Teksti: Aija Palomäki

Profilointityökalu tiedon laadun arvioinnissa

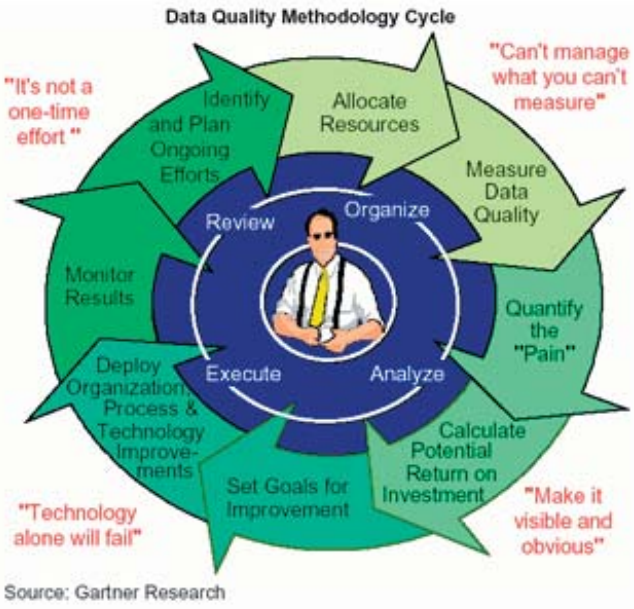
Aija Palomäki toimii informaatioarkkitehtinä Nokiassa. Hän työskentelee tällä hetkellä tiedon profiloinnin parissa sekä johtaa SAP-järjestelmän tietomallin sovittamista Nokian yritystason tietomalliin. Aija on DAMA Finlandin perustajajäsen ja toimii vuonna 2006 myös DAMA Finlandin varapuheenjohtajana.

Nykyaikaiset profilointityökalut ovat tehokkaita numeronmurskaajia - niiden avulla tietovarasto- ja integraatiohankkeet voivat arvioida lähdejärjestelmien tiedon rakennetta ja laatua kattavasti jo hankkeen alkuvaiheissa. Vaikka hyödyt ovat nähtävissä, uudentyyppisen työkalun tuonti yritykseen saattaa kuitenkin olla odotettua haastavampaa. Tiedon laadun pitäminen hyvänä pysyvästi – vastakohtana kertaluonteiselle, jotakin tiettyä tarkoitusta varten tehdylle assessoinnille – edellyttää pysyvien roolien ja vastuiden asettamista paikoilleen.

Tiedon laadusta

Tiedon laadun pitäminen hyvänä pysyvästi on haastavaa (Kuva 1). Se edellyttää pysyvien roolien ja vastuiden asettamista paikoilleen. Keskeisin näistä rooleista on tiedon huoltaja, nokialaisessa kielessä ”Global Concept Owner”, joka vastaa tiedosta ja sen rakenteesta koko sen elinkaaren ajan – esimerkiksi tuotetiedosta tuotteen suunnittelusta valmistukseen, myyntiin, toimitukseen ja jopa huoltoon asti.

Muistettakoon kuitenkin, että tiedon laatu ja ”hyvyys” on aina suhteellinen käsite, ja oikeastaan ainoa luotettava tiedon laadun mittari on se, palveleeko tieto nykytilassaan yrityksen liiketoimintaa riittävästi, vai muodostuuko esimerkiksi tiedon oikeellisuudesta, ajantasaisuudesta tai saatavuudesta este onnistuneelle liiketoiminnalle. Tiedon laadun parantaminen on usein merkittävä ja pitkäjänteinen investointi joka voidaan perustella vain kustannussäästöillä tai uusilla liiketoiminnan mahdollisuuksilla.



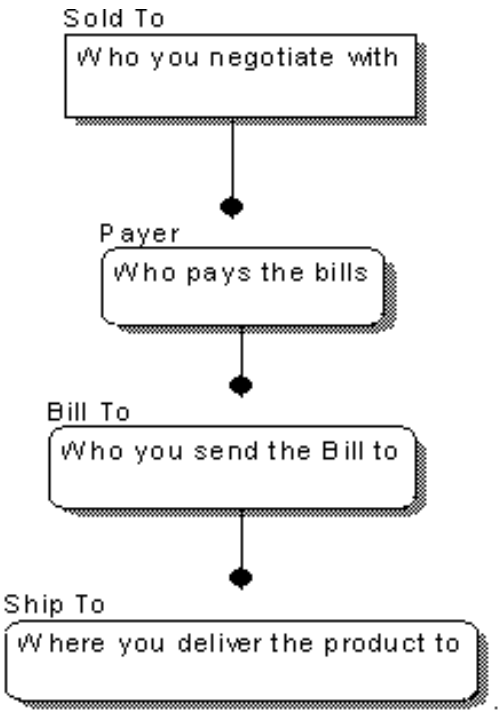
Kuva 1. Tiedon laadun jatkuva arviointi. Lähde: Gartner Research.

Lisäksi, suuressa yrityksessä tiedon laatu näyttäytyy erilaisena siitä riippuen, mistä liiketoimintaprosessista käsin asiaa tarkastellaan: tavaran toimitusprosessin kannalta riittävä tiedon laatu esimerkiksi tuotteiden sarjanumeroiden tallentamisen suhteen, saattaakin muodostua pullonkaulaksi huoltoprosessin onnistumiselle, mikäli riittävän usea sarjanumero on jäänyt tallentamatta! Usein laatuparannukset olisi tehtävä siinä liiketoimintaprosessissa, joka ei itse suoranaisesti ko. laadunparannuksesta hyödy.

Jos tiedon elinkaarta tarkastellaan ravintoketjuna siten, että tieto syntyy ravintoketjun alkupäässä ja tiedon viimeinen hyödyntäjä on ravintoketjun huipulla, tiedon laadun parannusvaatimukset kohdistuvat useimmiten juuri ravintoketjun alkupäähän, ja suurimmat laatuongelmat koetaan huipulla. Ravintoketjun alkupään motivoiminen laatuparannuksiin on aina oma haasteensa, koska rahalliset ja liiketoimintahyödyt materialisoituvat vasta huipulla.

Profilointityökalu voi osoittautua hyödylliseksi apuvälineeksi matkalla pysyvään laadunparannukseen – sen avulla voidaan nostaa esiin ilmeisimmät tiedon laatuun liittyvät ongelmat.

Kuva 2. Jotkin logistiikkatapahtuman osapuolet ovat 1:N relaatiossa toistensa suhteen. Lähde: Data Management issues - During and After a Major SAP Implementation Anthony Harris, Enterprise Data Architect, Air Product



Tiedon profiloinnista yleisesti

Tiedon profiloinnilla tarkoitetaan yleensä prosessia, jossa tutkitaan jonkin olemassa olevan tietokannan sisältöä, keräten tilastollista ynnä muuta tietoa kannan taulujen sarakkeiden arvoista.

Sarakkeista tyypillisesti haettavaa tietoa ovat:

1) Sarakkeiden todelliset arvojoukot: vastaavatko kannasta löydetyt arvot sitä mitä sarakkeille on alun perin määritelty, esim. ovatko sarjanumero-kentän arvot numeerisia vai onko joukossa merkkijonoja

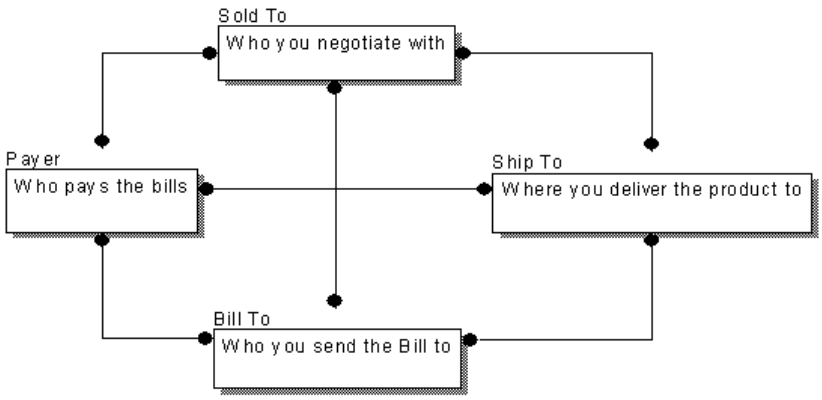
2) Sarakkeiden todelliset tietotyypit

3) Sarakkeiden sisäinen rakenne, esimerkiksi suomalaisten puhelinnumeroiden tulisi noudattaa tiettyä rakennetta

4) Tietyn arvon esiintymistiheys, esimerkiksi jos suurin osa yrityksen asiakaskuntaa sijaitsee Kaliforniassa, useimmiten esiintyvän osavaltiokoodin tulisi olla ”CA”.

5) Tilastolliset tiedot sarakkeen arvoista, esimerkiksi minimiarvo, maksimiarvo, keskiarvo ja mediaani

Kuva 3. Kaikki logistiikkatapahtuman osapuolet ovat M:N relaatiossa toistensa suhteen. Lähde: Data Management issues - During and After a Major SAP Implementation Anthony Harris, Enterprise Data Architect, Air Product



Lisäksi profilointityökalut pystyvät löytämään riippuvuuksia eri sarakkeiden arvojen välillä sekä silloin kun sarakkeet sijaitsevat samassa taulussa, että silloin kun ne ovat eri tauluissa.

Näiden analyysien tuloksena usea profilointityökalu pystyy ehdottamaan kolmannessa normaalimuodossa olevaa versiota kannan rakenteesta sekä tarvittaessa myös populoimaan uuden rakenteen mukaisen kantaversion vanhan kannan arvoilla.

Profiloinnin hyötyjä

Yhä useampi yritys on valinnut strategiakseen pakettiohjelmistojen laajamittaisen hyödyntämisen. Usein pakettiohjelmistoilla myös hallitaan yrityksen kaikkein keskeisimpiä tietoja, ja tarve hyödyntää näitä tietoja eri puolilla organisaatiota on suuri.

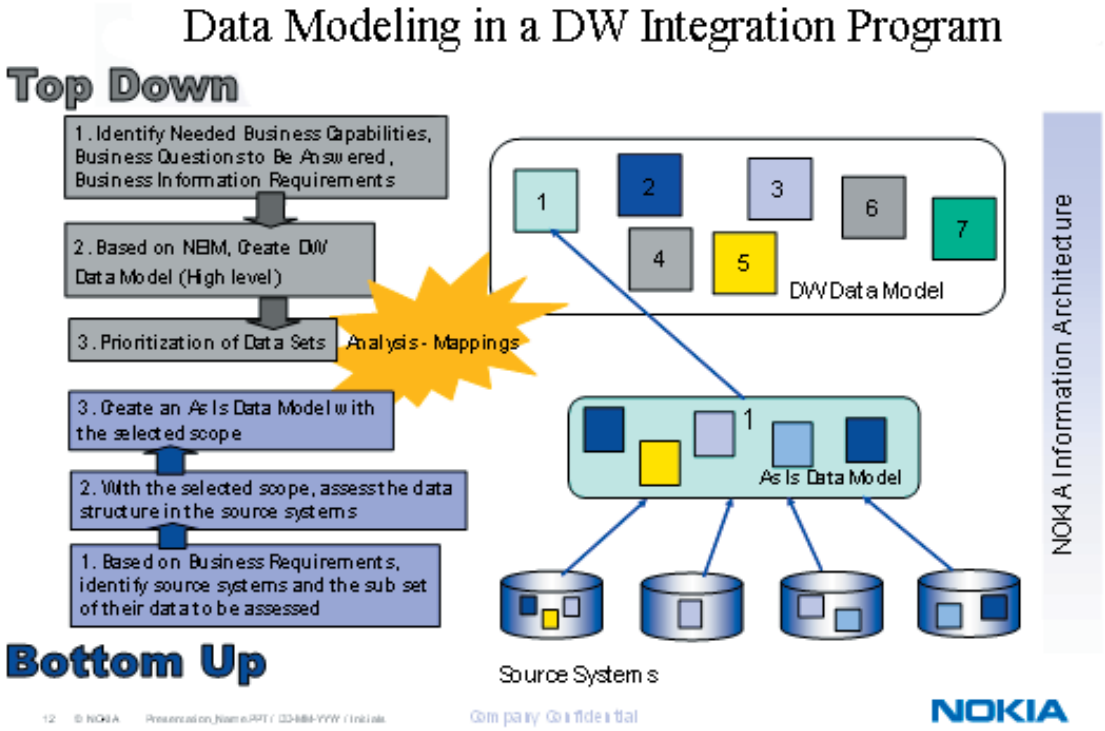
Pakettiohjelmistot on kuitenkin rakennettu yleiskäyttöisiksi, jolloin ne perustuvat sisällöllisesti varsin laajaan ja rikkaaseen tietomalliin: käytettävissä on suuri joukko tauluja, joiden sarakkeiden määrä on tyypillisesti useita kymmeniä, jopa satoja.

Pakettiohjelmiston konfigurointi ja käyttöönotto yrityksessä on usein karsintaa: päätetään mitkä tarjolla olevista tietomallin osista otetaan käyttöön ja mitkä jätetään tyhjiksi. Näitä päätöksiä ei aina dokumentoida riittävän huolellisesti ja vuosien mittaan, kun ohjelmiston käyttö laajenee, myös lisätauluja ja sarakkeita otetaan käyttöön. Joitain sarakkeita saatetaan myös käyttää useampaankin eri tarkoitukseen, esimerkiksi siitä riippuen, mikä organisaatioyksikkö niitä käyttää.

Pakettiohjelmisto sisältää useimmiten paketin valmistajan dokumentoiman tietomallin – paketin omistajan räätälöimää tietomallia se ei kuitenkaan välttämättä sisällä.

Tarkastellaanpa esimerkiksi hetki liiketoimintatapahtuman osapuolia niin kuin ne SAP-järjestelmässä määritellään. Konfiguroimaton SAP-järjestelmä määrittelee seuraavat logistiikan osapuolet: Customer, SoldTo, BillTo, Payer ja ShipTo. SAP ei kuitenkaan määrittele näiden osapuolten suhteita toisiinsa, vaan nämä relaatiot määritellään vasta kun SAP-järjestelmää konfiguroidaan asiakkaan tarpeisiin.

Lopputulema voi olla yhtä hyvin se, että SoldTo, Payer, BillTo, ShipTo -osapuolet ovat yhden suhde moneen -relaatiossa toisiinsa (Kuva 2) tai se, että niillä kaikilla on monta-moneen relaatio toisiinsa (Kuva 3). Nopein ja tehokkain tapa selvittää kuinka ko. konfiguraatio on tehty, on profiloida SAP-järjestelmän sisältämät osapuolitideot työkalun avulla.



Kuva 4. Tiedon mallinnus -prosessi ylhäältä alaspäin ja alhaalta ylöspäin

Näistä syistä profilointi on merkittävä vaihtoehto jokaiselle hankkeelle, joka tarvitsee käyttöönsä pakettiohjelmiston hallinnoimia tietoja. Profiloinnilla voidaan varmistaa nopeasti ja tehokkaasti, mitkä sarakkeet ovat käytössä, mitkä ovat sarakkeiden käytössä olevat arvojoukot, mitkä ovat sarakkeiden todelliset tietotyypit, ja mitkä ovat taulujen relaatiot suhteessa toisiinsa.

Lisäksi paketin käytössä olevasta tietomallista voidaan luoda kolmannessa normaalimuodossa oleva versio, joka on hyvä lähtökohta esim. tietovaraston atomaarisen tason tietomalliksi – sehän kuvaa jo valmiiksi sitä tietorakennetta, mikä lähdejärjestelmästä voidaan saada tietovarastoon.

Pakettiohjelmisto tai ei, profilointi tarjoaa joka tapauksessa ainutlaatuisen tavan varmistaa, onko kulloisenkin hankkeen tavoittelema liiketoimintahyöty saavutettavissa lähdejärjestelmiin talletettujen tietojen todellisten arvojen perusteella. Tällainen ”järjellisyystarkistus” olisikin syytä tehdä proaktiivisesti jo hankkeen alkuvaiheessa.

Profilointi voi paljastaa senkaltaisia tiedon saatavuus- tai laatuongelmia, että hankkeessa tavoiteltujen hyötyjen saavuttaminen voi olla mahdotonta tai edellyttää aikaaviepiä muutoksia liiketoimintaprosessissa, lähdejärjestelmien toteutuksessa tai asianosaisten henkilöiden rooleissa ja vastuissa. Tällöin on syytä toteuttaa aikaavievät. muutokset ensin ja käynnistää tietojärjestelmähanke vasta, kun onnistumisen edellytykset ovat olemassa. Tietojärjestelmähankkeen viivästäminen on tällöin toki parempi vaihtoehto kuin se, että hankalat tiedon laatuongelmat löydetään vasta järjestelmätestauksen yhteydessä, jolloin kalliit järjestelmäinvestoinnit on jo tehty.

Profiloinnin esteitä

Profilointityökalun käyttöönotto yrityksessä on aina työtapojen ja -kulttuurin muutos ja sellaisena muutoshallinnallinen haaste. Profiloinnin avulla halutaan ehkä tutkia yrityksen tuloksen ja toiminnan kannalta sensitiivistä tietoa, tai sensitiivistä tietoa on tutkimuksen kohteena olevassa tietokannassa, vaikka juuri sitä ei tutkittaisikaan. Profilointityökalun kytkeminen tuotantokäytössä olevaan kantaan on usein käytännössä mahdotonta suorityskykyrasitteen takia. Tällöin joudutaan pyytämään koko kannan varmuuskopiota ja siirtämään se tutkimusympäristöön. Niinpä moni tiedon omistaja ei välttämättä myönnä ilomielin lupaa kannan kopion siirtämiseen, mikäli profiloinnin reunaehdot eivät ole riittävän hyvin selvillä.

Käytännön kokemus osoittaa, että profilointipyynnön yhteydessä on tärkeää keskustella huolellisesti läpi ainakin seuraavat seikat:

1. Tarvitaanko tiedon todellisia arvoja, eikö kannan dokumentointiin tutustuminen riitä?
2. Mitä tietoa tarkkaan ottaen halutaan tutkia?
3. Minne tieto halutaan viedä?
4. Ketkä saavat pääsyn kannan kopioon?
5. Miten profilointi tapahtuu?
6. Kuinka profiloinnin tuloksia käytetään?
7. Kuinka kauan analyysi kestää ja paljonko liiketoiminnan asiantuntijoiden aikaa siihen tarvitaan?
8. Mitä kannan kopiolle tapahtuu analyysin jälkeen?

1 Profilointi on ainoa tapa selvittää, mitkä sarakkeet ovat todella käytössä, mitkä ovat niiden todelliset arvojoukot ja tietotyypit. Näiden selvittäminen ei onnistu lukemalla kannan dokumententteja.

Lisäksi profiloinnilla saadaan arvokasta tietoa lähdejärjestelmän tiedon todellisesta rakenteesta sekä tietojen suhteesta toisiinsa, mikä nopeuttaa tietojärjestelmähankkeen tietomallinnustyötä merkittävästi perinteiseen haastattelupohjaiseen ”ylhäältä alaspäin” -mallinnukseen verrattuna, ks. kuva 4. Profilointi paljastaa myös duplikaatit ja tiedon laatuongelmat ja on siten ehdoton edellytys hankkeen suunnittelulle ja riskinhallinnalle.

2 Koko kannan profilointi on harvoin tarpeen, vaan mieluummin on selvitettävä hankkeen kannalta olennaiset tiedot ja keskityttävä niihin. Haluttu profiloinnin kohde on kommunikoitava tiedon omistajalle liiketoiminnan, ei atk-toteutuksen kielellä. Sensitiiviset tiedot voidaan peittää kopiossa, tai profilointityökalun käyttäjäksi osoitetaan henkilöt, joilla on valtuudet käsitellä ko. tietoa.

3 Tutkimusympäristön turvallisuus- ym. järjestelyt tulee käydä läpi tiedon omistajan kanssa.

4 Profilointiyökalun käyttäjät tulee määritellä. Heidän tulee olla mieluiten yrityksen sisäisiä työntekijöitä ja heillä tulee olla nimetty työnjohtaja, joka vastaa henkilökohtaisesti siitä, ettei tiedon väärinkäytöksiä tapahdu. Muilla henkilöillä ei saa olla pääsyä tutkimusympäristöön eikä tutkimuksen kohteena olevaan kannan kopioon.

5 Käytössä olevan profilointityökalun nimi ja sen käyttötapa tulee käydä läpi. Yleensä työkalu analysoi ensin kaikkien taulujen kaikki sarkkeet ja tämä työvaihe kestää päiviä. Seuraavaksi voidaan alkaa käsitellä löydöksiä yhdessä liiketoiminta-asiantuntijoiden kanssa sekä selvittää erilaisia spesifisiä tiedon laatuun liittyviä kysymyksiä.

6 Profiloinnin tuloksilla tulee olla sama tietoturvallisuusluokitus kuin profiloinnin kohteena olevalla tiedolla. Tulokset käydään oletusarvoisesti läpi vain tiedon omistajan kanssa, ja ainoastaan tiedon omistajan luvalla muilla foorumeilla. Tuloksia tullaan kuitenkin käyttämään hankkeen suunnittelussa.

7 Profiloinnin kesto riippuu analysoitavista tietomääristä ja siksi siihen ei voi antaa yleispätevää vastausta. Liiketoiminnan asiantuntijoiden panosta tarvitaan profiloinnissa esiintulleiden avointen kysymysten selvittämiseen, nyrkkisääntönä noin tunnin verran päivässä profiloinnin keston ajan.

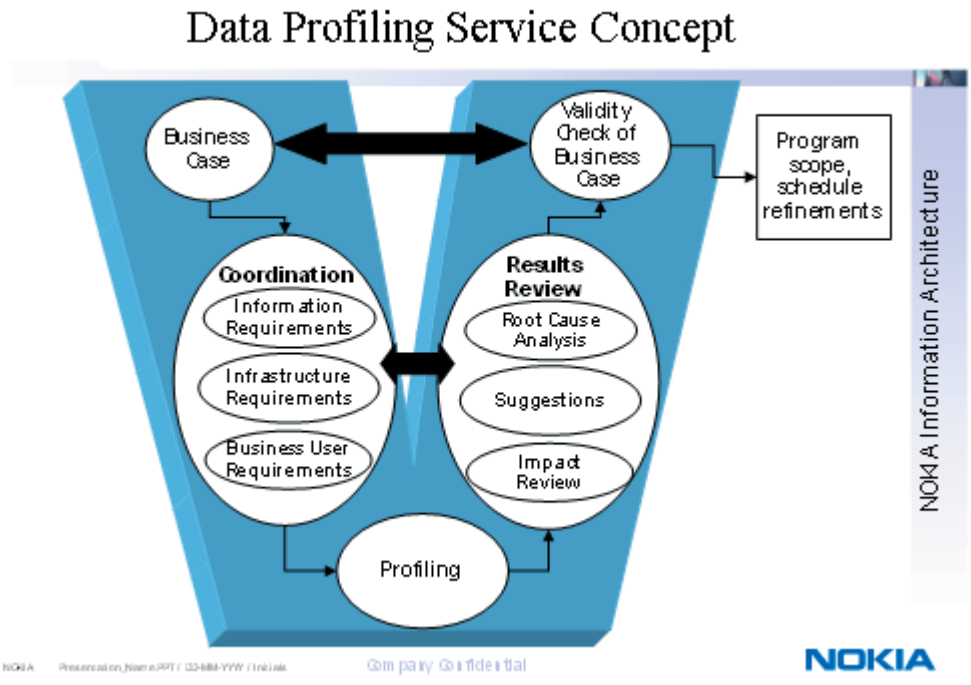
8 Kun profilointi on valmis, kannan kopiot tuhotaan. Myös kaikki kannan tietoja sisältävät raportit ja otokset tuhotaan välittömästi.

Profiloinnin organisoinnista

Suuressa yrityksessä hankekohtainen profilointi kannattaa järjestää yrityksen sisäisenä palveluna, jonka puoleen integraatio- ja tietovarastointihankkeet voivat kääntyä varmistaakseen hankkeensa tavoittelemat rahalliset hyödyt sekä hankkeen toteutus- ja riskienhallintasuunnitelman (Kuva 5).

Palvelu tarvitsee päällikön, nokialaisessa kielessä ”Service Manager”, joka vastaa profilointi- ja tietopyyntöjen koordinoinnista ja niputtamisesta siten, että samoja tiedon omistajia ja asiantuntijoita ei turhaan vaivata moneen kertaan. Lisäksi tarvitaan asianomistaja, ”Concept Owner”, joka vastaa profilointilöydösten analysoinnista yhdessä liiketoiminnan asiantuntijoiden kanssa sekä sopii korjaavista toimista tiedon omistajien kanssa. Varsinaisen teknisen profiloinnin tekevät tehtävään koulutetut tietokannan hoitajat.

Kuva 5. Profiloinnin organisointi V-mallin mukaisesti



Yhteenveto

Profiloinnin hyödyt ovat selvät – automatisoidun työkalun avulla laatua voidaan arvoida kattavasti ja tehokkaasti. Lopullisen arvion laadun hyvyydestä voivat kuitenkin antaa vain liiketoiminnan edustajat. Tiedon laadun pitäminen yhteisesti sovitulla tasolla edellyttää pysyvien liiketoimintaroolien ja -vastuiden asettamista paikoilleen sekä mahdollisesti myös muutoksia prosesseissa ja tietojärjestelmien toteutuksessa. Puhumme siis isoista ja aikaa vievistä asioista.