

# DATAKVALITETSRAPPORT

powered by **GEOMATIC**



# INDHOLD

- Indledning
- Ledelsesresumé
- Datafangst
  - Om datafangst
  - Personnavne
  - Adresser
  - Telefonnumre
  - CPR-numre
  - Dubletter
- Datavedligeholdelse
  - Om datavedligeholdelse
  - Vedligeholdelse via officielle kilder
  - Personnavne
  - Adresser
  - Telefonnumre
  - Ændring af livsstatus
  - Forbehold mod kontakt
- Om Geomatic

# Indledning

I vores digitale verden anvender næsten alle organisationer data i deres forretning til analyser, marketing automation, effektivisering, automatisering eller til optimering af kundeoplevelsen; data er fuldstændigt integreret i den måde vi arbejder på, og de er kernen i alle disse aktiviteter. Men hvordan står det egentlig til med datakvaliteten?

Ted Friedman, Vice President hos Gartner udtrykker det således (2018): *“As organizations accelerate their digital business efforts, poor data quality is a major contributor to a crisis in information trust and business value, negatively impacting financial performance.”*

Selvom det er indlysende for de fleste, at datakvalitet betyder noget, er det fortsat en stor udfordring for mange organisationer at få kontrol over data, og det skyldes ofte at:

1. Det er vanskeligt at få et overblik over, hvor slemt det står til
2. Det er svært at beregne omkostningen ved dårlig datakvalitet
3. Det kan være vanskeligt at finde den rette datakvalitetsløsning

Med datakvalitetsrapporten giver vi dig en grundlæggende indsigt i dine egne data. Du får benchmarks, estimater på omkostningen og anbefalinger til tiltag og forbedringer.

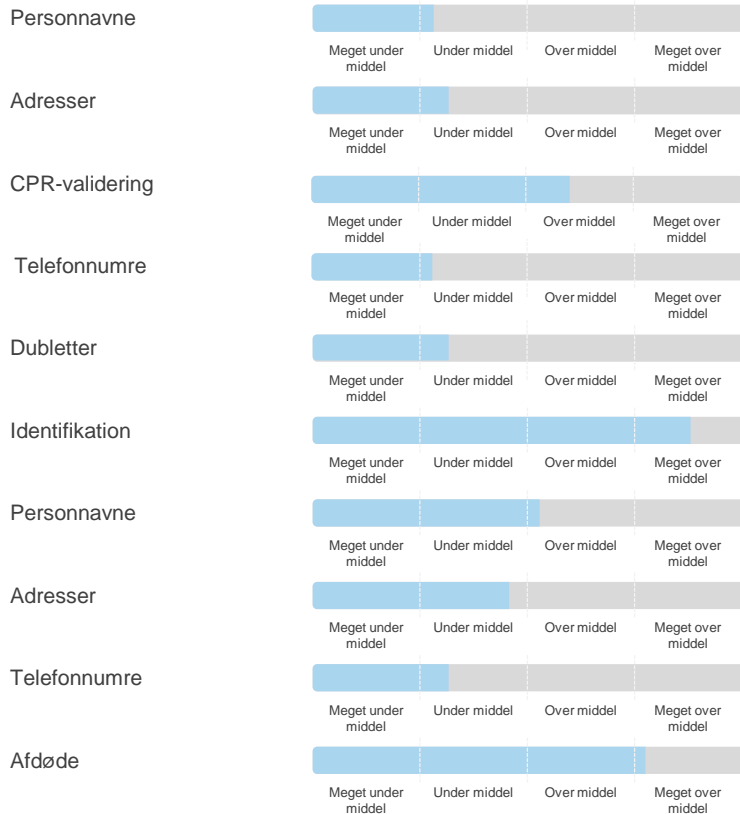
Vores fokus er på stamdata på privatpersoner; kunder og medlemmer, for hvad mange ofte glemmer, så er stamdata de eneste data, der kan give transaktionelle og ustrukturerede data kontekst.

Rapporten indeholder først et ledelsesresumé, som aggregerer de væsentligste fund og konklusioner i rapporten, herefter går vi i dybden med analyser af datakvaliteten under datafangsten og i datavedligeholdelsen, og til slut deler vi vores anbefaling til en langtidsholdbar datakvalitetsløsning.

God læselyst!

# LEDELSESRESUMÉ

# Overblik



Analysen viser, at der er fejl på hver 5. personnavn

12,7 % af adresserne kunne kun identificeres delvist eller slet ikke

Bemærk, at 66 personer ser ud til at have et CPR-nummer, der tilhører en anden person

Næsten hver 5. person har ikke et mobilnummer

Der er 1.244 dubletter. Mange vil sandsynligvis kræve manuel håndtering.

Efter bearbejdning og reduktion af fejl og mangler ligger identifikationsraten på 97 %

5,8 % af personnavnene afspejler ikke virkeligheden

6,7 % af adresserne er ikke korrekte pga. flytninger. Hyppigere opdatering vil eliminere dette

Næsten hvert 5. telefonnummer tilhører en anden person

Gode rutiner

# Estimat på ”prisen” for dårlig datakvalitet

## Omkostning ved dårlig datakvalitet i datafangsten

Antal 100 % korrekte rekords = 42.300

Datakvalitetsindeks = 42,3

### ”The Rule of 10”

Det koster 10 gange så meget at udføre en aktivitet, når datakvaliteten er dårlig i forhold til, hvis den var perfekt

Dine omkostninger ved 10.000 records:

$(4.230 * 1) + (5.770 * 1 * 10) = 61.930$  kr.

Dine omkostninger ved samme aktivitet, hvis alle data var perfekte:

$(10.000 * 1) = 10.000$  kr.



## 3 anbefalede hovedindsatsområder

1

### Udfordringer med datafangsten

Analysen viser, at XX er udfordret på kvaliteten af både, navne, adresser, telefonnumre og dubletter og analysen peger stærkt i retning af problemer opstået under datafangsten.

Sammenlignet med anbefalingerne ligger datakvaliteten **under middel**.

**Vi anbefaler** at stramme op på processerne i datafangsten samt en integration med API eller Retlify. Det vil sikre, at data oprettes korrekt fra start.

2

### 13,8 % af telefonnumrene ejes af en anden person

Analysen viser, at XX er udfordret på både indhentning og vedligeholdelse af telefonnumre.

Sammenlignet med anbefalingerne ligger datakvaliteten **under middel**.

**Vi anbefaler** en skræddersyet telefonvalideringsløsning, der fylder hullerne ud, sikrer den rigtige ejer og overvåger ændringer.

3

### Opdateringsfrekvensen

Analysen viser, at 6,7 % af kunderne er flyttet siden sidste opdatering af addressedata.

Sammenlignet med anbefalingerne får det adressekvaliteten til at ligge **under middel**.

**Vi anbefaler** en hyppigere opdateringsfrekvens.

# ANALYSE AF DATAFANGSTEN



# Fokus på datafangsten - hvorfor?

Datafangsten er den proces, hvormed nye data kommer ind i et system, en database eller i en organisation for allerførste gang, og et andet ord for de processer, I som organisation anvender, når I opretter nye kunder eller medlemmer.

I de fleste organisationer sker datafangsten på et eller flere af følgende steder:

1. I et kunde-, service- eller kontaktcenter, hvor mange medarbejdere betjener kunder eller medlemmer, og hvor datafangst er et af deres hovedarbejdsområder
2. I en butik eller på en anden lokation, hvor organisationen møder sine kunder eller medlemmer ansigt til ansigt, og hvor datafangst kun udgør en mindre del af medarbejdernes opgaver
3. Online, hvor nye medlemmer eller kunder ved egen hjælp giver organisationen adgang til deres data i en kontaktformular eller i fritekstfelter

Datafangsten er en kritisk proces, fordi det er her, fundamentet for jeres datakvalitet lægges. Det fundament kan være stærkt eller svagt afhængig af, hvilke tiltag I har taget for at sikre, at data kommer korrekt ind fra start. Det er under datafangsten, der skabes dubletter, fordi system eller medarbejder ikke kan identificere en kunde eller et medlem, der allerede er oprettet. Ofte er årsagen til fejl i data under oprettelsen. Undersøgelser viser, at mere end 75 % af alle fejl i data stammer fra datafangsten, og det er fejl, som flyder direkte ud i organisationen.

Datafangsten er en af de processer, mange organisationer nedprioriterer, selvom der er en stærk sammenhæng mellem kvaliteten af datafangsten, kvaliteten af data og organisationens evne til at skabe resultater og nå sine forretningsmæssige mål.

Derfor stiller vi skarpt på kvaliteten af jeres data under datafangsten her i første afsnit af rapporten.

# Datafangst: Personnavne (1 af 2)

I denne analyse kigger vi på struktur og indhold i navnedata, som vi har modtaget dem. Vi analyserer i hvor høj grad data er komplette, konsistente og korrekte på **8 kvalitetsparametre** og ser efter trends og karakteristika i datasættet. Analysen giver indsigt i kvaliteten af datafangstprocessen. Til analyserne anvendes egne værktøjer

## RESULTAT AF ANALYSEN

Udfyldte navnefeltene = **97,2 %**

**8,5 %** mangler enten for- eller efternavn

**2,2 %** indeholder dobbeltnavne

**1,3 %** indeholder initialer

**6,0 %** er ikke en person

**4,7 %** indeholder UPPER CASE

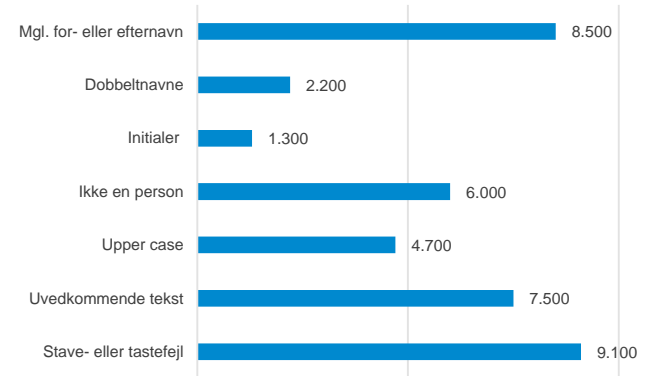
**7,2 %** indeholder uvedkommende tekst

**9,1 %** indeholder stave- eller tastefejl

## KONKLUSION

Analysen viser, at XX er udfordret på kvaliteten af navnedata og analysen peger stærkt i retning af problemer opstået under datafangsten. Sammenlignet med anbefalingerne ligger datakvaliteten under middel.

Resultat i antal



## SÆRLIGE BEMÆRKNINGER

Der er oprettet testcases og dødsbonavne i navnedata. Samlet set <150 stk.

# Datafangst: Personnavne (2 af 2)

Denne side indeholder konkrete eksempler på navne kvalitetsmæssige udfordringer i jeres datasæt

## Stave- og tastefejl

Fejl (fornavn)	Korrigeret (fornavn)	Fejl (efternavn)	Korrigeret (efternavn)
Willy	Villi	Brorson	Bresson
Thovald	Thorvald	Dybahl	Dubdahl
Christian	Kristian	Vildt	Vindt
Luis	Louis	Romdorf	Rosendof
Jedan	Ceren	Friss	Friis
Coony	Conny	Vistup	Vistrup
Elvin	Erwin	Akselsen	Axelsen
Annelise	Anna Lis		

## Uvedkommende tekst

Fornavn	Efternavn
Bente MORS	Sørensen
Ulla REG	Liljedahl
Hans (ubekendt på adressen)	Dahl
Jørgen	Nielsen test
Janus MORS	Uldahl
Sabine (ukendt på adressen)	Dahl
Ida	Kristensen MORS
Jonas	Jensen MORS

- MORS = 2.653
- Test = 2.202
- Ubekendt på adressen = 1.298
- REG = 204

## Ikke-genkendelige personnavne

- Villa Hansen
- Tryxx L. Toft
- Guldsmed Sørensen
- Testi Testesen
- Unknow Fogh
- Fælleshuset
- 111662 xx 5839

customer_id	output_name
603401324	Test Testensen
603360629	Test Thomsen
602537938	Test Anders

# Datafangst: Adresser (1 af 2)

I denne analyse kigger vi på struktur og indhold i adressedata, som vi har modtaget dem. Vi analyserer i hvor høj grad data er komplette, konsistente og korrekte på **6 kvalitetsparametre** og ser efter trends og karakteristika i datasættet. Analysen giver indsigt i kvaliteten af datafangstprocessen. Til analyserne anvendes egne værktøjer

## RESULTAT AF ANALYSEN

Kan adressen identificeres i adresseregistret

- Eksakt match = 87,31 %
- Delvist match = 8,53 %
- Ingen match = 4,16 %

**4,79 %** mangelfuldt vejnavn

**8,92 %** mangelfuldt husnr., bogstav eller etageinfo

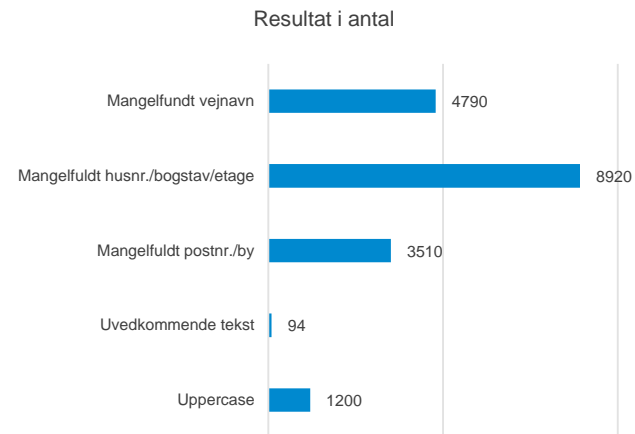
**3,51 %** mangelfuldt postnummer og/eller by

**<0,1 %** indeholder uvedkommende tekst

**1,2 %** indeholder UPPER CASE

## KONKLUSION

Analysen viser, at XX er udfordret på kvaliteten af adressedata og analysen peger stærkt i retning af problemer opstået under datafangsten. Sammenlignet med anbefalingerne ligger datakvaliteten under middel.



## SÆRLIGE BEMÆRKNINGER

Ingen

# Datafangst: Adresser (2 af 2)

Denne side indeholder konkrete eksempler på adressekvalitetsmæssige udfordringer i jeres eget datasæt

## Stave- og tastefejl

Adresser med stave- og slåfejl	Adresser korrigeret for fejl
Aleen 105	Alleen 105
Noder Fasanvej	Nordre Fasanvej
Halmond Alle	Helmond Allé
Kastrupælundgase 64	Kastruplunggade 64
Prinsse Maries Alle 32	Prinsesse Maries Allé 32
Kærehevevej 34	Kærehavevej 34
Sjælør Bulevard	Sjælør Boulevard
Peter Bellisvej 2	Peder Billesvej 2

## Eksempler på fejl og mangler

Input data			Korrekt data		
Herlevgårdsvej 27	2730	Herlev	Herlevgårdsvej 17 B, 2 th	2730	Herlev
Søndermarken 36	3050	Espergærde	Søndermarken 36, st tv	3050	Espergærde
arildvej 5	6900	Varde	Storegade 55 B, 2	6800	Varde
	5200	Bogense	Abildvej 5, st tv	5400	Bogense
	7330	Brande	P. Sabroes Allé 2, st	7330	Brande
Salvej 20 B	845	Hammel	Sallvej 20 D	8450	Hammel
Hyldegaards tværvvej		Charlottenlund	Hyldegårds Tværvvej 49, 6 th	2920	Charlottenlund

## Uvedkommende tekst

Top 3 uvedkommende tekst: 1) Ukendt = 3.498, 2) Test = 1.287, 3) HH = 994

- Har ringet til hende den 17/2 2015 Ønsker ikke HA (bruger dem ikke)
- From: 9/7/2017 2:30:00 PM, To: 9/7/2017 3:30:00 PM, Phonenummer: 29533138
- Ukendt Adr. - Brev kommet retur

# Datafangst: CPR-numre

I denne analyse kigger vi på CPR-numrene, som vi har modtaget dem. Vi validerer og analyserer på **6 kvalitetsparametre** og ser efter trends og karakteristika i datasættet. Analysen giver indsigt i kvaliteten af datafangstprocessen. Vi anvender egne værktøjer og CPR-registret

## RESULTAT AF ANALYSEN

Udfyldte CPR-felter = **88 %**

< **0,1 %** har et ugyldigt format

< **0,1 %** består ikke modulus-11\*

< **0,1 %** tilhører muligvis en anden

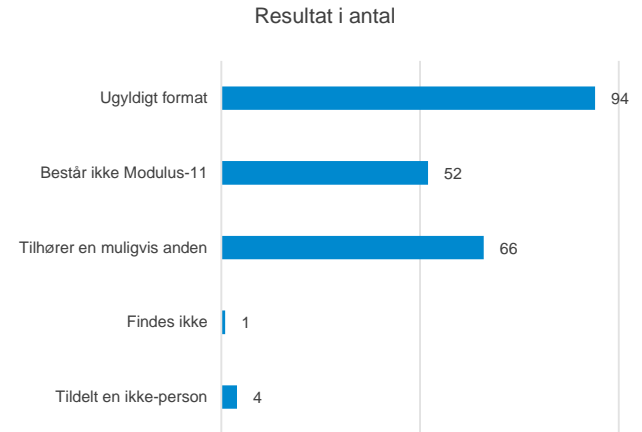
< **0,1 %** findes ikke i CPR-registret

< **0,1 %** er tildelt en ikke-person

*\*Modulus-11 er en metode til generering af et kontrolciffer, som anvendes til at kontrollere for tastefejl. Eksempler på ugyldigt format: Forkert antal cifre, ugyldig fødselsdato*

## KONKLUSION

Analysen viser, at XX har gode processer for datafangst på CPR-numre. Dog bemærkes det, at kun 88 % har et udfyldt CPR-nummerfelt, og at 66 numre ser ud til at tilhøre en anden person. Sammenlignet med anbefalingerne ligger datakvaliteten over middel.



## SÆRLIGE BEMÆRKNINGER

Ingen

# Datafangst: Telefonnumre

I denne analyse kigger vi på struktur og indhold i teledata, som vi har modtaget dem. Vi analyserer i hvor høj grad data er komplette, konsistente og korrekte på **8 kvalitetsparametre** og ser efter trends og karakteristika i datasættet. Analysen giver indsigt i kvaliteten af datafangstprocessen. Til analyserne anvendes egne værktøjer og det nationale telefonnummerregister

## RESULTAT AF ANALYSEN

Har mindst ét telefonnummer = **91,8 %**

Har 2+ telefonnumre = **28,2 %**

Har intet telefonnummer = **8,2 %**

**18,3 %** har intet mobilnr. (**18.300**)

**36,1 %** af numrene er identiske hos 2+ (**36.100**)

**1,0 %** af numrene registeret som mobilnr. er fastnetnr. (**1.000**)

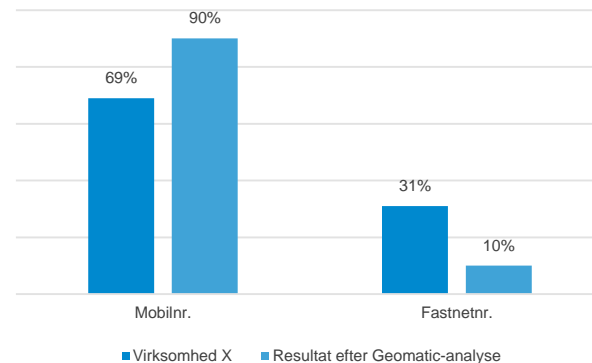
**0,2 %** af numrene har ikke et dansk format (**200**)

**<0,1 %** er dubletter (**37**)

## KONKLUSION

Analysen viser, at XX er udfordret på kvaliteten på telefonnumre, og analysen peger stærkt i retning af problemer opstået under datafangsten. Sammenlignet med anbefalingerne ligger datakvaliteten en del under middel.

Fordelingen mellem fastnet og mobil før og efter analysen af nummertype



## SÆRLIGE BEMÆRKNINGER

Da det nationale telefonnummerregister kun dækker 65-75 % af de danske nummerejere, kan tallene være højere. Det kan betyde, at den dårlige kvalitet og konsekvensen heraf er mere omfattende.

# Datafangst: Doubletletter

I denne analyse identificerer vi doubletletter på tværs af datasættet. Vi anvender vores egne værktøjer. I analysen anvender vi både historisk og aktuel navne- og adressedata fra CPR-registret på den enkelte person, som er helt essentielt for, at analysen lykkes

Der er **2.448 doubletletter** i datasættet:

- Antal tvillingepar = 1.224
- Antal trillingepar+ = 0

**1.224** records kan fjernes

## KONKLUSION

Analysen viser, at XX er udfordret på doubletbarrieren, og analysen peger stærkt i retning af problemer opstået under datafangsten. Sammenlignet med anbefalingerne ligger datakvaliteten under middel.

### i Vidste du, at...

doubletletter er blandt de største datamæssige udfordringer på tværs af brancher – den mest tidskrævende at løse på bagkant, men nem at undgå

Derfor er doubletletter så svære: Dataindholdet i doubletletter kan indtage mange former. De kan være næsten identiske, de kan være 100 % forskellige og alt midt imellem. Eksempel på en doublet:

- Ulla Borck, I.C Sørensen Vej 5, 2200 København N
- Ulla Sørensen, Borgmester Hansens Vej 92, 4300 Holbæk

Personen er flyttet mellem de to oprettelser og har opgivet mellemnavn i det første tilfælde og efternavn i det andet tilfælde



# ANALYSE AF DATAVEDLIGEHOEDEN



## Fokus på datavedligeholdelsen - i et lidt større perspektiv

Datavedligeholdelsen er den proces, hvor en organisation sikrer, at data afspejler virkeligheden så godt som muligt. I modsætning til datafangsten, er det en proces, som prioriteres af de fleste organisationer, men det er de færreste, der har en egentlig datakvalitetsstrategi.

Mange organisationer har en datavedligeholdelsesproces, hvor data opdateres med hjælp fra officielle kilder i forbindelse med større aktiviteter, fx en kampagne eller før udsendelsen af et medlemsmagasin hver 3. måned. Men data er i dag totalt integreret i den måde, vi arbejder på, og konsekvensen af denne tilgang til datakvalitet er, at alle aktiviteter, der anvender data imellem to opdateringer, benytter data, der forfalder og forældes. Det koster på top og bund med langt højere beløb, end de fleste organisationer er klar over.

Kvaliteten af datavedligeholdelsen hænger direkte sammen med kvaliteten af datafangsten. Er der ikke fokus på, at nye kunder og medlemmer oprettes med korrekte og komplette data, kan det umuliggøre en senere vedligeholdelse og opdatering via officielle kilder. Forudsætningen for at kunne modtage ændringer og opdateringer er nemlig, at kunden eller medlemmet kan identificeres først, og det stiller krav til data.

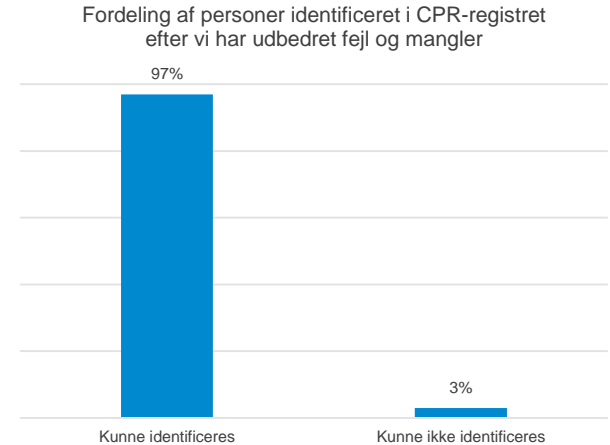
Forestil dig, hvilken effekt det kunne have på dine resultater i arbejdet med kundeoplevelsen, analyser, marketing automation, churn etc., hvis du kunne stole på data til alle aktiviteter, fordi de blev opdateret ugentligt eller månedligt.

# Er datakvaliteten god nok til at kunne opdateres via officielle kilder?

CPR-registret én af vores kilder til opdatering af data på privatpersoner. Datakvaliteten er afgørende for, om CPR-registrets systemer kan identificere og genkende en person, og dermed også en forudsætning for at kunne modtage ændringer og opdateringer. CPR-registret tillader ikke mange udsving i datakvaliteten, og derfor gennemgår og udbedrer vi fejl og mangler i alle datasæt, før vi sender dem til CPR-registret

## Primære årsager til manglende identifikation

- Markante fejl i vejnavn/husnummer
- Ikke muligt at placere lejligheder i etageejendomme



## KONKLUSION

Trods udfordringer i datafangsten var kvaliteten tilstrækkelig til, at vi kunne udbedre de fejl og mangler, der var. Sammenlignet med anbefalingerne ligger genkendelsesprocessen i den absolutte top.



Vi vurderer, at det er muligt at identificere og rette data på mindst 50 % af de ikke-genkendte via [Retrify](#). Herefter vil de kunne genkendes i CPR-registrets systemer

# Datavedligeholdelse: Personnavne

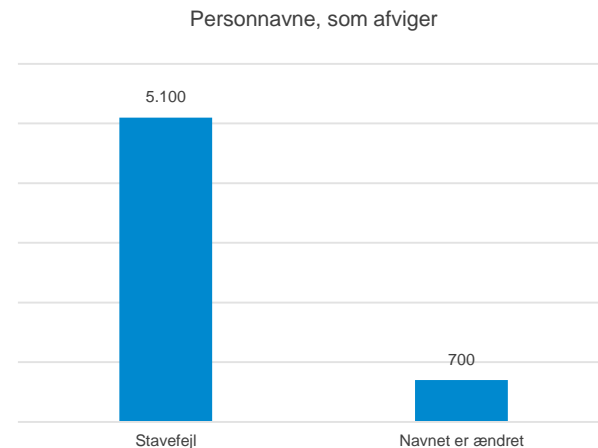
I denne analyse kigger vi på, om navnedata afspejler virkeligheden, og om data er korrekte. Har en person skiftet navn, vil det fremgå af denne analyse. Analysen giver indsigt i kvaliteten af jeres opdaterings- og vedligeholdelsesrutiner. Til analysen anvendes egne værktøjer og CPR-registret

**5.800** personnavnene var ikke korrekte, svarende til **5,8 %**.

Afvigelserne bestod i enten stavefejl eller i en reel ændring af fornavn og/eller efternavn.

## KONKLUSION

Data opdateres kun kvartalsvis og det afspejler sig i analysen. Sammenlignet med anbefalingerne ligger datakvaliteten under middel.



**i** Vidste du, at...  
der er ca. 50.000 voksne danskere, der ændrer navn hvert år

# Datavedligeholdelse: Adresser

I denne analyse kigger vi på, om adressedata afspejler virkeligheden, og om data er korrekte. Analysen giver indsigt i kvaliteten af jeres opdaterings- og vedligeholdelsesrutiner. Til analysen anvendes egne værktøjer i kombination med data fra CPR-registret

**6.706** adresser var ikke korrekte, svarende til **6,7 %**.

## Årsag til afvigelserne

- Personen er flyttet
- Adressen kunne ikke genkendes
- Der var mindre fejl i adressen

Af de 6.706 afvigelser er 5.867 (87,5 %) flytninger, hvoraf er 83 (1,4 %) sket for mere end et år siden.

## KONKLUSION

Data opdateres kun kvartalsvis og det afspejler sig i analysen. Vi anbefaler at afklare årsagen til de historiske flytninger, som ikke er blevet opfanget tidligere. Sammenlignet med anbefalingerne ligger datakvaliteten under middel.

Årstal	Antal historiske flytninger
2018	9
2017	10
2016	10
2015	6
2014	4
2013	9
2012	1
2011	5
2010	3
2009	2
2008	1
2006	3
2004	1
2003	1
2002	1
2001	3
1999	1
1995	1
1991	2
1990	1
1989	1
1983	2
1976	1
1975	1
1974	1
1971	1
1959	1

83

# Datavedligeholdelse: Telefonnumre

I denne analyse kigger vi på om telefonnumrene eksisterer, og om de afspejler virkeligheden, dvs. om personen er registreret som ejer af et nummer. Analysen giver indsigt i kvaliteten af jeres opdaterings- og vedligeholdelsesrutiner. Til analysen anvendes egne værktøjer i kombination med data fra det nationale telefonnummerregister, som dækker 65-75 % af danskerne

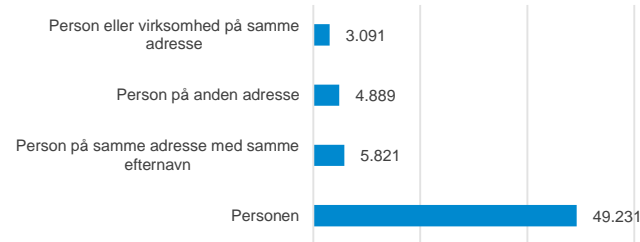
**13.801** telefonnumre tilhører en anden person end angivet i datasættet, svarende til **13,8 %**.

Telefonnummerregistret har oplysninger på **30,1 %** af de 18.300 personer, der manglede mobilnummer, svarende til, at **5.508** personer kan beriges med et mobilnummer.

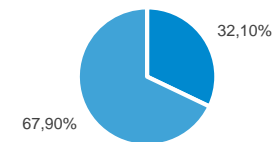
## KONKLUSION

Analysen viser et meget usædvanligt billede, hvor 13,8 % af telefonnumrene tilhører en anden person. Det indikerer dårlige opdaterings- og vedligeholdelsesrutiner, og det betyder, at næsten 1/5 af telefonnumrene ikke kan vedligeholdes via det officielle register. Sammenlignet med anbefalingerne ligger datakvaliteten under middel

Registreret ejer af telefonnummeret i antal personer



Andelen af numre, som var registreret i det nationale telefonnummerregister



• Nummeret findes i teleregistret • Findes ikke i teleregistret

# Datavedligeholdelse: Ændring af livsstatus

I denne analyse kigger vi efter personer i datasættet med en alvorligt ændret livsstatus, dvs. personer, der er forsvundet, udrejst eller afgået ved døden. Analysen giver indsigt i kvaliteten af jeres opdaterings- og vedligeholdelsesrutiner. Til analyserne anvendes egne værktøjer i kombination med data fra CPR-registret

Der er **149 personer** med en alvorlig ændret livsstatus, svarende til **1,5 %**.

Den ændrede status fordeler sig således:

- 32 personer er afgået ved døden
- 112 personer er udrejst
- 5 personer er forsvundet

## KONKLUSION

Analysen viser, at XX har gode opdaterings- og vedligeholdelsesrutiner omfattende ændringer i livsstatus. Sammenlignet med anbefalingerne ligger vedligeholdelsen over middel.



**Vidste du, at...**

der er ca. 55.000 danskere, der årligt afgår ved døden

# Datavedligeholdelse: Personer, der har taget forbehold mod kontakt

I denne analyse kigger vi efter personer i datasættet, der har taget forbehold mod kontakt, dvs. personer, der har registreret sig på Robinsonlisten, eller som er navne- og adressebeskyttet. Analysen kan give ny indsigt, men siger ikke noget om kvaliteten af jeres opdaterings- og vedligeholdelsesrutiner. Til analyserne anvendes egne værktøjer i kombination med data fra CPR-registret

## ROBINSONLISTEN

**21.736** er registreret på Robinsonlisten, svarende til, at **20,1 %** har takket nej til uanmodet markedsføring. Det svarer nogenlunde til den andel, der er i befolkningen.

De fleste har antageligt givet positivt tilsagn til jer, og I har derfor formodentlig mulighed for at henvende jer med markedsføringsbudskaber. Hvis der ikke i samme grad er givet positivt tilsagn til konkurrenterne, kan man forsigtigt antage, at I har en konkurrencemæssig fordel.

Registrering på Robinsonlisten ophæves, når en person selv afmelder sig.

## NAVNE- OG ADRESSEBESKYTTELSE

**1.148** har registreret sig som navne- og adressebeskyttet, svarende til 0,11 %. Det svarer også nogenlunde til den andel, der er i befolkningen.

Personer, der er navne- og adressebeskyttede kan ikke registreres på Robinsonlisten, og deres personlige data kan heller ikke opdateres fra officielle datakilder.

Denne beskyttelse skal fornyes hvert år.



# OM GEOMATIC

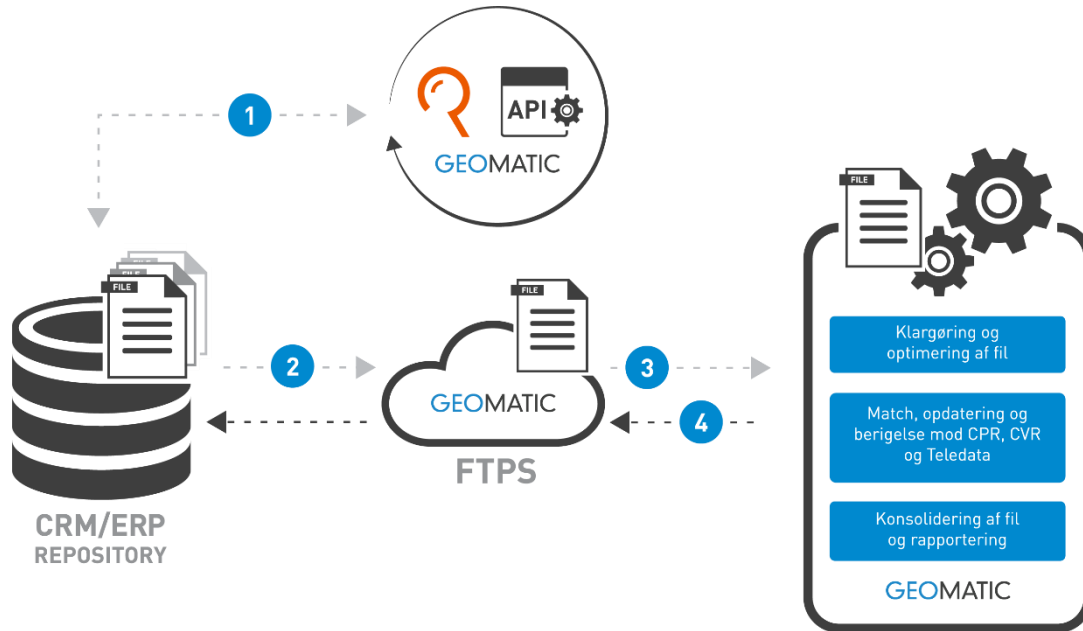
**Geomatic** er det data- og analysehus med speciale i data science og Master Data Management-løsninger, som en lang række af Nordens førende virksomheder foretrækker at samarbejde med, når de vil sætte deres egne og offentligt tilgængelige data bedre i spil på måder, som både betaler sig, og som samtidig giver dem større sikkerhed, nemmere nytteværdi og bedre beslutninger

Vi rådgiver om hele processen fra dataindsamling, validering, strukturering og berigelse til behandling, analyse, segmentering og operationalisering af data

Læs mere på [www.geomatic.dk](http://www.geomatic.dk)



# Vores model for en langtidsholdbar datakvalitetsløsning



## 1 Datafangst

- Verificering og oprettelse af nye kunder med støtte fra API eller applikationen Retrify
- Instant opdatering eller berigelse efter behov

## 2 Automatiseret data vedligeholdelse (INPUT)

- Automatiseret filudveksling til sikker FTP-server for opdatering af data i CPR, CVR og Teledata
- Filformat kan skræddersyes
- Mandatory dataelementer for sikker match

## 3 Automatiseret datavask

- Fil(er)ne behandles via en række automatiserede processer for maksimering af match hitraten. Dette inkluderer bl.a. fonetiske regler, alternative stavemåder, geokodning og korrigering af fejl i person- og vejnavne og CPR-numre.
- Match, opdatering og berigelse
- Konsolidering af fil(er), dubletmarkeringer og ændringsmarkeringer
- Kun ÆNDRINGER

## 4 Automatiseret data vedligeholdelse (OUTPUT)

- Data returneres til den sikre FTP server, hvorfra de hentes
- Der kan være behov for ressourcer ifm. med opdateringer

# Bliv klogere

Vi udgiver løbende analyser, guides og juridiske vejledninger om data, forretningsdrivende dataanalyse, datadrevet marketing og dataetik

Læs med på vores blog:

<https://geomatic.dk/dk/blogs/>

Se mere om vores løsninger:

<https://geomatic.dk/dk/solutions/>

