches are required for integration of structured and unstructured information – while the former requires the well-known tools and techniques of database domain, the latter requires much more flexible techniques, based on text analysis, text mining, fuzzy classification and intense participation of the users. It also has to be noted that forced integration of unstructured information using structured techniques creates a risk of stripping down the significant part of contained meanings. While dealing with information integration problems, a possible help would be the use of related experiences from the field of experience management or lessons learned.