

Управление на качеството на бизнес данните в информационна система

доц. д-р Валентин Кисимов
УНСС, катедра „Информационни
технологии и комуникации”

Резюме: Многообразието на приложения, които се изпълняват в една информационна система, води до различно ниво на непълно качество на данните в нея – от частично липсващи данни през дублирани данни с различно съдържание до противоречиви данни на бизнес услугата. Съвременните интегрирани системи на бизнес партньори, използващи Бизнес-с-Бизнес информационни потоци, води също до несъответстващи информационни съдържания. В зависимост от нивото на непълно качество на данните загубите за бизнеса на фирмата могат да варират, но те винаги присъстват.

В статия е предложена нов метод за създаване и поддържане качеството на данните според формулирани бизнес правила, както и са предложени решения на базата на информационни и комуникационни технологии (ИКТ) за подпомагане на функционирането на този метод. Предложеното ИКТ решение включва автоматизирани стъпки за следене качеството на данните, както и ИКТ решения, подпомагащи процесите на управление на качеството на данните, за които се изисква човешка намеса от страна на бизнес специалистите. Разработена е класификация на бизнес данните според принципа на съхранение и според принципа на логическо значение, като за всяка класификация са предложени различни решения.

В статията са изготвени показатели за оценка и управление на качеството на данните, като за всеки показател е разработено техническо решение за изпълнение. Цялостното ИКТ решение, обезпечаващо предложения метод, е описано като концептуална архитектура, помощни обслужващи системи и специално разработен формироваател на бизнес съобщения, служещ за управление качеството на транзитно преминаващите бизнес данни. За всяка от задачите на предложения метод са представени технически решения, което прави предложения метод директно използваем за реални практически нужди. Посочени са отделните хардуерни и програмни компоненти, чрез които е извършено тестване на определени компоненти от ИКТ решението на предложения метод.

Ключови думи: качество на данни (КД), класификация на бизнес данни, метод за динамично управление на КД, показатели за оценка на КД, ОУС за КД, ИКТ система за КД, формироваател на бизнес съобщения.

JEL: C6, C63, C8, C81, D8, D85.

1. Въведение и цели

Според докладу на Teradata (отдел на NCR) и на Intelligent Solutions Inc, USA, [10][12] проблемът за лошото качество на данните струва на бизнеса на САЩ повече от 600 милиарда долара на година. Явно за която и да е компания в света – малка или голяма, проблемът за качеството на данни-

те е много сериозен. Проведеният анализ на източниците на проблеми в качеството на данните показва, че те са свързани с процесите по въвеждане на данни, с процесите по интегриране на данни, пристигащи от различни разпределени подсистеми на същата информационна система, или от системите на бизнес партньори и съдружници [5]. Източник на проблеми в качеството на данните е и обединението и сливането на няколко компании в една [4]. На всички тези условия е свидетел и българският пазар, поради което качеството на данните не е само проблем на чуждите фирми, но и на българските.

По принцип фирмите се нуждаят да имат информационни системи, в които стратегическите данни са акуратни и бизнесът може да разчита на тях. За да се управлява качеството на данните, трябва да се изясни какво е то „качество на бизнес данните“ и как влияе на съответния бизнес, защото едно ниво на качество на данните при една корпорация може да не е достатъчно ниво на качество в друга корпорация. Решаването на проблемите за качеството на бизнес данните изисква многомерен подход и е много повече от чисто техническо решение [3]. Освен това решаването на този проблем не е еднократен акт, той има еднократни стъпки, но изисква непрекъснато да се поддържа съответното ниво на качество на бизнес данните, свързано с всяка постъпваща в информационната система данна.

Известни са няколко метода и модела за оценка на качеството на техническите данни [14][15], които са в две направления – ръчен анализ върху съществуващи и съхранени данни и автоматична оценка на данните след тяхното съхранение. Тези методи и модели не са свързани с бизнес значението на данните и могат да се прилагат само на ниво технически данни. Вторият проблем на тези методи и модели е, че те се отнасят за статични данни – вече съхранени върху

носител, а не за данни по време на тяхната транзакционност (обработка и придвижване към мястото за съхранение).

Целта на настоящата статия е да предложи нов метод за създаване и поддържане качеството на бизнес данните според формулирани бизнес правила, които е приложим за статични (вече съхранени) и динамични (в процес на транзакция) данни. Статията също има за цел да предложи решение на базата на информационни и комуникационни технологии (ИКТ) за подпомагане на функционирането на този метод. Предложеното ИКТ решение включва автоматизирани стъпки за следене на качеството на данните, както и ИКТ решения, подпомагащи стъпките в управлението на качеството на данните, за които се изисква човешка намеса от страна на бизнес специалисти (решенията на бизнес специалистите не може да се генерират автоматично от ИКТ системи; ИКТ системите може само да подпомагат бизнес ръководителите при вземането на техните решения). Предложеният метод и ИКТ решение поддържат качеството на данните, управлявано от динамично изменящи се бизнес правила, осъществявайки по този начин качеството на данните, съответстващо на динамичните бизнес нужди на реалния живот.

2. Качеството на данните е бизнес проблем, не е технически

Понятието „доверие в данните“ е понятие, което се променя, и зависи от конкретните в момента бизнес правила. Информационната система трябва да има „верни данни“ в динамичен аспект, в зависимост от динамиката на бизнес правилата и изискванията [1].

Често едни данни се създават от едно приложение, а други приложения ползват тези

данни [7]. Също така е често явление две приложения, създадени от независими фирми, да водят до записване на едни и същи данни няколко пъти, но в различна степен на пълнота и акуратност, създавайки различни степени на „бизнес истината“. Съвременните Бизнес-с-Бизнес интеграции водят до автоматично приемане на данни от партньор в бизнеса, които данни невинаги отговарят напълно на бизнес целите [6]. Проблемът за качеството на данните е проблем на цялостното управление на информационната система и на цялостното предназначение на информационната система [8][9][13].

От техническа гледна точка качеството на данните се изразява основно в *приемане, съхранение, защита, предоставяне* и *извличане* на данните, като от техническа гледна точка стойността на данните не е обект на внимание. Стойността на отделните данни, семантиката на тяхната обвързаност, отношенията и взаимодействието между отделните данни, пълнотата на всички необходими данни за един процес е чисто бизнес въпрос.

От бизнес гледна точка качеството на данните се изразява основно в *пълнота* (доколко всичките данни, нужни на едни бизнес процес, са налични), *съгласуваност* (доколко взаимно допълващите се данни за даден бизнес процес са в непротиворечиви стойности една с друга), *съответствие* (доколко данните за един бизнес процес са в рамките на допустими граници), *акуратност* (доколко точността на данните съответства с нуждите на даден бизнес процес), *навременност* (доколко времето, през което данните са създадени, е адекватно на времето, когато данните трябва да се извлекат за съответен бизнес процес).

Също така бизнес управлението на фирмата трябва да е наясно колко струва определен

процент некачествени данни и точно за кои данни. Ето защо качеството на данните е бизнес въпрос, като целта на ИКТ решението е да подпомогне този процес.

3. Бизнес данните – основа за управление качеството на данните

Бизнес процесите се базират на бизнес данни. Бизнес данните трябва да бъдат добре идентифицирани, със стандартизирани формати на полета, да имат съответствие с бизнес процесите, в които се използват, да са без дублиране и да имат съответната бизнес референтна интеграция, чрез което да се осигурява тяхното качество.

За да се извърши фокусиране върху качеството на бизнес данните, ще направим детайлна класификация на данните и от това ще следва как да се осигурява тяхното качество.

В една информационна система данните може да се класифицират по различни принципи:

- според принципа на запазване;
- според принципа на логическо значение.

Според принципа на запазване данните могат да се определят като:

- а) съхранени;
- б) производни;
- в) пристигащи.

Съхранените данни се записват и пазят във файлова система, в база от данни или в хранилище за неструктурни данни; записват се чрез програмни средства по време на изпълнение на програми. *Производните данни* се създават динамично само когато е необходимо тяхното използване. Те се получават

В резултат на обработка на съхранените данни и тяхното проявление е в момента на използването им. *Престигащите данни* са в комуникационните мрежи и са данните по пътя към съответните системи за управление на данни. Това са данни, които ще се превърнат в съхранени на носител от данни, или са заявки за създаване на производни данни. При наличност на добра система за управление качеството на данните *престигащите данни* трябва да се проверяват за условията на качество и след проверката да се предават за записване (при коректно спазване условията за качество) или да се изхвърлят (при некоректност от тяхна страна в условията за качество).

Според принципа на логическо значение на данните те могат да се определят като:

- a) технически елементарни данни;
- b) бизнес елементарни данни;
- c) технически агрегирани данни;
- d) бизнес междинни данни:
 - бизнес агрегирани данни,
 - метаданни,
- e) бизнес крайни данни.

Техническите елементарни данни са най-ниското ниво на агрегация на данни, без да има в тях бизнес значение, като това са обикновено поле от запис на файл, съдържание на колона на таблица в БД, променлива в обект от програма и те не носят бизнес значение.

Бизнес елементарните данни също представляват данни от най-ниското ниво на агрегация, но като самостоятелни данни те имат бизнес значение, например име на фирма, цена на артикул, пълен адрес на клиент.

Технически агрегираните данни представляват съвкупност от елементарни данни, кои-

то нямат бизнес значение, и агрегацията на тези данни има само техническо значение, например таблицата SYSTEM.SYSCOLSTATS от базата от данни IBM DB/2, използваща индекси в DB/2, е агрегирана информация, състояща се от много колони, представляващи бизнес и технически елементарни данни (като цена на артикул, номер на клиент, име на фирма).

Бизнес агрегираните данни също представляват съвкупност от елементарни данни, но които като съвкупност имат бизнес значение. Например таблицата „Доставчик“ в БД има елементарни технически данни (вътрешен номер на доставчик) и елементарни бизнес данни (име на доставчик, адрес на доставчик, e-mail за контакт на доставчика), но като цяло тя е една бизнес съвкупност. Бизнес агрегираните данни може да се оформят в йерархия.

Метаданните са данни за данните. Това са данни, които описват съвкупност от елементарни данни, имат някакъв бизнес смисъл на абстракция и служат за управление на тези елементарни данни. Метаданните се различават от агрегираните данни по това, че в тях може да се съдържа допълнителна информация, която я няма в съставлящите я елементарни данни, като описатели, указатели, управляеми речници и друга управляваща информация. В допълнителното описание към метаданните може да се включи и уъркфлоу (последователно/паралелни потоци от процеси) към следващи метаданни. От метаданни може да се създаде по-висока йерархия метаданни.

Бизнес агрегираните данни и метаданните се обединяват в понятието *бизнес междинни данни*, т.е. това са данни, които не са бизнес елементарни данни, но същевременно могат да се използват за създаване на по-високо ниво бизнес данни – или на други бизнес меж-

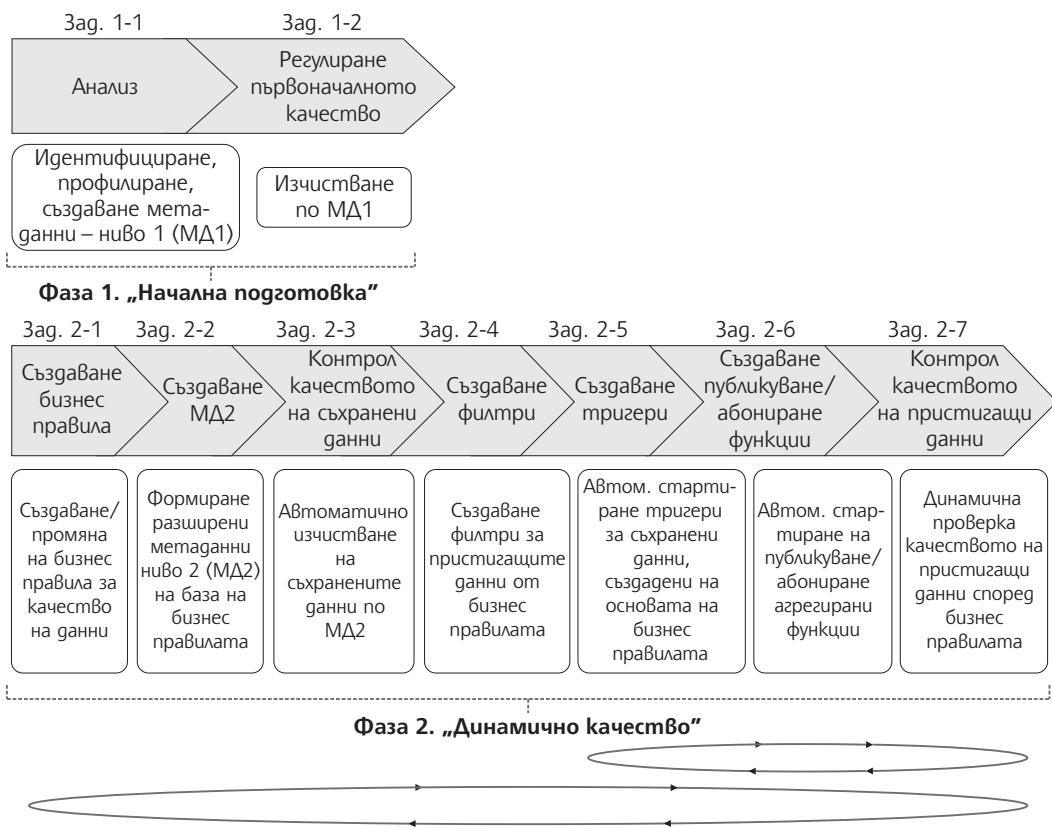
динни данни, или на бизнес крайни данни. *Бизнес крайните данни* са бизнес данни, които се използват за формиране на краен бизнес резултат, като бизнес отчети, бизнес графики, бизнес тенденции и т.н. Обикновено бизнес крайните данни не се съдържат на носител.

Техническите данни са само обслужващи данни на бизнес процесите. Когато става въпрос за качеството на данните, трябва да се разбира качеството на бизнес данните. Изхождайки от факта, че бизнес крайните данни са само резултатни данни, за управление качеството на данните ще използваме три типа бизнес данни – бизнес елементарни данни, бизнес агрегирани данни и метаданни.

4. Метод за динамично бизнес управление качеството на данните

От автора е разработен метод за динамично бизнес управление на качеството на данните, който се състои от 2 фази и е представен на фигура 1.

Фаза 1 е начална фаза, защото се изпълнява преди същинското използване на метода. Тази фаза има за цел да извърши първоначално изчистване на данните на информационната система. По принцип всяка съществуваща информационна система има съвкупност



Фигура 1

от данни, които са в някаква степен на ненапълно добро качество. Посредством тази фаза се прави начално изчистване на качеството на данните в информационната система. В тази фаза се изпълняват 2 задачи. Първата задача (заг. 1-1) е идентифициране, профилиране и създаване на метаданни-1 (МД1), която има за цел да анализира съществуващата информационна архитектура на информационната система, да създаде съответните профайлове на данните, от които да се изградят метаданни. Наричаме тези метаданни начални (МД1), защото те се изграждат, преди да започне същинското използване на метода. Втората задача на първата фаза (заг. 1-2) извършва началното изчистване на данните в системата, като метаданните (МД1) имат водеща роля в изчистването.

Втората фаза – „Динамично качество“, е същинската част от предложения метод. Тази фаза се изпълнява всеки път, когато се извършат промени в бизнес правилата за управление качеството на данните, или се създават нови правила. Затова цялата фаза може да се нарече циклична. Същевременно една част от фазата (задачи 2-5, 2-6 и 2-7) работи постоянно и автоматично и следи непрекъснато за качеството на пристигащите данни. Тази част от втора фаза е вътрешно циклична на цялата фаза.

Задача 2-1 е същинското формиране на бизнес правилата за управление качеството на данните. Тази задача също е свързана с промяна на съществуващи правила, когато се прилага при цикличното ѝ използване. Бизнес правилата за качество на данните се описват на бизнес езика BPEL (Business Process Execution Language), на който работят бизнес специалистите и на който стандартно се описват всички бизнес процеси. Това не е технически език за програмиране, а бизнес език за описване на бизнес

активности. Използвайки BPEL за описание правилата за качество на данните, се постига бизнес контролиране и управление качеството на данните. Предложената ИКТ система, която ще подпомага управлението на качеството на данните, използва интерпретатор на BPEL езика, така че с изменение на бизнес правилата за качество на данните (чрез BPEL кодиране), управлението на данните ще се променя синхронно с новите правила.

Задача 2-2 има за цел да формира нова съвкупност от разширени метаданни – метаданни-2 (МД2), получени на база на МД1 и от бизнес правилата на управление качеството на данните (загадени в задача 2-1). МД2 е основа за работата на фаза 2 на метода. На основа на данните от МД2 се извършва еднократно изчистване на информацията в информационната система (задача 2-3). На нейна основа се създават филтри за селектиране на пристигащи данни (задача 2-4). Целта на филтрите е да изберат от множеството данни, постъпващи към хранилищата на данни, онези приложни и бизнес съобщения, в които пристигат данни, подлежащи за проверка от бизнес правилата за управление на качеството на данните. В предложената ИКТ система за поддържане на метода, на база на IP пакетите, в мрежата се прави формиране на приложни съобщения (например HTTP, SAP, FTP, MQSeries и т.н.), като от едно или няколко приложни съобщения се създават бизнес съобщения. В предложения метод от пристигащите бизнес съобщения се извличат съответните бизнес данни, които се анализират за съответно качество (според BPEL правилата). Тези бизнес данни са пристигащи данни и са един от видовете данни, подлежащи на контрол на качество от предложения метод. Формираните в тази задача филтри имат за задача да извлекат само онези бизнес съ-

общения, в които има бизнес данни за проверка. Филтрите също така селектират бизнес данните от бизнес съобщенията и се създават на база на XML описания.

Задача 2-5 служи да формира тригери в съществуващите бази от данни, за да се извлече информация от записани или за записващи се в БД бизнес данни. Тригерите се активизират в зависимост от настъпване на определени промени, свързани с тези данни, като логиката за тригерите съответства на бизнес правилата от XQuery описанието. Създаването на тригерите е програмиране, което обикновено се извършва на Java или на някой специфичен за конкретната БД език. Указание, за кои данни да се създадат тригери, следва от XQuery описанието. Възможно е да се създаде еднократно тригер за една данна, а след това той да се използва за управление на различни нива на качество на тази данна. Тригерите може да информират директно централната система за управление качеството на пристигащите данни – „Динамична проверка качеството на пристигащи данни“, но може и да предават резултати на процедура за публикуване/абониране. Тази процедура (задача 2-6) има за цел да агрегира информация от различни тригери, постъпващи асинхронно, и има „абонамент“ за информация, която трябва да се събере от няколко тригера. Този „абонамент“ представлява съвкупност от правила, произтичащи от XQuery бизнес правилата за качество на данните. Същевременно отделните тригери, които подават информация към тази процедура, „публикуват“ своите резултати там. Когато информацията от различните тригери се агрегира според бизнес правилата, тази агрегирана информация се предава на блока „Динамична проверка качеството на пристигащи данни“. Последната задача – задача 2-7, извършва същинския контрол на пристигащите данни.

5. Показатели за оценка и управление качеството на бизнес данните

В настоящата статия се предлага съвкупност от показатели за оценяване качеството на данните, които са водещи в работата на предложения метод за управление качеството на данните, както и във функционирането на ИКТ системата, която поддържа самото управление на качеството на данните. Тези показатели са следните.

- *Пълнота на данни (Completeness)*

Това представлява параметър, измерващ съществуването или отсъствието на данни. Например при банките адресът на клиентите се взема от други отчети – от сметката за плащане на вода, на телефон, на ганък сгради и пр., и тези данни невинаги са пълни. Пълнотата на такива данни е средно 95 %, но 5 % непълнота могат да доведат до милиони загуби на банката. За да се осъществи пълнота на данните, те трябва да се изчистят. Изчистването може да доведе до 3 възможности: коригиране на непълните данни с използване на така наречени „верни данни“; запазване на данните в непълен вид, но се създава отделен запис, отговарящ на съгласуваните – изчистени данни; отстраняване на непълните данни.

- *Съгласуваност на данни (Consistency)*

Съгласувани данни са тези, при които при възможно наличие на дублиране на данни, те са с еднакво и налично съдържание. Съгласуваността може да бъде транзакционна съгласуваност – за времето на изпълнение на една транзакция, или съхранена съгласуваност (съгласуваност на данни, записани в някакви системи за управление на данни – файлове, БД, складове от данни и т.н.).

- *Наличност на данни (Availability)*

Това изисква данните да са налични (например в документ за нареждане полето „Заявена бройка“ не може да е празно). Основно в този показател за качество е наличността на данните, идващи от външна корпорация, да отговарят на бизнес връзките с тази корпорация.

- *Съответност на данни (Relevance)*

Този показател изисква стойностите на данните да попадат в приемлив обхват или да са от определена типизирана съвкупност.

- *Акуратност/прецизност на данни (Accuracy/Precision)*

Акуратността се отнася до това, колко близко стойността на данните е до „верните“ данни. Прецизността е измерението на тази близост. Така наречените „верни“ данни могат да се задават само от позиция на коректно бизнес разбиране за същността на данните.

- *Навременност/свежест на данни (Timeliness/Freshness)*

Данните от борсата, отпечатани и публикувани, са винаги стари, докато данните за исторически събития са винаги свежи, навремени. В бизнес информационните системи данните обикновено са стари, но за някои бизнес процеси тази „старост на данните“ е невалидна. Този параметър използва времето за записване на данните и времето, когато данните се смятат актуални. Разликата между тези времена би показала дали данните са свежи, или са стари.

6. ИТ система за бизнес управление качеството на данните

6.1. Концептуална архитектура на ИКТ решението

Концептуалната архитектура на предложеното ИКТ решение, което да поддържа предложения метод, е представено на фигура 2.

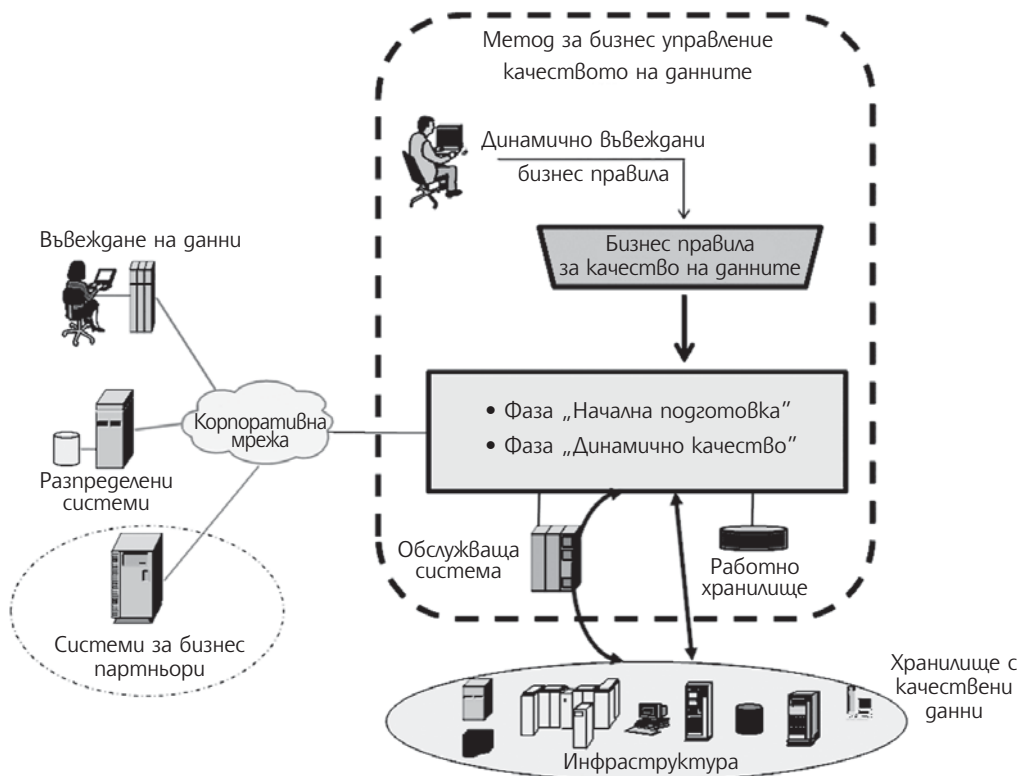
В инфраструктурата се намират всички хранилища на съхранени данни. Данните в тези хранилища и пристигащите данни трябва да се контролират за качество чрез бизнес правила. Пристигащите данни идват от корпоративната мрежа, което означава: от системите за въвеждане на данни; от другите системи на разпределената архитектура на информационната система; от други партньорски ИКТ системи, работещи на принципа „Бизнес-с-Бизнес“. Предложената ИКТ се състои от: блок, наречен „Бизнес правила за качество на данните“, който съхранява бизнес правилата, описани първоначално на BPEL; блок, който управлява двете фази на метода – „Начална подготовка“ и „Динамично качество“; обслужваща система, която извършва основни процеси, насочени към управление качеството на данните; и работно хранилище, служещо за междинно запазване на работни данни. Предложената ИКТ система подпомага бизнес специалиста в управление на процесите, свързани с качество на данните, както и извършва автоматични процеси за динамично управление качеството на данните. В резултат от работата на предложената ИКТ система хранилищата с данни ще имат качество на данните, зададено от бизнес правила.

Бизнес правилата се задават чрез програмиране на BPEL от бизнес специалисти и това е изходната точка на бизнес управле-

нието на данни (задача 2-1). Задачите 1-1 и 2-1 се извършват с помощта на обслужващата система под управлението на бизнес специалист по данни. Задачи 1-2 и 2-3 се извършват автоматично от обслужващата система. Задачи 2-4, 2-5, 2-6 и 2-7 изискват програмиране, съответстващо на зададените бизнес правила. Това са най-продължителните процеси, които, в зависимост от сложността на бизнес правилата, може да отнемат първоначално няколко човеко-месеца. Впоследствие, при промяна на бизнес правилата, прекодирането на тези четири задачи става в значително по-кратки срокове – няколко човеко-седмици. Обикновено критичните бизнес данни не се променят с промяна на бизнес правилата за качество на данните, променят се само условията на

проверка на тяхното качество. Това означава, че при промяна на BPEL бизнес правилата, в интервал на една или няколко седмици ще заработи ИКТ системата за контролиране това качество в автоматичен режим.

Използваният език за описание на бизнес критериите за управление качеството на данните – BPEL, е стандартен език за моделиране на бизнес процеси, който може да използва услугите на ориентирана към услуги архитектура (ОУА) [2]. Езикът BPEL е език за оркестриране (взаимодействие между различните услуги, имайки централизирано управление), а не за хореографиране (взаимодействие на услуги по двойки с определено взаимодействие между двойките и с продължителни и твърдо определени



Фигура 2

състояния в двойките). Езикът BPEL дава възможност да се изградят агрегати от услуги, разглеждани в езика също като услуги. Езикът BPEL е стандартизиран и се поддържа в момента от основните софтуерни производители в света и от създателите на отворен код.

Предложеното ИКТ решение използва една BPEL услуга за един тип проверка качество на данни, базирано на един тип показател

за оценка на качеството. Използват се 6-те показателя за оценка на качеството, описани по-горе, като в предложената системата са създадени шест BPEL услуги. Например дефиниран е показател „Пълнота на данните“ и е създадена една BPEL услуга за пълнота на данните. Това, кои данни ще се проверяват за пълнота и с какви форми на пълнота, се задава чрез параметри на BPEL услугата. За всяка една от 6-те BPEL услуги могат да се задават неограничено

Таблица 1

Но BPEL услуга	Показател за оценка качеството на данните	Описание на BPEL услугата	Параметри в BPEL услугата
1	Пълнота на данни	Изчистване по „верни данни“ със зададен начин за привършване	<ul style="list-style-type: none"> „Верни данни“ (константа или променлива) Привършване (коригиране, отстраняване, запазване със страничен запис)
2	Съгласуваност на данни	Сравнява съдържанието на дублирани данни	<ul style="list-style-type: none"> Тип на съгласуване (съхранени – съхранени данни; съхранени – пристигащи данни) Съхранени данни (тип-данни-В-МД2) Пристигащи данни (партньор, тип-МД2, тип-данни-В-МД2)
3	Наличност на данни	Сравнение за наличност на данни – съхранени или пристигащи данни	<ul style="list-style-type: none"> Данни за сравнение (съхранени/пристигащи) Условие за наличност (! = null; != 0; != space) Пристигащи данни (партньор, тип-МД2, тип-данни-В-МД2)
4	Съответствие на данни	Сравнение за попадане на данните в определен обсег	<ul style="list-style-type: none"> Тип на съответствие (съхранени данни; пристигащи данни) Съхранени данни (тип-данни-В-МД2) Пристигащи данни (партньор, тип-МД2, тип-данни-В-МД2)
5	Акуратност на данни	<p>Определя период от време (ПВ): $t_2 - t_1$; $t_{\text{текущо}} - t_n$.</p> <ul style="list-style-type: none"> За ПВ: текуща акуратност (ТекАк) = $\sqrt{((\text{ТекАк} - \text{ВярнаАк} / \text{ВярнаАк})^2)}$ $\text{ДопАк} - \text{ТекАк} < 0$ 	<ul style="list-style-type: none"> Вярна акуратност (ВярнаАк) Допустима акуратност (ДопАк) Период от време ($t_1 \div t_2$) Начален момент (t_n)
6	Навременност на данни	<ul style="list-style-type: none"> Записване времето на пристигане ($t_{\text{прист}}$) с името на данната Проверка $t_{\text{тек}} - t_{\text{прист}} < t_{\text{навр}}$ 	<ul style="list-style-type: none"> $t_{\text{навременност}}$ ($t_{\text{навр}}$) Пристигащи данни (партньор, тип-МД2, тип-данни-В-МД2)

количество изисквания за качество – множество параметри. Когато в езика BPEL се специфицира използване на една услуга за управление качеството на данните, броят параметри не е ограничен. В предложената ИКТ система за всяка BPEL услуга съществува буфер, където се зареждат параметрите на услугата, и по време на изпълнение параметрите се вземат от буфера за по-голяма оперативност. Реализирането на 6-те BPEL услуги е във формата на обикновени Java обекти – POJO (Plain Old Java Object). В предложената архитектура се използва компонентът ориентирана към услуги архитектура (OYA), която осигурява изпълнението на основните предложени фази на метода. Тази архитектура се базира на продукта ServiceMix като OYA, който поддържа стандарт JSR181 за EJB/POJO. По този начин всичките 6 POJO кодове се превръщат на услуги в OYA.

В таблица 1 са представени 6-те BPEL услуги за управление качеството на данните, заедно със спецификацията на възможните използвани параметри в тези BPEL услуги.

6.2. IBM Information Server като обслужваща система на ИКТ системата за бизнес управление качеството на данните

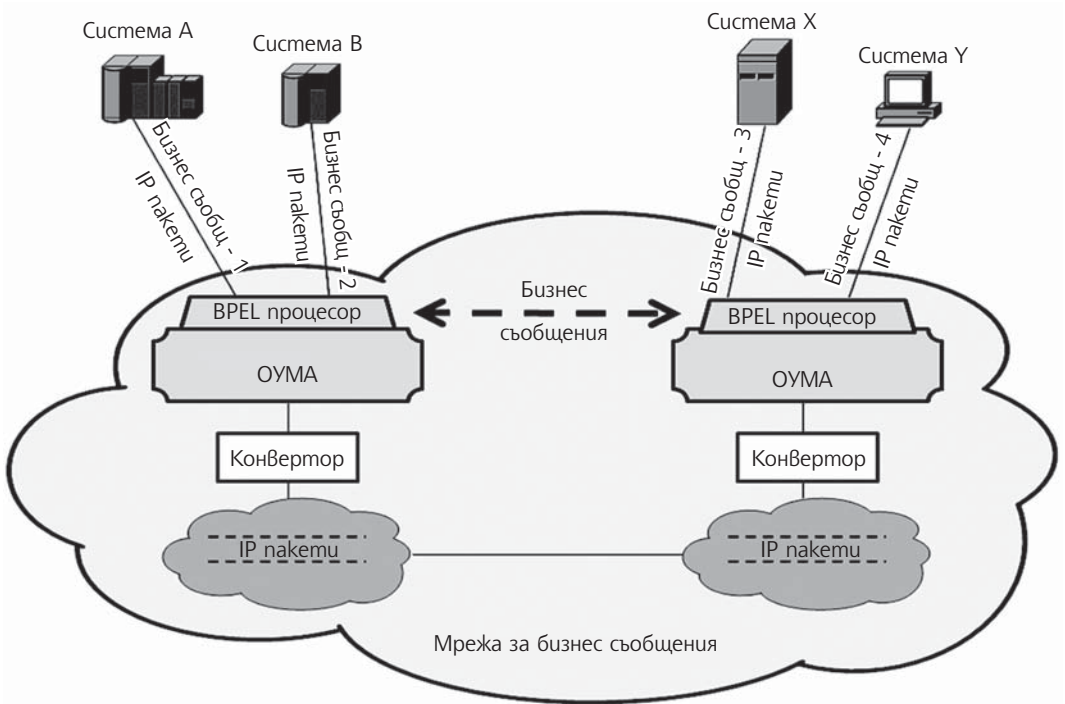
Предложената ИКТ система използва IBM Information Server [11] като обслужващ сървър. IBM Information Server извършва профилиране, изчистване и интегрира данните от различни хетерогенни системи. Той извършва интегриране, водено от метаданни. IBM Information Server предоставя множество услуги под формата на уеб услуги, които групирани представляват: подобряване на качеството на данните – използвайки главно изчистването на данни; опростяване на преобразуването на данни – извършвано чрез ETL процеси, главно използвани за

хранилища за данни; осигуряване на единна представа за данни, ползвайки структурни и неструктурни данни.

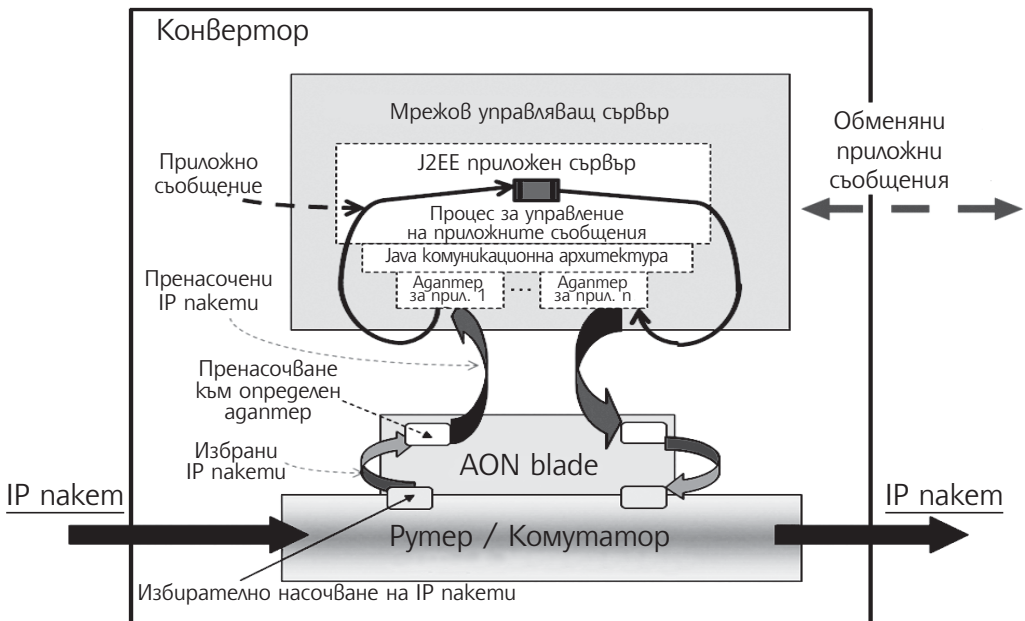
Компонентите на IBM Information Server са: WebSphere Information Analyzer за определяне на релации между данни в различни физически среди; WebSphere Business Glossary за съхраняване описанието на метаданните; WebSphere Quality Stage за изчистване на данните, базирано на правила; WebSphere Data Stage, извършващ основно ETL процесите; WebSphere Federation Server, осигуряващ комуникации с различни системи за съхранение на структурирана и неструктурирана информация, независимо от формата на данните и вида на тяхното приложно управление; WebSphere Information Server Directory, служещ за публикуване на съществуващите за използване уеб услуги; WebSphere Metadata Server, представляващ хранилище за метаданните, като същевременно осигурява и услуги за метаданни – метаданни за достъп, метаданни за интегриране, метаданни за въвеждане и извеждане, анализ на метаданни, търсене и справки за метаданни.

6.3. Формировател на бизнес съобщения

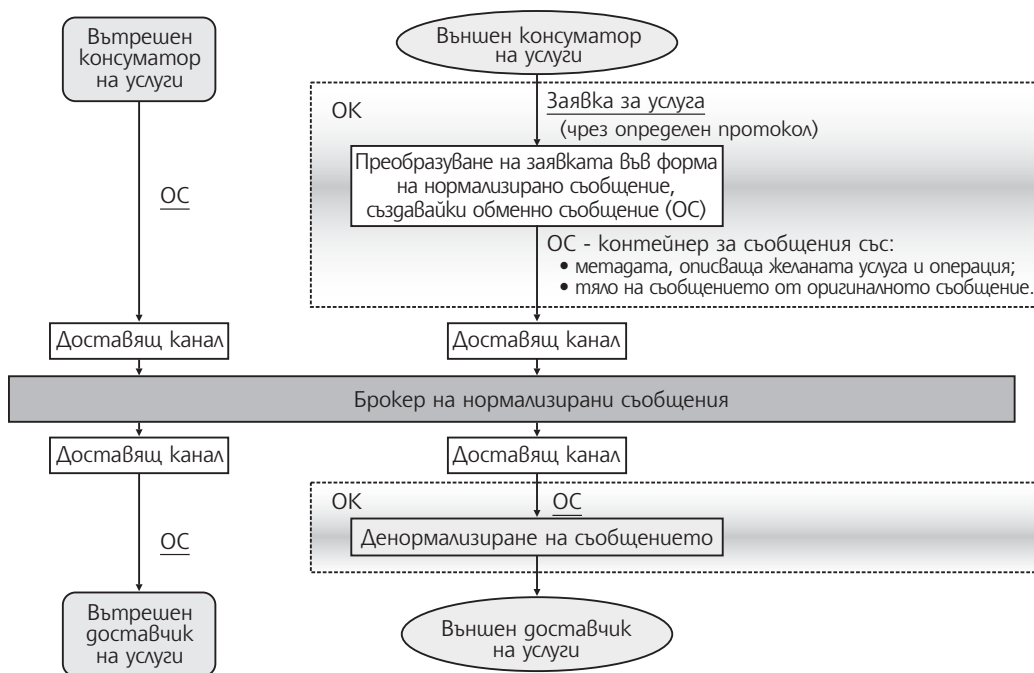
За обработване на пристигащи съобщения предложената ИКТ система използва формировател на бизнес съобщения. Това е съвкупност от устройства и процедури, водещи до създаване и управление на бизнес съобщения. Мрежовите IP пакети се преобразуват в приложни съобщения (използвайки устройството конвертор), които след това се преобразуват в бизнес съобщения (използвайки устройство, базирано на ориентираната към услуги архитектура – OYA). Поставайки OYA в инфраструктурната архитектура, т.е. на междинно ниво в 7-те нива на ISO модела, тази OYA ще



Фигура 3



Фигура 4



Фигура 5

наричаме ориентирана към услуги междинна архитектура (ОУМА) – фигура 3. В проектираното решение за ОУМА се използва отново ServiceMix продуктът (продукт с отворен код), в който BPEL процесорът е вградена компонента.

Устройството конвертор е съвкупност от 3 различни устройства: Cisco комутатор или рутер, към който се включва Cisco AON blade, и мрежов управляващ сървър. AON blade е специална карта, в която е зареден AON софтуер за извличане на IP пакети и преобразуване според определени изисквания. AON се програмира като J2EE приложен сървър и в настоящата разработка приложното осигуряване на AON работи като прокси сървър за пренасочване на пакетите към мрежовия управляващ сървър – фигура 4.

ОУМА използва корпоративна шина за услуги (КШУ), работеща на принципа на JBI (Java Business Integration) – стандарт JSR208. Начинът за преобразуване на съобщения в бизнес услуги е показан на фигура 5.

Всеки външен консуматор, такива като бизнес съобщения, се превръщат в обвързани компоненти (ОК) посредством превръщането на съобщението в обменно съобщение (ОС) – чрез добавяне на мета-данни като заглавни данни към съобщението. Така полученото обменно съобщение са подава на доставящия канал, който го предава като нормализирано съобщение на брокера на нормализирани съобщения. Последният представлява рутер на нормализирани съобщения.

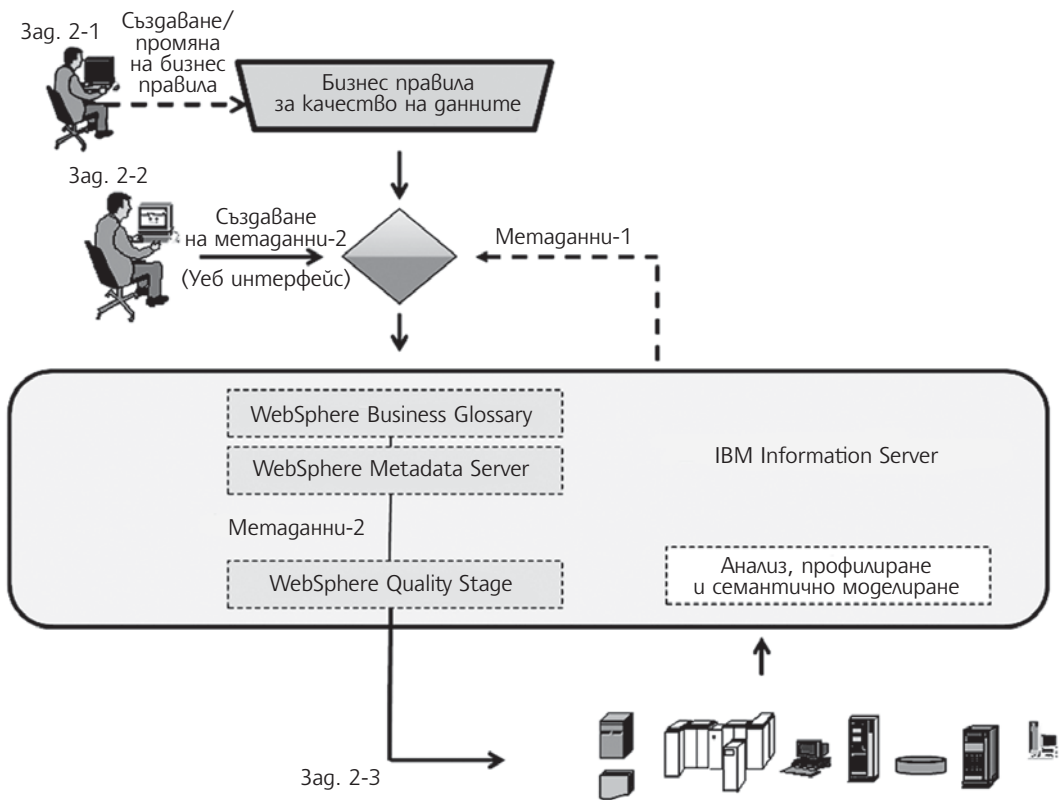
6.4. Технологично изпълнение на задачите на предложения метод

За изпълнение на задачите 2-1, 2-2 и 2-3 се изисква използване на обслужващия сървър – IBM Information Server, като тези задачи се осъществяват под управлението на бизнес специалист по качество на данните (фигура 6).

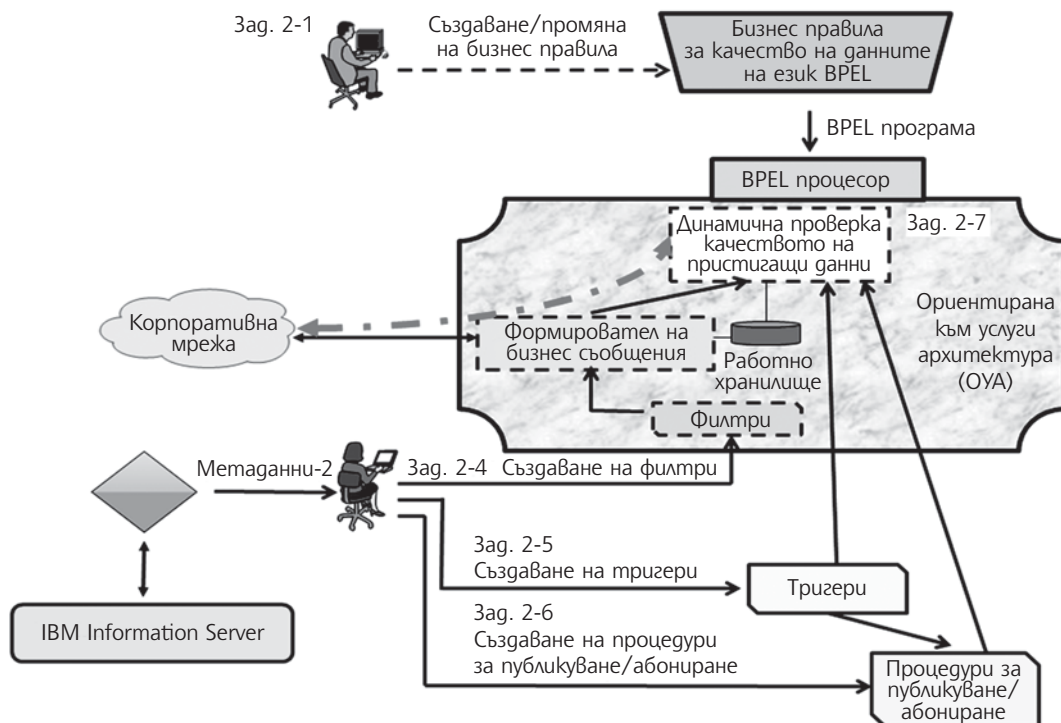
Първите две задачи изискват бизнес виждането на специалиста, затова се извършват под негово управление. След формиране на МД2 (края на задача 2-2), изпълнението до края на задача 2-3 (изчистването на данните според бизнес правилата) се извършва автоматично. Намесата на бизнес специалиста се свежда до стартиране на съответните

задачи и до подаване на корелации между бизнес данни, които IBM Information Server не може да улови автоматично чрез аналитичните си и профилиращи механизми. Изпълнението на останалите задачи на фаза 2 на метода е показано на фигура 7.

Задачите 2-4, 2-5 и 2-6 изискват съответно програмиране, което се прилага при изменение на бизнес правилата в BPEL. Задачата 2-7 изисква първоначално програмиране, като при изменения на BPEL бизнес правилата за качество на данните допълнително програмиране не се изисква. Програмирането на задача 2-7 се счита като част от първоначалното създаване на системата.



Фигура 6



Фигура 7

Литература

1. Кисимов, В., Р. Николов, Бизнес интелегентни инфраструктури – основа за бизнес интелегентни системи в реално време, Международна научна конференция по повод 40-годишнината на специалност „Информатика“, под научната редакция на проф. А. Бъчваров, Сборник доклади, С., 2007.
2. Велев, Д., Е. Денчев, К. Стефанова, В. Лазарова, М. Цанева, А. Мурджева, SOA – Основно направление в развитието на софтуерните технологии, Международна научна конференция по повод 40-годишнината на специалност „Информатика“, под научната редакция на проф. А. Бъчваров, Сборник доклади, С., 2007.
3. Стефанова, К., В. Кисимов, М. Цанева, В. Лазарова, А. Мурджева, Д. Велев, Е. Денчев, Д. Кабакчиева, Проектиране на Център за компетентност по бизнес интелегентност, Международна научна конференция по повод 40-годишнината на специалност „Информатика“, под научната редакция на проф. А. Бъчваров, Сборник доклади, С., 2007.
4. Денчев, Е., К. Стефанова, М. Цанева, Д. Велев, В. Кисимов, А. Мурджева, В. Лазарова, Проблеми и решения при избор на ERP система, Международна научна конференция по повод 40-годишнината на специалност „Информатика“, под научната редакция на проф. А. Бъчваров, Сборник доклади, С., 2007.
5. Цанева, М., В. Кисимов, А. Мурджева, Е. Денчев, Д. Велев, В. Лазарова, К. Стефанова, Интегриране на бизнес приложения – ос

новно предизвикателство към системното програмиране, Международна научна конференция по повод 40-годишнината на специалност „Информатика“, под научната редакция на проф. А. Бъчваров, Сборник доклади, С., 2007.

6. Лазарова, В., Е. Денчев, Д. Велев, А. Мурджева, В. Кисимов, К. Стефанова, М. Цанева, Критерии за определяне на потребителската ефективност на информационните системи в интернет, Международна научна конференция по повод 40-годишнината на специалност „Информатика“, под научната редакция на проф. А. Бъчваров, Сборник доклади, С., 2007.

7. Мурджева, А., К. Стефанова, М. Цанева, В. Лазарова, Е. Денчев, Д. Велев, Разпределяне на бизнес логиката в многослойни приложения и съвременните технологии, Международна научна конференция по повод 40-годишнината на специалност „Информатика“, под научната редакция на проф. А. Бъчваров, Сборник доклади, С., 2007.

8. Poor Quality Data: The sure way to lose business and attract auditors, Gartner conference, SA, 2006.

9. Data Integration, The Nucleus of the Data Delivery Evolution, Gartner Conference, SA, 2007.

10. Data Quality Management: Often Overlooked Key to Affordable, High Quality Patient Care, Darryl McDonald, Teradata, NCR.

11. IBM Information Server, IBM, http://www-304.ibm.com/jct03002c/software/data/integration/info_server/overview.html

12. US Public Company Accounting reform and corporate responsibility, http://www4.law.cornell.edu/uscode/15/usc_sup_01_15_10_98.html

13. Basel II: Revised international capital framework, <http://www.bis.org/publ/bcbsca.htm>

14. Batini, C., M. Scannapieca, DataQuality, SpringerLink, 2006.

15. Pilar, A., M. Lachlan, Quality Measurement and Assessment models including Data Provenance to grade Data sources, <http://www.macs.hw.ac.uk/~pilar/research/ATINERv3.pdf> 