



10 'fjellvettregler'
En diskusjon om informasjonskvalitet

Per Anders Waaler
per.anders.waaler@bWise.no

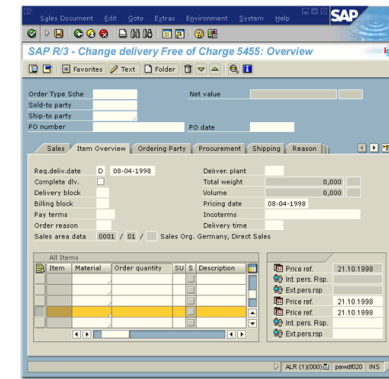
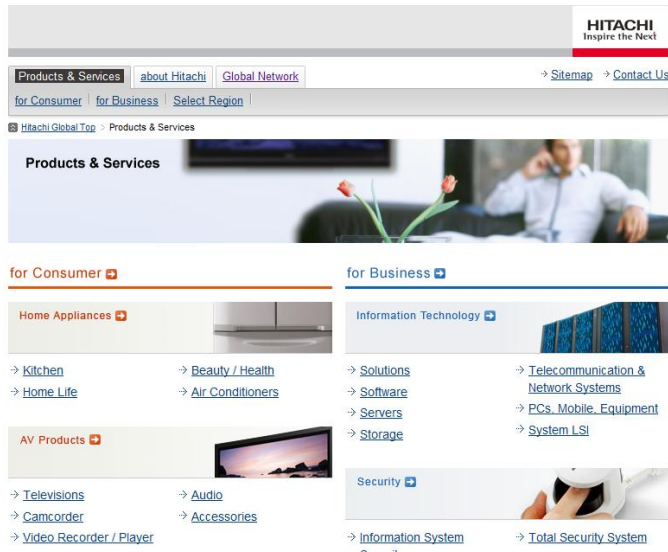
Agenda

- 1. Hvem er bWise og hva kan vi om datakvalitet?**
 - 1.1. Dette forklares best ved å se på noen kundeeeksempler

- 2. Definisjoner er fint, så vi vet hvor vi har hverandre**
 - 2.1. Er kvalitet målbart?
 - 2.2. Kan vi enes om hva som er 'bra' og ikke?

- 3. Regler**
 - 3.1. Våre regler
 - 3.2. Har du en god en?

Kundeeksempel 1. Hitachi



- Global kunde av SAP. Vi jobbet med Hitachi Japan (HQ)
- Vi fant samme kunde registrert opptil 12 ganger
- Konsolidering av kundeaktivitet (ordre, reskontro) i bakkant
- De hadde 34 andre 'plants' med tilsvarende utfordring
- De tok problemet ved roten, data-entry, i alle prosesser som fanger kundedata

Kundeeksempel 2. Schibsted / Finn



FINN bedrifter

Forside Person Bedrift Kart

Bedrifter
Søk på restaurant, snekker, hotell, læsesmed...
snekker i buskerud

SBS
Snekker Service AS

3 921 treff [Tilspil 1 treff i nettbutikk](#) [Sortering](#) [søkeord](#) | [firmanavn](#)

Brannsikring
Postboks 193, 3340 Åmot, tlf. 95015411
Vi utfører alt innen salg og montering av Norgeslignen, stiger, feiestiger, varmepumper, tak- og blikkenslagerarbeid, snekkertjenester, maling, stiger og stillaser. Ta kontakt for befaring og prisstilbud.
[Mer info](#) | [E-post](#) | [Reaksjon](#) | [Roller](#) | [Endre oppføring](#)

Elementmontasje AS www.elementmontasje.no
Industriveien 1, 3300 Hokksund, tlf. 32750808
Utleie av kraner og løfteutstyr.
[Mer info](#) | [Hjemmeside](#) | [E-post](#) | [Reaksjon](#) | [Roller](#) | [Endre oppføring](#)

Hemsedal Alarm og Vaktmesterservice AS www.hemsedal-alarm.no
Rampa Kjøpesenter, 3560 Hemsedal, tlf. 32067000
Din totalentreprenør av hus og hytte. Vi tar ansvaret for det du eier i Hemsedal!
[Mer info](#) | [Hjemmeside](#) | [Reaksjon](#) | [Roller](#) | [Endre oppføring](#)

Velg fylke

Bustend	3 919
Alvershus	1
Oslo	1

Velg bransje

Blyg- og tømmerlønester	3 750
Blyg- og tømmermestere	70
Snekkertjenester	62
Jernvarer	24
Møbelsnekkere	11
Lathing	8

[+ Vis flere](#)

- Ny kunde for SAP, fokus på BO Data Integrator
- Drøyt 50 informasjonskilder skal settes sammen
- Ingen sikker nøkkel, ingen fasit på adresse etc.
- Mange indikative datasett
 - Brønnøysund
 - Selskapets egen Webside
 - Konsensus i kildene

Kundeeksempel 3. BestBuy.com

The screenshot shows the BestBuy.com website interface. At the top, there's a navigation bar with 'BEST BUY' logo, 'Weekly Ad', 'Store Locator', 'Outlet Center', 'Gift Center', and 'Small Business Center'. Below this is a search bar with 'SEARCH FOR Keyword or Item #' and a 'GO' button. A banner below the search bar reads 'NO INTEREST on HDTV, home theater and Geek Squad home theater installation purchases totaling \$999 and up if paid in full within 3 years. Exclusions apply. Learn more >'. The main content area shows search results for 'lcd tv' with 212 items found. On the left, there's a 'Shop Category' sidebar with 'TV & Video (191)' selected. Below it are 'Narrow your results by:' sections for 'Customer Reviews' (Top-Rated 58) and 'Current Offers' (On Sale 84, Special Offers 107, Free Shipping 115, Online Only 65, Financing Offers 122, Outlet Center 11). A 'Brand' section lists Samsung (45), LG (24), Insignia (23), Sony (22), and Toshiba (22). The main product grid displays three items:

Product	Price
Insignia® - 32" Class 720p Flat-Panel LCD HDTV/DVD Combo	Our Price: \$469.99
Sharp AQUOS 52" Class 1080p 120Hz Flat-Panel LCD HDTV	Our Price: \$1,299.99
Dynex® - 19" Class / 720p / 60Hz / LCD HDTV	Our Price: \$179.99

Below this, there's a 'MAGNOLIA HOME THEATER' section with three items:

Product	Price
LG - 47" Class / 1080p / 240Hz / LCD HDTV	Sale: \$1,349.98
LG - 42" Class / 1080p / 240Hz / LCD HDTV	Our Price: \$1,399.98
Samsung - 55" Class / 1080p / 240Hz / LCD-LED HDTV	Sale: \$3,009.98

- BestBuy hadde brukt 3500 timer på analyse av datakvalitet
- Denne sa noe om hvor dårlig det var fatt med dem – særlig på kundedata
- Men den sa ingenting om forbedringstiltak
- Vi brukte ca 500 timer + BODI på
 - Vask og korreksjon, kopling
 - Endring av prosess for datafangst
 - Datatransport til Webben

Datakvalitetsanalyse på mindre enn 3500 timer

Column	Average string length	Median string length	Distincts	Nulls	Nulls %	Zeros	Zeros %	Blanks	Blanks %	Patterns
CustomerID	5	5	91	0	0%	n/a	n/a	0	0%	1
CompanyName	18	18	91	0	0%	n/a	n/a	0	0%	82
ContactName	13	14	91	0	0%	n/a	n/a	0	0%	42
ContactTitle	14	17	12	0	0%	n/a	n/a	0	0%	12
Address	18	18	91	0	0%	n/a	n/a	0	0%	82
City	8	8	69	0	0%	n/a	n/a	0	0%	18
Region	3	2	19	80	65.93%	n/a	n/a	0	0%	7
PostalCode	5	5	87	1	1.09%	n/a	n/a	0	0%	9
Country	5	6	21	0	0%	n/a	n/a	0	0%	8
Phone	12	12	91	0	0%	n/a	n/a	0	0%	20
Fax	12	12	20	22	24.17%	n/a	n/a	0	0%	20

Value	Records	% of total
99999	50	54.95%
9999	15	17.58%
99999-999	9	9.89%
999999	5	5.49%
9999999	2	3.30%
99999999	2	2.20%
999999999	2	2.20%
9999999999	1	1.10%
Other	1	1.10%

CustomerID	CompanyName	ContactName	ContactTitle	Address	City	Region	PostalCode	Country	Phone	Fax
CDMMI	Comércio Mineiro	Pedro Alonso	Sales Associate	Av. dos Lusiadas, 23	Sao Paulo	SP	05432-043	Brazil	(11) 555-7647	<Null>
FAMIA	Familia Arqubaldo	Ania Cruz	Marketing Assistant	Rua Orós, 92	Sao Paulo	SP	05442-030	Brazil	(11) 555-9857	<Null>
GOURL	Gourmet Lanchonetes	André Fonseca	Sales Associate	Av. Brasil, 442	Campinas	SP	04876-796	Brazil	(11) 555-9482	<Null>
HANAR	Hanari Carnes	Mario Pontes	Accounting Manager	Rua do Paço, 67	Rio de Janeiro	RJ	05454-876	Brazil	(21) 555-0091	(21) 555-8765
QUEDE	Que Delicia	Bernardo Batista	Accounting Manager	Rua da Paraifadora, 12	Rio de Janeiro	RJ	02389-673	Brazil	(21) 555-4252	(21) 555-4545
QUEEN	Queen Cozinha	Lúcia Carvalho	Marketing Assistant	Alameda dos Canários, 891	Sao Paulo	SP	05487-020	Brazil	(11) 555-1189	<Null>
RICAR	Ricardo Adocicados	Janete Limeira	Assistant Sales Agent	Av. Copacabana, 267	Rio de Janeiro	RJ	02389-890	Brazil	(21) 555-3412	<Null>
TRADH	Tradição Hipermercados	Anabela Domingues	Sales Representative	Av. Inês de Castro, 414	Sao Paulo	SP	05634-030	Brazil	(11) 555-2167	(11) 555-2168
WELLI	Wellington Importadora	Paula Parente	Sales Manager	Rua do Mercado, 12	Resende	SP	08737-363	Brazil	(14) 555-8122	<Null>

- Tekstfelt lengder
- Antall unike verdier
- Antall NULL
- Antall med BLANK
- Antall mønstre og hvilke

- Data Integrator har innebygget data-profileringsverktøy
- Du får se viktige egenskaper ved dine data
- Den viser mønstre og format
- Et supert startpunkt for forbedringsprosesser

Kundeeksempel 4. Nettbuss

■ Mange fritekstfelt i inputapplikasjonen

■ Stor idérikdom under inntasting

■ Dyrt å endre i applikasjon

■ Vi brukte BODI

- Fikset 97,8% av feilene
- Logger alt
- Write-back
- Regler i Excel

<u>Applikasjonsfelt</u>	<u>GyldigVerdi</u>	<u>KontrollVerdi</u>	<u>RegelType</u>	<u>DatabaseFelt</u>
Rullestolheis	J	J*,1,X	2	RULLESTOLHEIS
Rullestolheis	N	N*,0	2	RULLESTOLHEIS
Aircondition	SUTRAK	SUTRAK*,Sutrak*,sutrak*	2	AIRCONDITION
Aircondition	N	N*	2	AIRCONDITION
Antall sitteplasser	0..90	0..90	3	ANTSITTEPLASSER
Euro	EURO 0	Euro*0	1	EUROKLASSE
Euro	EURO 1	Euro*1	1	EUROKLASSE
Euro	EURO 2	Euro*2	1	EUROKLASSE
Euro	EURO 3	Euro*3	1	EUROKLASSE
Euro	EURO 4	Euro*4	1	EUROKLASSE
Euro	EURO 5	Euro*5	1	EUROKLASSE
Euro	EURO 6	Euro*6	1	EUROKLASSE
Euro	EEV	*EEV*	1	EUROKLASSE
Fartskr.type	DIGITAL	*Digital*	1	FARTSSKRIVERTYPE
Fartskr.type	ANALOG	*ANALOG*,*KIVE*	1	FARTSSKRIVERTYPE
Fartskr.type	N	NULL	1	FARTSSKRIVERTYPE
Antall ståplasser	0..70	0..70	3	ANTSTÅPLASSER
Tilleggsvarme	WEBASTO	WEBASTO*,	2	TILLEGGSVARMER
Tilleggsvarme	STROCO	STROCO*	2	TILLEGGSVARMER

Resultat

View Data - VASK_LOGG(Nettbus_Rapp_Base.DBO)

ABILNUMMER	FELT	NY_VERDI	GAMMEL_VERDI	BESKRIVELSE	AK
13214	Batterilader	N	NEI	Endret pga regel	J
13215	Batterilader	N	NEI	Endret pga regel	J
13215	Toalett	N	NEI	Endret pga regel	J
13215	Alkolås	N	NEI	Endret pga regel	J
13215	Tilleggsvarme	WEBASTO	WEBASTO THERMO 300	Endret pga regel	J
13215	Aircondition	N	NEI	Endret pga regel	J
13215	Guidestol	N	NEI	Endret pga regel	J
13215	Rullestolh.	N	NEI	Endret pga regel	J
13601	Rullestolh.	N	NEI	Endret pga regel	J
13601	Guidestol	N	NEI	Endret pga regel	J
13601	Aircondition	N	NEI	Endret pga regel	J
13601	Alkolås	N	NEI	Endret pga regel	J
13601	Toalett	N	NEI	Endret pga regel	J
13601	Batterilader	N	NEI	Endret pga regel	J
13602	Batterilader	N	NEI	Endret pga regel	J
13602	Toalett	N	NEI	Endret pga regel	J
13602	Alkolås	N	NEI	Endret pga regel	J
13602	Aircondition	N	NEI	Endret pga regel	J
13602	Guidestol	N	NEI	Endret pga regel	J
13602	Rullestolh.	N	NEI	Endret pga regel	J
13604	Rullestolh.	N	NEI	Endret pga regel	J
13604	Guidestol	N	NEI	Endret pga regel	J
13604	Aircondition	N	NEI	Endret pga regel	J
13604	Tilleggsvarme	WEBASTO	WEBASTO THERMO 300	Endret pga regel	J
13604	Alkolås	N	NEI	Endret pga regel	J
13604	Toalett	N	NEI	Endret pga regel	J
13604	Batterilader	N	NEI	Endret pga regel	J
14020	Rullestolh.	N	NEI	Endret pga regel	J
14022	Toalett	J	JA	Endret pga regel	J
14023	Rullestolh.	N	NEI	Endret pga regel	J
14023	Aircondition	N	NEI	Endret pga regel	J
14024	Fartskr.type	ANALOG	skive	Endret pga regel	J
14089	Guidestol	J	JA	Endret pga regel	J
14099	Guidestol	J	JA	Endret pga regel	J
14099	Fartskr.type	ANALOG	SKIVE	Endret pga regel	J
14102	Fartskr.type	ANALOG	SKIVE	Endret pga regel	J
14103	Fartskr.type	ANALOG	SKIVE	Endret pga regel	J
14104	Fartskr.type	ANALOG	SKIVE	Endret pga regel	J
14105	Fartskr.type	ANALOG	SKIVE	Endret pga regel	J
14105	Tilleggsvarme	EBERSPÄCHER	EBER.DSWS.NR.252091	Endret pga regel	J
14108	Fartskr.type	ANALOG	SKIVE	Endret pga regel	J
14109	Fartskr.type	ANALOG	SKIVE	Endret pga regel	J

Record: 1 of 1000 (sorted, filtered: (GAMMEL_VERDI IS NOT NULL) AND (BESKRI

Eksemplene viser ett vanlig problem

Kunder fra SAP ERP @ HQ

	id	lastname	firstname	address	city	state	zip
1	172-32-1176	White	Johnson	10932 Bigge Rd.	Menlo Park	CA	94025
2	213-46-8915	Green	Marjorie	309 63rd St. #411	Oakland	CA	94618
3	238-95-7766	Carson	Cheryl	589 Darwin Ln.	Berkeley	CA	94705
4	267-41-2394	O'Leary	Michael	22 Cleveland Av. #14	San Jose	CA	95128
5	274-80-9391	Straight	Dean	5420 College Av.	Oakland	CA	94609
6	341-22-1782	Smith	Meander	10 Mississippi Dr.	Lawrence	KS	66044
7	409-56-7008	Bennet	Abraham	6223 Bateman St.	Berkeley	CA	94705
8	427-17-2319	Dull	Ann	3410 Blonde St.	Palo Alto	CA	94301
9	472-27-2349	Gringlesby	Burt	PO Box 792	Palo Alto	CA	94301

Ren og pen kundeliste (med total-engasjement)

	id	lastname	firstname	address
1	172-32-1176	White	Johnson	10932 Bigge R
2	213-46-8915	Green	Marjorie	309 63rd St. #
3	238-95-7766	Carson	Cheryl	589 Darwin Ln
4	267-41-2394	O'Leary	Michael	22 Cleveland /
5	274-80-9391	Straight	Dean	5420 College /
6	341-22-1782	Smith	Meander	10 Mississippi
7	409-56-7008	Bennet	Abraham	6223 Bateman
8	427-17-2319	Dull	Ann	3410 Blonde S
9	472-27-2349	Gringlesby	Burt	PO Box 792
10	486-29-1786	Locksley	Charlene	18 Broadway /
11	527-72-3246	Greene	Morningstar	22 Graybar Hc
12	648-92-1872	Blotchet-Halls	Reginald	55 Hillsdale Bl
13	672-71-3249	Yokomoto	Akiko	3 Silver Ct.
14	712-45-1867	del Castillo	Innes	2286 Cram Pl.
15	722-51-5454	DeFrance	Michel	3 Balding Pl.
16	724-08-9931	Stringer	Dirk	5420 Telegrap
17	724-80-9391	MacFeather	Stearns	44 Upland Hts
18	756-20-7391	Kearson	Livia	5720 McAvley

Kunder fra nylig oppkjøpt 'Firma NN'

	id	lastname	firstname	address	city	state
1	807-91-6654	Panteley	Sylvia	1956 Arlington Pl.	Rockville	MD
2	527-72-3246	Greene	Morningstar	22 Graybar Hou...	Nashville	TN
3	722-51-5454	DeFrance	Michel	3 Balding Pl.	Gary	IN
4	712-45-1867	del Castillo	Innes	2286 Cram Pl. #.	Ann Arbor	MI
5	341-22-1782	Smith	Meander	10 Mississippi Dr.	Lawrence	KS
6	899-46-2035	Ringer	Anne	67 Seventh Av.	Salt Lake City	UT
7	998-72-3567	Ringer	Albert	67 Seventh Av.	Salt Lake City	UT
8	172-32-1176	White	Johnson	10932 Bigge Rd.	Menlo Park	CA
9	486-29-1786	Locksley	Charlene	18 Broadway Av.	San Francisco	CA
10	427-17-2319	Dull	Ann	3410 Blonde St.	Palo Alto	CA
11	846-92-7186	Hunter	Sheryl	3410 Blonde St.	Palo Alto	CA
12	672-71-3249	Yokomoto	Akiko	3 Silver Ct.	Walnut Creek	CA
13	724-08-9931	Stringer	Dirk	5420 Telegraph	Oakland	CA



Kunder Salesforce eller Siebel eller ...

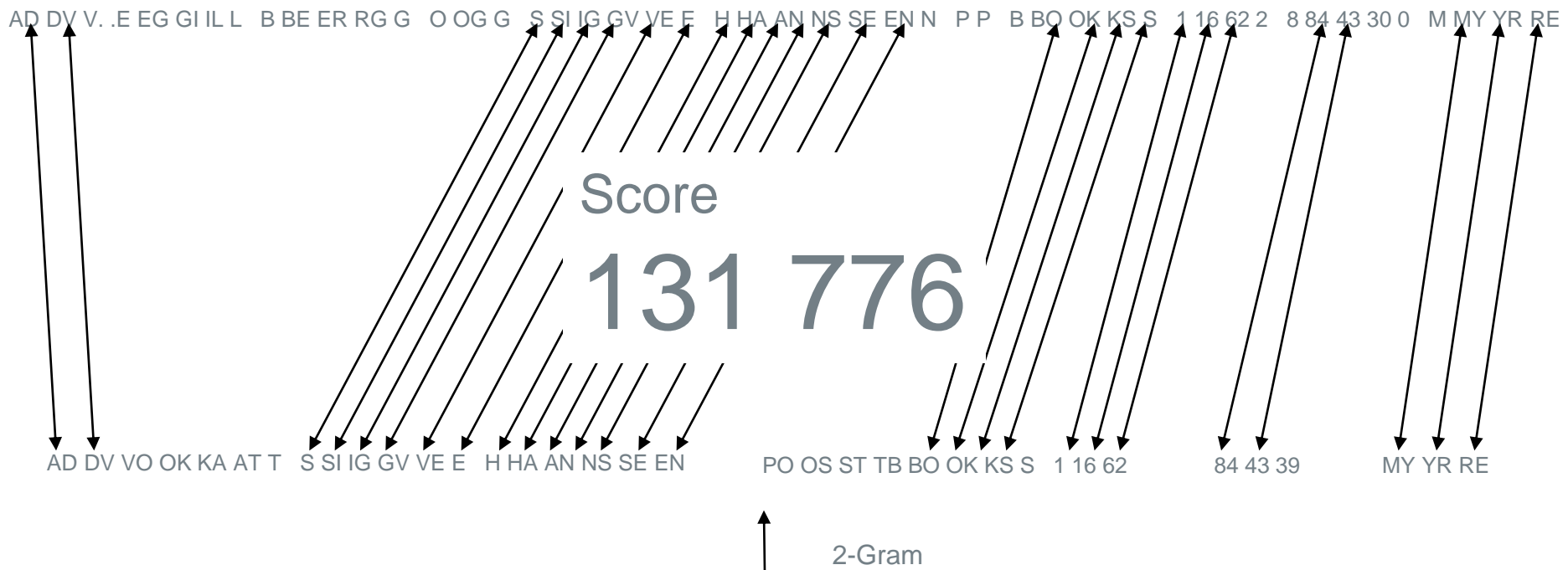
	id	lastname	firstname	address	city	state
1	807-91-6654	Panteley	Sylvia	1956 Arlington Pl.	Rockville	MD
2	527-72-3246	Greene	Morningstar	22 Graybar Hou...	Nashville	TN
3	722-51-5454	DeFrance	Michel	3 Balding Pl.	Gary	IN
4	712-45-1867	del Castillo	Innes	2286 Cram Pl. #.	Ann Arbor	MI
5	341-22-1782	Smith	Meander	10 Mississippi Dr.	Lawrence	KS
6	899-46-2035	Ringer	Anne	67 Seventh Av.	Salt Lake City	UT
7	998-72-3567	Ringer	Albert	67 Seventh Av.	Salt Lake City	UT
8	172-32-1176	White	Johnson	10932 Bigge Rd.	Menlo Park	CA
9	486-29-1786	Locksley	Charlene	18 Broadway Av.	San Francisco	CA
10	427-17-2319	Dull	Ann	3410 Blonde St.	Palo Alto	CA
11	846-92-7186	Hunter	Sheryl	3410 Blonde St.	Palo Alto	CA
12	672-71-3249	Yokomoto	Akiko	3 Silver Ct.	Walnut Creek	CA
13	724-08-9931	Stringer	Dirk	5420 Telegraph	Oakland	CA

PS. Dette er like gøy med feks produktkataloger, procurement systemer og HR-systemer

En av flere mulige løsninger

Kunde fra SAP @ HQ

"ADV. EGIL BERG OG SIGVE HANSEN, P BOKS 162, 8430 MYRE"



"ADVOKAT SIGVE HANSEN, POSTBOKS 162, 8439 MYRE"

Kunde fra nylig oppkjøpt 'Firma NN'

2.1 Definisjoner

- Den subjektive følelsen av 'god' kvalitet kommer dersom data-/informasjonsunderlaget er egnet til å støtte de beslutninger man står overfor
- Mer formelt må underlaget være;
 - **Korrekt.** Dette er jo selve basiskravet. Lettest å forstå, men ofte vanskeligst å oppnå. OG; det forutsetter at det finnes én 'korrekt' versjon av sannheten. I praksis er ulike definisjoner og benevnelser den største årsak til feil her.
 - **Komplett** Fordi viktige elementer som sammen skaper totalbildet gjerne registreres i ulike steg i forretningsprosessen, vil informasjonsbildet til tider være ukomplett. Dersom datavarehuset feks henter sitt underlag da, ...
 - **Tidsriktig** I vår arvede verden av systemporteføljer fra ulike tidsepoker, som i altfor liten grad er sanntidsintegreerte, er det ofte slik at datapunkter mangler eller ikke kommer frem før i morgen – etter at en batch (grøss) har kjørt.
 - **Konsistent** Betyr bare mangel på selvmotsigelser. Dette er ansett å være litt enklere innen IT enn i politikken. In-konsistens oppstår normal når noe går feil (nettet, strømmen, serveren) og dersom det ikke 'ryddes' kan vi drasse med feilen til evig tid.

10 regler. 1 til 5

- Analysér underliggende datakvalitet som et ledd i alle BI-relaterte prosjekter
 - Budsjettering, Konsolidering, Systemintegrasjon, Rapportering, KPI
- Aggregater er det letteste skjulestedet for tallfeil (pga manglende synlighet)
 - Sum(salg) pr ÅR, pr AVD, pr PRODUKT, pr PROD_GRUPPE
- Dubletter er de flestes største, dyreste og mest undervurderte problem
 - Sykdommen har kjent kur.
- Vi må også analysere relasjoner, ikke bare feltinnhold
 - Ordrelinjer uten ordrehode, Faktura til Kunde som ikke finnes
- NULL-verdier er fine for utviklere, og ofte helt ubrukelige for andre
 - Finn 'konsument-vennlige' måter å fremstille NULLS på

10 regler. 6 til 10

■ Murphys lov lever i beste velgående

- Hvis noe KAN gå feil, vil det gjøre det – på det verst tenkelige tidspunkt
- Hvis det IKKE gikk feil, vil det senere vise seg at det hadde vært best om det gjorde det

■ Teste, teste og teste.

- Feilsituasjonene må ofte spesialbehandles, test for avvik og kompletthet

■ Opprett kontrollrapporter når data skal fraktes mellom system

- Sum (saldo) Count(kunder) på begge sider for å sikre at overføringen er riktig

■ Involvér de som kjenner grunndata så tidlig som mulig

- Gi dem BO DI profiler-resultater eller liknende

■ Sats på interative forbedringsrunder, ikke 'big bang'

- Ingen har så langt klart å se alle problemer i starten, dette er en læringsprosess

"The scientific theory I like best is that the *rings of Saturn* are composed entirely of lost airline luggage."

- Mark

Russell.