

**Taru Väre**

# **MASTER DATA**

Alma Talent  
Helsinki 2019

Copyright © 2019 Alma Talent Oy ja kirjoittaja

ISBN 978-952-14-3580-5

ISBN 978-952-14-3581-2 (sähkökirja)

ISBN 978-952-14-3582-9 (verkkokirja)

Kansi: Outi Pallari

Taitto: Sirpa Puntti

BALTO print, Liettua 2019

Anna palautetta: *protoimitus@almamedia.fi*

Tilaa Master data -kirja Alma Talent Shopista: [shop.almatalent.fi](http://shop.almatalent.fi)

# Master datan hallinnan käsikirja

## liiketoiminnan kehittäjille ja päättäjille

Kaikessa liiketoiminnassa käytetään nykyään tavalla tai toisella dataa. Uudet tekniset innovaatiot, edistyneen analytiikan tuleminen, koneälyt ja muut kuumat trendit asettavat lisää mahdollisuuksia datan hyödyntämiselle. Jotta näistä uutuuksista voi saada oikeasti uutta liiketoimintaa tai parantaa jo olemassa olevaa toimintaa, perusasiat täytyy olla kunnossa. Mikään talo ei kestä uusia kerroksia, elleivät sen perustukset ja runko ole kunnossa. Tämä kirja antaa kaikille liiketoiminnan kehitystehtävissä toimiville henkilöille käytännön ohjeet siihen, miten varmistaa, että organisaation data on siinä kunnossa, että sen päälle voi rakentaa uutta liiketoimintaa.

Kirjan tarkoitus on paikata aukko suomalaisessa ammattikirjallisuudessa. Datanhallintaan ja erityisesti master datan käsittelyyn ei löydy kirjallisuutta, joka auttaisi liiketoiminnan ihmisiä ymmärtämään, mitä heidän pitäisi tehdä saadakseen datasta hyötyä.

### Kirjailijasta

Olen rakentanut omaa taloa usean vuoden ajan, joten rakentamiseen liittyvät piirteet ovat käytännössä tuttuja. Kun on kokenut omakotitalossa sotkun, jonka kellarissa pettänyt viemäriputki voi saada aikaan, on helppo verrata kokemusta siihen, mitä työssään näkee huonolaatuisen datan aiheuttavan.

Työelämässä olen toiminut 10 vuotta erilaisissa IT- ja data hankkeissa, tällä hetkellä senior konsulttina Talent Base Oy:ssä. Viimeiset kuusi vuotta olen puhtaasti keskittynyt master datan hallintaan ja sen eri osa-alueisiin. Palkkatyön ohessa toimin kansainvälisen datan hallinnan yhdistyksen DAMA:n Suomen jaoston hallituksessa. Lisätietoa yhdistyksestä löytyy [www.damafinland.fi](http://www.damafinland.fi). Olen suorittanut DAMA:n data-ammattilaisen CDMP Practitioner -sertifikaatin ensimmäisenä Suomessa.

## Kiitokset

Tämä kirja ei olisi syntynyt ilman sitä suurta kannustusta ja tukea.

Erityiset kiitokset haluan antaa:

puolisolleni Miikalle,

Sisarelleni Helille sekä

kollegoilleni Kealle, Sinille, Jounille ja Terolle.

Kiitos myös Wärtsilä Oyj:n master data tiimille, joka kertoi oman tarinansa master datan parissa.

DAMA-verkoston kautta sain myös suurta apua, joista erityiskiitos henkilöille Nicola Askham ja Aaron Zornes.

# Sisällys

<b>Master datan hallinnan käsikirja</b> .....	5
Kirjailijasta .....	5
Kiitokset .....	6
<b>JOHDANTO</b> .....	11
<b>DATA TERMINOLOGIA</b> .....	13
Sanastoa .....	14
Erilaiset datatyypit .....	15
<b>OSA 1 MASTER DATA TEKEE BISNEKSEN</b> .....	19
<b>1 Ei master dataa – ei liiketoimintaa</b> .....	21
Master datan määritelmä .....	23
Ajallinen master data .....	25
Rakenteellinen ja taloudellinen master data .....	26
Master datan tasot .....	27
Mistä master data koostuu .....	29
Master datan attribuutit .....	33
Master datan hallinta .....	37
<b>2 Master datan elämän tarkoitus</b> .....	42
Master datan hallinnan hyödyt .....	45
Tietosuoja-asetus ja master data .....	51
Miten kartoittaa master datan hallinnan hyötyjä .....	52
Datan hallinnan strategia .....	58

<b>3 Ylpeästi ainutlaatuinen master data</b> .....	62
Master datan omistaja .....	63
Master datan kehittäminen tehokkaasti .....	64
<b>OSA 2 MASTER DATAN HALLINTA KÄYTÄNNÖSSÄ</b> .....	67
<b>4 Datan rakennepiirrustukset</b> .....	69
Tietomallit ja niiden tarkoitus .....	70
Tietomallin muodostaminen .....	72
Datastandardi ja sen merkitys .....	78
Datastandardin muodostaminen .....	81
Valmiiden tietomallien ja standardien hyödyntäminen .....	86
<b>5 Datan kasvattaminen hyvässä ohjauksessa</b> .....	88
Master datan hallinnan prosessit .....	88
Master datan hallinnan prosessien kehittäminen .....	91
Datan elinkaari .....	96
<b>6 Helpon liikkumisen varmistaminen</b> .....	99
Organisaatio, jossa on vähän järjestelmiä .....	104
Organisaatio, jossa on keskimääräisesti eri järjestelmiä .....	107
Organisaatio, jossa on monipuolisesti erilaisia järjestelmiä .....	111
MDM-järjestelmä ja tietovarastointi .....	115
<b>7 Oikein suunnitellut kodit</b> .....	118
Master datan hallinnan järjestelmien ominaisuudet .....	119
Master datan hallinnan järjestelmät .....	134
<b>OSA 3 HYVIEN PÄÄTÖSTEN TEKEMINEN</b> .....	139
<b>8 Hyvä hallintamalli tuo hyviä päätöksiä</b> .....	141
Välttämätön hallintamalli .....	142
Mitä datan hallintamalli sisältää? .....	143
Datan hallintamallin merkitys .....	144
Datan hallintamallin pääkulmakivet .....	147
<b>9 Master datan ydinperhe</b> .....	151
Datan hallintamallin yleinen rakenne .....	151
Datan hallintamallin tyypilliset tuotokset .....	160

<b>10 Oman perhemallin muodostaminen</b> .....	164
Organisaatiokulttuurit .....	164
Vastuun sijoittaminen oikein .....	169
Datan hallintamallin vastuunjako .....	173
<b>11 Mallin vieminen käytäntöön</b> .....	178
Hallintamallin jalkauttamisen tärkeimmät periaatteet .....	179
Datan hallintamallin jalkauttamisen johtaja .....	182
Viesti tehokkaasti .....	184

## **OSA 4 LAADUKKAAN LOPPUTULOKSEN VARMISTAMINEN** .....

.....	189
<b>12 Hyvänlaatuinen master data</b> .....	191
Datan laadun hallinta .....	191
Miksi panostaa datan laatuun? .....	194
Mitä on hyvä datan laatu? .....	198
<b>13 Moniulotteinen laatu</b> .....	203
Sisällön kattavuus .....	204
Datan oikeellisuus .....	206
Ainutlaatuisuus .....	208
Vaatimustenmukaisuus .....	209
Yhteneväisyys .....	210
Eheys .....	212
Järkevyys .....	213
Ajanmukaisuus .....	214
Muutama sana mittaamisesta .....	216
<b>14 Laadun päättymätön tarina</b> .....	219
Jatkuvan parantamisen malli .....	221
Datan laadun virheiden juurisyitä .....	225
Datan laadun parantamisen työkalut .....	228
<b>Loppusanat: Miten päästä alkuun</b> .....	232

## **LIITE: MERKITYKSELLISIMMÄT**

<b>MDM-JÄRJESTELMIEN TOIMITTAJAT</b> .....	235
--	-----





# JOHDANTO

Törmäsin ensimmäisen kerran huonolaatuisen master datan aiheuttamiin ongelmiin vuonna 2008. Silloin yritykseen, jossa työskentelin, hankittiin uusi raportointijärjestelmä, johon yhdistettiin dataa raportointia varten useista eri käytössä olevista toiminnanohjausjärjestelmistä.

Työskentelin tuon uuden raportointijärjestelmän pääkäyttäjänä. Organisaatioon suunniteltiin uudistusta ja sitä varten tarvittiin erilaisia raportteja pohjatiedoksi. Työn piti olla yksinkertaista; yhdistelin isoihin raportteihin tuotetietoa. Kaikissa järjestelmissä oli lähtökohtana sama tuotehierarkia. Harmi vain, että hierarkiaa ei oltu käytetty samalla tavalla ja ne sisälsivät monia poikkeuksia. Muistan edelleen, miten tuskaista oli yhdistellä tuota dataa. En tiedä, mitä virheitä raportissa lopulta oli kaiken sen käsin tehdyn korjailun jälkeen. Toivottavasti kukaan ei kuitenkaan menettänyt työtään uudistuksessa näiden mahdollisten virheiden vuoksi.

Huhtikuussa vuonna 2017 talouslehti Forbes kirjoitti, että dataa syntyy vuosittain 16.3 ZB verran (zettabyte = triljoona gigatavua). Luku on täysin käsittämätön. Erilaiset laitteet – mobiililaitteet, älytelevisiot, autot, lentokoneet ja niin edelleen – tuottavat arvion mukaan hämmentävät 2.5 kvintiljoonaa tavua dataa joka päivä (luku, jossa on 30 nollaa).

Dataa siis tulvii valtavat määrät joka päivä.

Tätä valtavaa datamäärää kaikki haluavat hyödyntää. Ymmärtää paremmin omaa liiketoimintaa tai ympäröivää maailmaa. Kehittää uudenlaista liiketoimintaa, saada enemmän tuottoja, enemmän rahaa.

Mikäli datan laatu on kuitenkin sillä tasolla, että sitä hyödyntäköseen jonkun täytyy käsin yhdistellä asioita, ollaan nopeasti pahassa pulassa. Käsin tekeminen on aina hitaampaa kuin automatisoitu prosessi. Lisäksi se on alttiimpaa virheille. Mitä useammin dataa käsitellään manuaalisesti, sitä suurempi riski on tehdä virheitä.

Datan laadun parantaminen lähtee aina master datan laadun parantamisesta.

Edellä kerrottu hetki raporttien parissa oli minulle se, jolloin ymmärsin, että data itsessään on merkittävää. Vielä silloin en tiennyt, mitä master data on ja miksi se on niin tärkeää. Tässä kirjassa kerroin oman tarinani master datan maailmassa ja autan sinua ymmärtämään, mitä master data on, miksi se on tärkeää ja miten sitä voi hallita niin, että organisaationne voi saada täyden hyödyn datasta.

## DATA TERMINOLOGIA

Data on vaikea sana suomentaa. Suomen kielestä löytyy valitettavasti parhaana vastineena vain sana *tieto*, joka voi tarkoittaa montaa muutaakin asiaa. Itse ajattelen tiedon enemmänkin työstettynä informaationa. Data on kuitenkin informaation raaka-ainetta – pieniä tiedon palasia, joiden merkitys riippuu siitä, kuka niitä käyttää tai mihin. Informaatio taas on hiukan laajempia tiedon palasia, joihin on jo liitetty merkitystä. Informaatiota tulkitsemalla voidaan muodostaa ymmärrystä tilanteesta, tehdä päätelmiä ja päättää toimenpiteistä. Mutta tämä tiedon kokonaisketju lähtee liikkeelle niistä pienistä, joskus hyvin merkityksettömältä tuntuvista tiedon palasista, datasta.

Jos puhuisin tiedonhallinnasta, asettaisin lukijani vaarallisille vesille. Tiedonhallinta sanana viittaa suomenkielessä yhtä lailla datanhallintaan (data management), opitun tai hankitun tiedonhallintaan (knowledge management), tai vaikkapa dokumenttien hallintaan. Jotta tarkoittamani tiedonhallinta ymmärretään riittävän yksiselitteisesti, käytän tässä kirjassa osittain englanninkielistä ilmaisua datanhallinta.

Datan parissa työskentelevät tietävät, että maailmasta löytyy monta erilaista dataa määrittelevää termiä. Seuraavana on kirjassa esiintyvistä termeistä lyhyet määritelmät, joihin voit palata yhä uudelleen. Tämä osio on hyvä ainakin silmäillä läpi ennen kuin uppoudut itse kirjaan, jotta uudet termit eivät tunnu liian hämmentäviltä.

## Sanastoa

**Attribuutti.** Entiteettiä kuvaava tai selittävä ominaisuus.

Jokainen attribuutti kertoo entiteetistä yhden asian. Näiden attribuuttien eli ominaisuuksien avulla entiteetti voidaan tunnistaa ja yksilöidä. Attribuuttien tulee kattaa kaikki se data, jota liiketoiminta tarvitsee toimiakseen.

Koska entiteetti on yhtä todellisen elämän asiaa edustava käsite, jokaisella entiteetillä on ainutlaatuinen lista ominaisuuksia. Luonnollisella henkilöllä nämä ominaisuudet sisältävät esimerkiksi henkilötunnuksen tai syntymäpäivän, nimen ja kotiosoitteen. Oikeushenkilöllä ne voivat olla yritystunnus, toiminimi ja rekisteröintipäivä.

**Domain.** Looginen dataa ryhmittelevä kokonaisuus.

Domainit ovat master datan ylin taso. Ne kertovat erilaiset dataryhmät, joita voidaan hallita ja omistaa organisaatiossa omina kokonaisuuksinaan. Domainit liittyvät toisiinsa, mutta yksi entiteetti kuuluu aina vain yhteen domainiin.

Domain on erityisesti liiketoiminnan tapa ryhmitellä dataa. Siksi domainien nimitykset pitää päättää organisaation oman toiminnan mukaisesti.

Yleisimpiä domaineja ovat asiakas, tuote, materiaali ja sopimus.

**Entiteetti.** Tosielämän asia, jota data kuvaa.

Entiteetit ovat käsitteitä, jolla pyritään kuvaamaan todellista elämää mahdollisimman tarkasti. Data on todellisen elämän digitaalinen versio, joten jokaista todellisen elämän asiaa vastaan on oltava oma edustajansa eli entiteettinsä. Entiteetti on todellisen elämän asioita luokitteleva käsite, joka voidaan selvästi yksilöidä ja tunnistaa.

Esimerkiksi asiakas on yrityksen näkökulma henkilöihin ja toisiin yrityksiin, jotka ostavat heidän palveluitaan. Todellisessa elämässä henkilö on oma asiansa ja yritys on toinen. Yritys voi sisältää useita todellisen elämän asioita. Oikeushenkilö, lain edessä vastuullinen

yritys, on oma todellinen asia ja yrityksen toimipaikka on toinen. Näin asiakas voi koostua kolmesta eri entiteetistä, jotka ovat luonnollinen henkilö, oikeushenkilö ja toimipaikka.

**Relaatio.** Kahden domainin tai entiteetin välinen suhde.

Jokaisella domainilla ja entiteetillä on suhde vähintään yhteen domainiin tai entiteettiin. Usein näillä on suhde moneen eri domainiin tai entiteettiin. Relaatio ilmaisee tuon suhteen ja voi kuvata myös minkälainen riippuvuus näiden käsitteiden välillä on. Relaatioiden ilmaisemiseen käytetään erilaisia merkintätapoja, joita kutsutaan tietomallintamisessa notaatioiksi.

**Tietue.** Tietojärjestelmässä oleva yksi itsenäinen, looginen tietokokonaisuus.

Lähtökohtaisesti tietue on yksi entiteetin ilmentymä, joka sisältää kaikki yksilöidyn entiteetin ominaisuudet. Tietueella on yleensä joku ainutlaatuinen avaintieto, jolla se yksilöidään tietojärjestelmässä. Tällainen avaintieto voi olla esimerkiksi järjestelmän antama tunniste henkilölle.

Relaatiotietokannassa yhtä tietuetta edustaa yksi rivi. Jos relaatiotietokannasta viedään dataa Exceliin, rakenne on Excelissä sama. Usein tietuetasosta saatetaankin puhua rivinä ja kun tarkoitetaan yksittäistä attribuuttia, siitä puhutaan sarakkeena.

## Erilaiset datatyypit

**Analyttinen data.** Tilastotieteen ja matematiikan keinoin hankittua lisädataa organisaation tapahtumatiedoista. Tällöin organisaation omaan dataan saatetaan liittää ulkopuolista dataa uusien näkökulmien saamiseksi. Hyvien analyttisten tulosten saaminen vaatii kuitenkin, että tapahtumatiedot on luotettavasti yhdistetty master

dataan ja sopiviin referenssitietoihin, jotka ovat analytiikassa oleellisia dimensioita, eli ulottuvuuksia, joilla dataa ryhmitellään.

Eniten riskejä ja ongelmia aiheuttaa se tieto, jota et tiedä, että et tiedä. Jos siis et tiedä, missä kunnossa organisaatiosi referenssi- ja master data ovat, edistyneenkin analytiikan käyttäminen siihen dataan perustuen voi aiheuttaa suuria riskejä.

**Master data**, suomenkielisiltä kutsumanimiltään avain-, ydin- tai perustieto.

Liiketoiminnan ytimessä olevaa luonteeltaan kohtalaisen pysyvää dataa, jota käytetään organisaatiossa laajalti. Master data vastaa kysymyksiin: mistä toimintamme muodostuu – mitä organisaatiomme tekee (tuotteet, palvelut), missä toimintaa on (erityyppiset toimipaiikat, geolokaatiot) ja keiden kanssa toimintaa tehdään (asiakkaat, toimittajat, kumppanit, henkilökunta, muut osapuolet).

Master data voidaan jakaa myös erilaisiin tarkentaviin luokkiin. On rakenteellista master dataa, joka sisältää usein organisaation taloudelliset rakenteet ja hierarkiat. Joidenkin määritelmien mukaan on myös väliaikaista tai aikaan sidottua master dataa, joka voi sisältää esimerkiksi hintatiedot tai tuotteiden erilaiset kokoonpanot. Itse kutsun sitä ajalliseksi master dataksi, joka on transaktiodatan ja master datan välimuoto. Dataa, joka ei ole aivan master dataa, mutta enemmän kuin pelkästään transaktiodataa.

Master datan hallinta on kriittistä organisaation päivittäisen toiminnan sujumiselle sekä organisaation liiketoiminnan tapahtumien luotettavalle raportoinnille.

**Metadata** eli kuvailu- tai määrittelydata.

”Tietoa tiedosta” on yleisin metadatan määritelmä. Metadataa käytetään niin tiedostojen kuin dokumenttienkin hallinnassa ja sen määritelmä vaihtelee hiukan riippuen, mistä käyttötapauksesta

puhutaan. Tässä kirjassa metadataa käsitellään vain master datan hallinnan näkökannalta.

Metadataa annetaan master datalle kaikille sen eri osille. Domai-  
nille annetaan sanallinen määritelmä, esimerkiksi mikä on yksi asia-  
kas. Attribuuteille metadataa on sanallisen määritelmien lisäksi eri-  
laiset muodolliset määritelmät. Näitä ovat tieto siitä onko attribuutti  
tekstimuotoinen, numeerinen vai kenties päivämäärämuotoinen tie-  
to. Myös datan maksimi- tai minimipituudet, pakollisuus, tai vaati-  
mus tietystä muodosta ovat metadataa. Vaatimus tietystä muodosta  
on esimerkiksi vaatimus siitä, että datan on noudatettava suomalai-  
sen henkilötunnuksen muotoa.

Metadatan määrittely osana master datan hallintaa on kriittistä  
master datan laadun varmistamiseksi.

### **Referenssidata** eli luokittelu- ja viitetieto.

Tämä datatyyppe ri ryhmittelee tai luokittelee muuta dataa. Esi-  
merkkejä tällaisista luokitteludatasta on tuote- ja asiakasryhmät. Se  
voi myös antaa yleisesti käytettävän koodiston, listan ennalta määri-  
teltyjä tietoja, joita käytetään muun datan yhtenäisyyden varmista-  
miseksi. Yleisesti käytettyä referenssidataa ovat muun muassa maa-  
koodilistat, kuten ISO-standardin ISO 3166 lista maakoodeista, tai  
kotimainen postinumeroluettelo.

Referenssidatan määrittely ja hallinta osana master datan hallin-  
taa on tärkeää luotettavalle raportoinnille.

### **Transaktiodata** eli tapahtumatieto.

Tapahtumatieto on varsinaisesta liiketoiminnasta talteen jäävää  
dataa. Se sisältää kuittitietoja, osto- tai myyntisopimuksia, lähetyshis-  
toja, vastaanottokuittauksia, asiakaspalvelutilanteen kirjaukset, pal-  
velupyynnöt jne. Eli kaikki se, mitä organisaatiossa päivittäin teh-  
dään sen perustarkoituksen täyttämiseksi.

**Big dataksi** kutsuttu datatyyppejä on oikeastaan tapahtumatietoa, jota muodostuu nopeasti suurina määrinä.

Yksi esimerkki tästä on sensoridata, jota valmistavan teollisuuden laitteet voivat lähettää internetin yli valmistajalle, sen kumppaneille ja asiakkaalle itselleen. Big dataksi kutsutaan yleisesti myös ei-tyypillistä kuva- tai videomuotoista dataa, jota ei useinkaan voida tallettaa samoihin tietokantoihin kuin mitä tyypillistä organisaation tapahtumatietoa. Big dataksi kutsutaan usein myös (perinteisen) organisaation perustoimintojen ulkopuolella syntyvää dataa. Tällaista dataa syntyy sosiaalisessa mediassa huokeita määriä.

Erilaisen transaktiodatan hyödyntäminen raportoinnissa ja analysoinnissa auttaa organisaatiota ymmärtämään, mitä liiketoiminnassa tapahtuu ja miksi.