

Calitatea datelor tip adresă

Baza eficienței operaționale

În ce constă calitatea datelor?

Pur și simplu, calitatea datelor implică asigurarea acurateții, actualității, integrității și a coerenței datelor folosite de către o organizație, în același timp având grijă ca toți cei ce folosesc datele să aibă aceeași înțelegere asupra ceea ce ele reprezintă. De exemplu, dacă datele de vânzări includ sau nu vânzările interne? Sunt acestea măsurate în unități sau Euro, RON sau poate dolari? Sfera de aplicabilitate a calității datelor nu este limitată la datele generate de propriile operații ale unei organizații; aceasta trebuie să vizeze și datele obținute din surse externe. Unele definiții ale calității datelor includ, de asemenea, accesibilitatea care, deși nu este o parte strictă a definiției calității datelor, este o caracteristică dezirabilă pentru ca acestea să fie utilizabile.

Având în vedere accentul care se pune în prezent pe necesitatea de a menține o “vedere de 360 grade asupra clientului”, s-ar putea în mod eronat crede că acest lucru implică crearea unei masive înregistrări pentru fiecare client. Mai degrabă, calitatea datelor reflectă posibilitatea de a lega toate înregistrările aceluiasi client - o sarcină care poate fi îndeplinită cu ajutorul identificatorilor înregistrărilor asociate cu fiecare client în parte.

Municipalitățile care doresc să colecteze echitabil veniturile din impozite trebuie să se asigure că acestea nu sunt incorect colectate de către municipalitățile vecine. Software-ul răspunzător de calitatea datelor oferă acuratețea necesară autorităților fiscale competente, ajutând companiile să rămână în limite legale și asigurând astfel tratarea corectă a clienților și a municipalităților în ceea ce privește taxele și impozitele.

Importanța calității datelor

Organizațiile iau decizii pe baza datelor disponibile la acel moment. Dacă o organizație își poate spori, în întregime, calitatea datelor, își va putea îmbunătăți calitatea deciziilor și, prin urmare, eficacitatea și eficiența. Aceasta va permite acelei organizații să-și servească mai bine și mai profitabil părțile constituente, fie aceștia clienți, angajați,

parteneri de afaceri sau acționari.

Multă atenție este focalizată pe aspectul curățării datelor ca parte a calității datelor. Aceasta este, însă doar o parte a ecuației. Alte aspecte includ integrarea datelor și aplicarea de reguli consecvente.

În timp ce calitatea datelor era considerată cu ceva timp în urmă o inițiativă laudabilă dar nu vitală, organizațiile realizează acum că aceasta este o necesitate absolută în special pentru acele aplicații de importanță critică sau pentru a îndeplini normele guvernamentale de raportare și comunicare. De fapt, când este vorba de aplicații tip Homeland Security sau Solvency II calitatea datelor devine rapid o preocupare serioasă pentru corporații.

Care este magnitudinea problemei?

Potrivit unui raport publicat de The Data Warehousing Institute (TDWI) în 2002, “calitatea slabă a datelor de intrare” a costat impresionanta sumă de 611 miliarde dolari numai pentru operațiile de printare, poștă și utilizarea excesivă a personalului în Statele Unite ale Americii. TDWI citează mai multe exemple, inclusiv o companie de telecomunicații ale cărei erori din datele de intrare au dus la codificarea incorectă a conturilor, pierzând astfel 8 milioane dolari pe lună atunci când acesta nu a putut trimite facturile. Costul real este, fără îndoială, mult mai mare, deoarece suma calculată de 611 miliarde dolari nu include efectele secundare, cum ar fi cele asociate cu alienarea și pierderea de clienți.

Știm, de asemenea, că problema persistă. Într-un recent studiu Gartner din 2009, participanții estimează că slaba calitate a datelor costă companiile în medie 8.2 milioane dolari pe an. 22% din respondenți și-au calculat pierderi anuale de peste 20 milioane dolari, iar 4% din respondenți estimau pierderi de peste 100 milioane dolari. În timp ce pierderile de milioane de dolari sunt semnificative, analiștii Gartner cred că aceste cifre subestimează adevărata pierdere financiară a organizațiilor.

O atenție deosebită se acordă rezolvării

problemelor legate de consolidarea înregistrărilor pentru a putea avea o imagine unică asupra clientului (SCV sau single customer view) respectiv eliminării duplicatelor de corespondența adresată unei persoane. Acestea sunt doar câteva exemple de adăugare de valoare datelor care, odată aplicate, aduc beneficii organizațiilor.

Calitatea datelor vizează nu numai numele și adresa clientului ci și numele produselor și descrierea acestora, unitățile de măsură, codurile de procedură medicală și numerele de identificare ale pacienților, numerele de telefon, adresele de e-mail, codurile de mărfuri, numerele de vânzător și numere de identificare ale vehiculelor, pentru a numi doar câteva. De exemplu, dacă o societate de asigurare vinde unele dintre produsele sale în RON, altele în EURO și câteva în dolari, atunci potențialul de eroare este considerabil.

Vă amintiti că Mars Climate Orbiter (1998) s-a prăbușit din cauza unei conversii incorecte a din metri în feet și inches!

Efecte adiționale

Slaba calitate a datelor poate influența negativ modul în care o companie este percepută în piață. Prima impresie trebuie să fie una de calitate. La fel cum impresia inițială a vizitării birourilor unei companii poate fi influențată de zona recepției, imaginea companiei poate fi influențată de calitatea datelor ei, în special dacă au ca rezultat adrese de e-mail eronate, numere de facturi incorecte și erori în livrare.

Incapacitatea de a elimina numele și adresele redundante duce la costuri inutile cu **poșta**.

Destinatarii pentru care există adrese duplicate pot deveni frustrați și pot pune la îndoială operativitatea companiei în cazul în care primesc mai multe plicuri cu același conținut.

Dacă aceste adrese de email redundante persistă în inconsecvența numelui sau a adresei, frustrarea poate ajunge chiar și până la abordarea pe cale legală.

Adăugați acesteia și costurile de marketing și transport legate de adrese greșite de corespondență (dintre care unele nu vor fi returnate sau, în cazul în care vor fi returnate, pot fi acum considerate ca fiind deja utilizate și plătite ca atare), amplasarea problemelor începe acum să iasă în evidență. În plus, dacă un singur client este inclus în baza de date de mai multe ori, de fiecare dată cu un alt număr de identificare, compania nu va putea determina adevăratul volum de vânzări pe clientul respectiv. Organizația ar putea fi chiar

în situația stânjenitoare de a încerca să vândă un produs pe care clientul l-a procurat deja de la companie!

Tehnici și Soluții

Calitatea datelor este un proces care implică mai multe faze, de captare de date, integrare de date, profilare de date, curățare și augmentare (îmbogățire) de date.

Identificarea și colectarea datelor

O bază de date bună presupune identificarea și colectarea datelor la sursă. Aceasta poate însemna o varietate foarte mare de surse și metode de importare, care pot fi efectuate atât în timp real cât și pe module. Clienții care fac cumpărături online, problemele și rezoluțiile trimise către call center-uri, datele colectate la diverse case în momentul plăților, sunt exemple de surse primare de date.

Deși datele pot fi curățate în orice moment, este optim ca acuratețea datelor să fie asigurată de la început. Odată ce o eroare este introdusă în sistem, corectarea ulterioară este mai costisitoare și este mult mai bine să se evite apariția acestor erori de la bun început.

Este interesant de avut în vedere că pe baza "order forms" completate on-line, potențialii devin furnizori de date. Totuși, spre deosebire de cei instruiți în "data entry" respectiv în completarea fiecărei coloane a bazei de date, mulți consumatori sunt satisfăcuți prin a introduce numai datele necesare pentru a lansa comanda, restul datelor rămânând necompletate.

Identificarea și corectarea erorilor din start reduce apariția problemelor ulterioare, pentru aceasta fiind recomandată utilizarea unui software care poate verifica o adresă pentru a vedea dacă aceasta există în mod real.

O adresă de livrare invalidă denotă o problemă de introducere a datelor, dar într-un mediu potrivit ar putea fi corectată în timp real. Alte software-uri pot determina județul și jurisdicțiile fiscale locale și ratele fiscale aferente. Sistemele software sunt foarte eficiente când se implementează pe datele de intrare, integrate cu sistemul de procesare a tranzacțiilor, putând fi apoi disponibile în timp real. Chiar dacă există persoane calificate care introduc datele în mod profesional, cum este cazul în care se plasează o comandă prin telefon, este important ca datele să fie colectate într-un format adecvat atât pentru utilizarea lor imediată (de exemplu, pentru transport și plată) cât și pentru utilizarea pe

termen lung (de exemplu, în scopuri analitice, cum ar fi identificarea viitoarelor posibilități de up-sell și cross-sell).

O organizație de servicii medicale care a achiziționat alte organizații mai mici cu același profil a descoperit că una din acestea a fost neglijentă în procedurile de introducere a datelor. Numerele de asistență socială folosite de pacienți pentru identificare au fost introduse în spațiul dedicat numelui pacientului. La prima vedere se pare că plasarea numărului era aleator, ori înaintea numelui, ori la sfârșitul câmpului de prenume, sau chiar între ele. Doar în momentul în care organizația a achiziționat un software dotat cu un program de recunoaștere a tipului de eroare s-a reușit furnizarea de informații corecte despre fiecare pacient în parte.

Integrarea datelor

Organizațiile colectează date din surse multiple, fiecare dintre acestea având formatul lor propriu. Integrarea de date presupune combinarea datelor din aceste surse dispersate. Acestea pot include datele conținute în aplicații Enterprise, baze de date orientate pe un singur obiect, cum ar fi jurnalele de apel de la un birou de asistență pentru clienți. În timp ce multe organizații au încercat să dezvolte propriile lor programe de software pentru integrarea datelor, chiar și cei care au reușit, și-au dat repede seama că aceasta devine o sarcină fără de sfârșit deoarece modulele componente necesită menținerea și sincronizarea permanentă cu fiecare nouă versiune a aplicației software. Cele mai multe organizații au descoperit că este mai eficient să utilizeze un software de integrare produs de terți, decât să încerce să dezvolte și să mențină codul propriu.

Profilarea datelor

Deseori documentația fișierului (metadata) nu

reflectă cu exactitate conținutul efectiv al fișierului. Acest lucru se întâmplă adesea atunci când programatorii au modificat continuu aplicația proprie, adăugând câmpuri noi bazei de date sau alipind acesteia date noi fără a documenta acest lucru.

Prin examinarea individuală a fiecărei coloane și câmp în parte, instrumentele de profilare a datelor pot determina dacă datele sunt concludente și conforme cu conținutul lor, inclusiv tipul de date, valorile admise și intervalele de valoare. În plus, ele pot valida legăturile intra-record și inter-fișier. Verificarea că sexul unei persoane nu vine în contradicție cu formula de adresare asociată cu numele lui, sau a ei, este un exemplu de dependență intra-record; verificarea faptului că o parte din numărul unei comenzi se potrivește cu o parte din numărul de pe fișierul master este un exemplu de dependență inter-fișier.

Un exemplu este verificarea dacă salariul unui angajat aparține intervalului specificat în codul funcției sale. Pentru a descoperi dacă aceste date sunt în concordanță, trebuie să fie comparat codul corespunzător salariului cu cel corespunzător funcției angajatului.

Profilarea datelor însoțește sau precede adesea procesul de curățare a datelor în scopul validării rezultatelor. Cu cât se profilează mai devreme datele cu atât mai repede se pot descoperi și corecta eventualele inadvertențe.

Curățarea datelor

Odată ce datele au fost colectate și adunate din diverse surse, ele trebuiesc curățate pentru a elimina erorile și inconsecvențele. Verificarea și corectarea datelor pot minimiza erorile, dar nu garantează că surprind erorile din interiorul sistemului.

V-ați întâlnit cu o situație asemănătoare?

Problemă: După ce a primit cele mai recente situații și statistici trimestriale din companie, șeful de vânzări este oarecum alarmat să vadă că numărul total de clienți a scăzut de la 700 la 600 în ultimul trimestru. Cu toate acestea, după ce a cerut agenților de vânzări o informare cu privire la numărul de clienți noi și numărul de clienți pierduți, numerele indică faptul că societatea a obținut 60 de noi clienți în timp ce pierde numai 10, câștig net de 50. Este raportul greșit?

Concluzie: Raportul este corect. Aparent numărul de clienți nu era adevărat, deoarece compania a descoperit că unii clienți erau reprezentați de mai multe ori în baza de date. În timpul acestui trimestru s-au descoperit 150 de "clienți duplicați". După curățarea și consolidarea bazei de date, numărul adevărat al clienților este de 550, și nu 700 cât a fost raportat trimestrul trecut. Adăugarea câștigului net de 50 de clienți care au apărut în cursul trimestrului la adevăratul număr 550 la sfârșitul trimestrului anterior, demonstrează că numărul corect este într-adevăr 600.

Spre exemplu, dacă fiecare department își crează propriul fișier de furnizori și folosește diferiți identificatori pentru aceiași furnizor, este necesară consolidarea acestor fișiere pentru a regăsi acei furnizori care par a fi diferiți, dar de fapt sunt unul singur. Curățarea datelor trebuie să asigure că valorile datelor sunt actuale.

Cand comparăm două înregistrări pentru a determina dacă conțin aceleași valori, este de preferat ca datele să aibă același format. Numerele de telefon, data nașterii, numele, adresele, codul numeric personal (spre exemplu client, angajat, furnizor, produs, cod de procedură, etc.) pot fi mai ușor comparate dacă au valori standardizate, adică au un anumit format. Dacă diverse unelte de integrare a datelor pot converti datele din diverse surse la un format dorit—spre exemplu pentru numerele de telefon pot fi eliminate caracterele ne-numerice cum sunt parantezele, punctele zecimale sau virgulele—abilitățile lor de curățare avansată a datelor sunt totuși limitate. Ele nu pot verifica, spre exemplu, dacă prefixul unui număr de telefon fix este corespunzător codului poștal al adresei respective.

Multe dintre uneltele pure de integrare a datelor nu pot analiza direct adresele după valori standard (exemplu St de la Stradă) sau să convertească o înregistrare ne-structurată în câmpuri separate (exemplu împărțirea unui câmp de nume, în câmpuri separate pentru Nume, Prenume sau Titlu) sau să efectueze compararea datelor astfel încât să identifice că 'Strava de la 2/3B/16 Libertății' este probabil aceeași persoană cu 'Neluțu Marcel Strava de la Nr.2, Bl.3, Sc.B, Ap.16 Bvd Libertății'. Numai o unealtă de curățare a datelor specifică țării poate recunoaște că în București, Splai este un alt nume pentru Splaiul Independenței.

Mergând mai departe, abilitatea de a lega membrii individuali ai unei familii din aceeași locuință este în mod evident funcția uneltei de curățare a datelor și nu a celei de integrare.

O soluție bună trebuie să rezolve cu acuratețe majoritatea problemelor legate de curățarea datelor și să le identifice pe cele care nu pot fi curățate pentru a fi investigate individual. Unele produse permit utilizatorilor să specifice un prag de ambiguitate înainte de a fi solicitată implicarea operatorului uman.

Puterea soft-ului ce asigură calitatea datelor este în mod cert crescută prin integrarea într-o

tehnologie robustă și de abilitatea de a permite împărțirea sarcinilor. Acest lucru este, de obicei, realizat fie printr-un parteneriat tehnic strâns între furnizorii de date de calitate și integratorii de date, fie având un singur furnizor pentru amândouă, caz în care, în general, integrarea este mai transparentă și mai strictă.

În sumar, curățarea datelor implică:

- Conversia câmpurilor la un format (un proces adesea realizat odată cu faza de integrare a datelor).
- Selectarea înregistrărilor astfel încât să se poată converti grupurile de câmpuri nestructurate în componentele lor individuale
- Identificarea și corectarea erorilor
- Eliminarea inconsecventelor
- Analiza înregistrărilor pentru a elimina duplicatele
- Completarea valorilor lipsă

Îmbunătățirea datelor

Companiile colectează date și din propriul sistem, fiind nevoie în mod frecvent ca bazele de date să fie îmbogațite cu alte date din surse externe. Date demografice, informații fiscale, date geografice cum ar fi latitudinea/longitudinea, date de recensământ, informații de credit și stil de viață sunt categorii de informații care pot îmbogăți înregistrările clientului.

Datele clientului beneficiază în mod frecvent de valoare adăugată prin alipirea de date geografice. Acestea pot fi folosite pentru a determina dacă un asigurat este în vecinătatea unui râu sau a unei zone inundabile, sau dacă un client care dorește servicii ADSL este situat la o distanță optimă de punctul de comutare.

În timp ce atenția a fost concentrată pe creșterea valorii datelor prin îmbogățirea înregistrărilor clientului pentru scopuri de marketing cum ar fi identificarea potențialilor clienți, aplicabilitatea datelor s-a extins la mai multe arii. Acestea pot include înregistrări medicale folosite pentru studierea și descoperirea posibilelor caracteristici comune într-un grup de pacienți cu același diagnostic sau boală sau chiar și aplicații de securitate care implică identificarea suspectilor de terrorism.

Astăzi, calitatea datelor este mai importantă ca niciodată

În trecut, multe organizații nu conștientizau cu adevărat nevoia datelor de calitate până când această necesitate nu devenea evidentă după

eșuarea inițiativei unei rentabilizări, de obicei legată de managementul relației cu clienții. Dacă organizației îi ia timp îndelungat și efort suplimentar să rezolve cu succes dilema legată de calitatea datelor, momentul este pierdut, iar încrederea și buna reputație a sponsorilor săi sunt de multe ori grav afectate.

O organizație care privește în perspectivă trebuie să includă calitatea datelor ca și parte integrantă pentru operațiunile sale zilnice. Deși, în mod cert, acest lucru nu poate fi obținut peste noapte, inițiativa oficială recentă de reglementare și securitate solicită compatibilitatea cu Solvency II.

Tari de pe toate continentele au reguli stricte legate de acuratețea și actualitatea datelor: spre exemplu, Departamentul Trezoreriei al Oficiului de Control al Activelor Străine ale SUA (OFAC), Sarbanes-Oxley, USA Patriot Act, Health Insurance Portability și Accountability Act (HIPAA) [Solvency II în Europa], toate solicită baze de date solid fundamentate.

Preocupările pentru siguranța publică și confidențialitatea individuală vor provoca, chiar și pentru acele companii rămase în urmă, să admită că un program eficient de curățare a datelor va deveni rapid o cerință obligatorie.

În timp ce mulți cred că Solvency II se adresează în special asigurărilor de sănătate, aceasta include, de asemenea, cerințele de raportare care vizează identificarea și eliminarea fraudelor, în parte prin recunoașterea cererilor necorespunzătoare și inconsecvențe în plata daunelor. Abilitatea de a face legătura cu precizie între ofertanții de servicii medicale, procedurile aplicate și medicație se vor baza pe date consistente, de o înaltă calitate.

Oficiul de Control al Activelor Straine (OFAC) al

Departamentul Trezoreriei SUA administrează și impune sancțiuni economice împotriva țărilor și grupurilor de indivizi, cum ar fi teroriști sau traficanți de droguri. OFAC publică o listă a Națiunilor Special Desemnate (SDNs) și a Persoanelor Blocate. Companiilor și persoanelor americane le este, în general, interzis să aibe contact cu orice persoană sau organizație de pe listă și sunt pasibili de primirea unor sancțiuni pentru încălcarea acestor reguli.

Sarbanes-Oxley cere persoanelor cu funcții de CEO și CFO să certifice exactitatea declarațiilor fiscale ale companiilor acestora. Cu toate acestea, dacă datele care stau la baza sistemului intern de control sunt sărace, nu contează cât este de bine făcut soft-ul, cât de încrezători se vor simți acei executivi să certifice rapoartele rezultate?

The Patriot Act cere instituțiilor financiare să verifice identitatea clienților și tranzacțiile realizate de aceștia pentru a identifica, de exemplu, posibilele activități de spălare de bani.

Acestea par să cuprindă o suită de reguli și reglementări care privesc orice organizație cu conexiuni internaționale este și va fi sub presiunea crescută de a menține date curate și auditabile.

Un vânzător online care solicita utilizatorilor să se înregistreze înainte de plasarea unei comenzi considera că are 18 milioane de clienți. Compania era pe cale de a face o investiție majoră într-un sistem CRM performant când a descoperit că mulți clienți aveau numele unor personaje din filmele de animație sau numele unor politicieni renumiți; unul dintre cele mai des folosite era numele lui George Bush de pe 1600 Pennsylvania Avenue. După eliminarea intrărilor evident fictive, numărul intrărilor valide s-a redus la aproximativ 6.5

S-a întâmplat vreodată în organizația dumneavoastră?

Problema: Un client sună la centrul de suport al companiei. Specialistul de la Centru accesează baza de date cu clienții să vadă istoricul clientului. Ca specialist, Suportul cere detalii despre natura problemei curente, dar clientul țipă la telefon: "este aceeași problema de care m-am plâns și săptămâna trecută într-un e-mail. De ce sunt nevoit să explic din nou?!"

Cauza: Compania crede că implementarea unei soluții CRM va permite urmărirea tuturor acțiunilor legate de un client sau un prospect, totuși e-mail-ul clientului a fost inițial trimis de pe computerul de

serviciu și nu de pe cel de acasă. Deoarece înregistrările clientului din baza de date a companiei pot interoga numele, adresa, numărul de telefon și e-mail-ul personal (de acasă), adresa de e-mail de serviciu nu se lega de aceste înregistrări. În baza lui de date era, totuși, alt set de înregistrări care conțineau adresa de e-mail de serviciu, dar care se presupuneau a fi ale altui client cu același nume! Pur și simplu întrebând clientul dacă a raportat deja problema putea fi ajutat suportul tehnic să facă conexiunea între cele două seturi de înregistrări și să-l ajute prompt cu rezolvarea problemei.

milioane. Folosind un software de de-duplicare și cerând simpla înregistrare din nou atunci când utilizatorii își uitau parolele, numărul clienților unici a fost redus la mai puțin de 2.5 milioane. Software-ul CRM deja existent a putut procesa cu ușurință acest volum, iar achiziționarea unui software nou și costisitor a fost amânată pentru o perioadă nedefinită.

Calitatea datelor este un proces continuu și un mod de viață pentru corporatii

Deși calitatea datelor poate să fie imperativ necesară, cu siguranță ea nu este un proces de moment. Chiar dacă, într-o lume ideală, toate datele ar fi verificate și curățate înainte de a intra în sistemul operațional, trăim într-o lume în care lucrurile se schimbă. Dacă o femeie se căsătorește și își schimbă numele de familie, dacă un client se mută la o nouă adresă sau dacă un furnizor fuzionează sau este preluat de un altul, o companie ar trebui să aibe posibilitatea să țină cont de aceste schimbări mai degrabă decât să presupună că orice tranzacție nouă vine de la un client nou. Calitatea datelor ar trebui să fie incorporată într-un sistem integrat de aplicații. De exemplu, în USA, cele mai multe sisteme de navigare on-line au fost la început concentrate pe colectarea adreselor în cel mai acceptabil format pentru Serviciile Poștale U.S. sau pentru alte servicii poștale străine; ele nu au fost astfel proiectate din perspectiva de a integra întreaga activitate a unui client.

În cazul în care nu există deja, un program riguros de management al datelor incluzând definiții și reguli bine conturate ar trebui implementate. Aplicațiile dezvoltate intern ar trebui să adere la aceste norme iar software-urile folosite ar trebui evaluate în funcție de cât de bine se conformează acestora sau cât de ușor sunt de modificat pentru a le respecta. În plus, pentru evitarea problemelor de nume și adresă menționate anterior, un program de management al datelor ar fi prevenit deficiențele metrice asociate cu problema Mars Orbiter anterior citată. Dacă definiția datelor ar include și atribute ca unitatea de măsură, acest lucru ar oferi consecvență astfel încât atunci când se consolidează numerele să nu existe riscul comparării de ‘mere cu pere’ sau și mai probabil ‘euro cu dolari’, ‘metri cu inci’.

Auditările ar trebui să vizeze verificarea calității datelor. Adăugând un câmp numit ‘date cu ultimele actualizări’ la fiecare fișier intrat (de ex. angajați, clienți, produse, vânzări, structură organizatorică) ajută la menținerea ‘prospețimii’

datelor și poate, de asemenea, să fie utilizat pentru extragerea selecției de auditat. În plus, un istoric al evoluției datelor ar trebui menținut astfel încât datele sursă să fie ușor de urmărit. O modalitate simplă de a realiza acest lucru constă în păstrarea valorilor originale ale datelor și crearea de câmpuri noi pentru valorile noi standardizate.

Componentele Soluției de Asigurare a Calității Datelor

Începeți cu începutul: O soluție solidă de asigurare a calitatii datelor începe de la sursa datelor de la care se pot minimiza erorile prin standardizare și verificare. De vreme ce standardizarea numelor și a adreselor sunt de la sine înțelese, pașii următori ar trebui făcuți pentru a preveni ca aceeași entitate (de ex., clienți, părți, angajați, vânzători, proceduri medicale, etc.) să fie reprezentată de mai multe ori în aceeași bază de date, de fiecare dată cu alt identificator-cheie.

Verificați Ipotezele: O investiție în profilarea datelor valorează cât o mie de ipoteze.

Un produs bun de asigurare a calității datelor va oferi o gama largă de instrumente și extensii opționale. Funcțiile de standardizare, administrare și de-duplicare, ar trebui să poată rula pe o varietate largă de platforme de hardware și să fie compatibile cu elementele opționale, dar nu mai puțin importante, de integrare, profilare și îmbunătățire a datelor. Această întregă suită ar trebui să poată opera de asemenea în timp real și în flux continuu. Operarea în timp real este cea mai potrivită pentru colectarea datelor, în timp ce operarea în flux continuu se potrivește mai bine verificărilor periodice și auditărilor. Ar trebui să existe un aspect și funcționalități comune pentru toate capacitățile care să permită unei adrese să fie verificate cu o singură interfață.

Să ofere posibilitatea unui audit test: Chiar și cele mai bune programe software specializate pot produce rezultate rezonabile dar incorecte care vor necesita investigație adițională. Un mod simplu de lucru constă în reținerea și nu înlocuirea valorilor originale cu ‘valorilor standardizate’.

Menținerea Actualității Datelor : Trebuie să admitem că oamenii își vor schimba numele și adresele în timp ce furnizorii, de asemenea, se vor muta sau se vor uni cu alte organizații. Adițional, numele străzilor și codurile poștale se pot schimba. Este necesară monitorizarea consecventă a acestor aspecte prin intermediul unor auditări, bazate pe

actualizări periodice folosind date îmbunătățite de la terți care asta au ca obiect de activitate.

Implementați bunele tehnici de management al datelor: Un program de management la nivel de organizație care să asigure consecvență în formatele standard ale bazelor de date, ale unităților de măsură și normele aplicabile vă ajută foarte mult la prevenirea și identificarea inconsecvenței datelor.

Datele reprezintă bunuri majore ale companiilor și ele trebuie tratate ca atare.

Managementul datelor implică, de asemenea, alegerea identificatorilor cheie. De exemplu, în timp ce unele organizații folosesc numărul de telefon ca și identificator cheie al clienților, date valoroase ale clienților sunt adesea pierdute când clientul se mută sau își schimbă numărul de telefon.

Să fie capabil să integreze o largă varietate de surse: Deoarece multe organizații folosesc pachete de software de la mulți furnizori, abilitatea de a "captura datele de la sursă" trebuie extinsă în vederea acceptării unei varietăți de surse. Chiar dacă o organizație dorește să-și dezvolte un software propriu de integrare de date, acesta va trebui să țină pasul cu fluxul constant de modificări pe care aplicația furnizorului le lansează în fiecare an, în particular acelea care modifică structura stabilită a fișierelor de date. Din fericire, software-urile de integrare de date sunt create pentru abordarea acestei probleme atât la nivel de baze de date cât și la nivel de aplicații specifice.

Luați în considerare companiile specializate în îmbunătățirea datelor: Este puțin probabil ca propriul dvs. sistem să acopere toate informațiile pe care le doriți. Furnizorii externi de date îmbunătățite și îmbogățite pot fi o sursă valoroasă pentru obținerea de clienți noi sau de date de noi furnizori. Asta poate include numere de telefon, suprapuneri de date geografice, demografice și informații despre stilul de viață al consumatorilor.

Oferă o varietate de căi de a accesa Software-uri dedicate calității datelor: În mod ideal, un

software cu un concept corespunzător în privința calității datelor ar trebui să fie disponibil la nevoie. Tehnologia potrivită ar trebui să fie oferită la alegere ca un serviciu găzduit sau ca o ofertă de licențiere. Adicional, ar trebui să funcționeze în timp real dar și să ofere posibilitatea implementării unui flux continuu. În timp ce majoritatea instrumentelor oferă propriile interfețe de programare a aplicațiilor particularizate, ar trebui să existe o capacitate curentă sau planificare pe termen scurt pentru disponibilitatea ca serviciu web.

Lucrați cu furnizori de date experimentați: Așa cum apelații la un centru de suport al unui furnizor nu vor să li se spună că sursa problemei este alt furnizor pe care ei trebuie să-l contacteze personal pentru rezolvarea problemei, tot astfel companiile vor să comunice cu furnizori cu experiență. Aceștia vor fi capabili să își asume responsabilitatea pentru soluția completă având în vedere că la randul lor se bazează pe parteneriate puternice menite să acopere eventuale nevoi aditionale.

Sumar

Calitatea datelor stă la baza tuturor proceselor operaționale și analitice ale unei organizații. Calitatea datelor va avea un impact major în randamentul și eficacitatea organizației. Altădată ignorată, calitatea datelor este recunoscută azi de multe organizații ca fiind parte din activitățile de fiecare zi și ca nefiind limitată la inițiativele centrate pe client. Aceste organizații realizează în același timp că acest proces trebuie să fie unul continuu și nu unul de moment.

Din fericire, cu cât buna practică în calitatea datelor este mai mult implementată cu atât mai ușoară devine menținerea calității. Calitatea datelor include atât curățarea cât și integrarea datelor, reguli consecvente și o mare doză de 'bun simț'. Succesul implică nu numai uneltele de software și bunele practici în managementul datelor, dar, cel mai important, implică totală dedicare în a face din calitatea datelor prioritatea numărul unu.

SMARTaddress

Curățarea &/sau Geocodarea datelor

SMARTaddress® a fost special proiectat să curețe și să geocodeze adresele românești. În conformitate cu cele mai bune practici în domeniu, se bazează pe date de intrare din surse multiple care includ Poșta Română.

SMARTaddress are la bază un instrument sofisticat de extracție a adreselor care separa date ne-structurate în componente separate (județe, localități, tip de stradă, nume de stradă, și numere de casă).

Datele tip adresă sunt apoi comparate cu cea mai cuprinzătoare și standardizată bază de date pentru România iar datele de ieșire sunt disponibile cu și fără diacritice și cu coduri poștale.

În plus, SMARTaddress încorporează un geocoder care înglobează coordonate pentru toate adresele din România.

Pitney Bowes

Business Insight

Data Integration

Sagent® Dataflow (Spectrum) de la Pitney Bowes este un motor de integrare puternic și flexibil care aduce împreună date din surse eterogene într-o vizualizare unică. Acesta oferă posibilitatea utilizării unei game largi de instrumente de transformare pentru a lega și a curăța componentele non-adresă ale datelor în vederea valorificării lor optime.

Odată ce datele sunt transformate, prin folosirea uneltelor de analiză Sagent Data Flow (Spectrum) se pot extrage informații importante și rapoarte utile. Astfel, organizațiile beneficiază de o mai bună înțelegere a datelor critice pentru afacerea sa iar clienții săi au posibilitatea de a lua decizii informate bazate pe date reale.

Interfața intuitivă face ca utilizatorul final să stăpânească soluția mai rapid decât cu oricare alt instrument BI. O dezvoltare cu mare impact vizual, ajută la crearea de date sofisticate, rapid și eficient.

Despre Geo Strategies

Geo Strategies a fost fondată în 1993 și a devenit lider în furnizarea de soluții de informație și analiza referențiată geografic pentru consumatori și zone în România.

Expertiza de bază este axată pe produsele de informație, instrumente analitice și de modelare, segmentarea consumatorilor, servicii de date, consultanță, training și management de proiecte.

Această expertiză este aliniată celor mai bune practici și metodologii internaționale în domeniu, la care s-au adăugat inovații semnificative și valoare care îi conferă relevanță și acționabilitate în context local.

Geo Strategies este partener Pitney Bowes, Experian și Navteq (parte a Nokia group) și dezvoltă produse și servicii de management al calității datelor (curățare, manipulare și integrare de date), profilare socio și geo-demografică, sisteme de țargetare, baze de date și acoperire cu hărți la standarde consacrate pentru utilizare în GIS și aplicații Business Intelligence.

Serviciile de curățare, îmbogățire și integrare de date, segmentarea Mosaic a consumatorilor, analizele și profilele Micromarketer pentru zone granulare de acțiune, sunt doar câteva dintre produsele și serviciile oferite de Geo Strategies pentru România.

Contactați-ne la info@geo-strategies.com sau sunați la (+44 1223 205080 sau +40 269 210832) pentru a discuta cerințele dumneavoastră.

Pentru a afla mai mult despre ceea ce Geo Strategies poate face pentru afacerea dumneavoastră în România vizitați www.geo-address.com www.mosaic-blue-strategies.com

Geo Strategies

