

Metodenotat – elmåler data

Baggrund og løsning

Danmarks Statistik (DST) har fået til opgave at etablere et datasæt over elforbruget i Danmark med afsæt i data fra Energinets DataHub. Data gøres tilgængeligt for forskere gennem DST's Forskningservice, hvor elforbruget vil kunne kobles til DST's øvrige registre på adresseniveau. Koblingen af elforbruget til DST's registre afhænger både af adresseangivelserne i Energinets DataHub og af de øvrige registres adresseangivelser.

I dette notat beskrives DST's fremgangsmetode i udarbejdelsen af datasættet. Desuden beskrives de modtagne data, mulige koblinger til de øvrige registre, afgrænsninger og ikke validerede.

Udover datasættet for elforbruget i Danmark udarbejder DST to tabeller med 1) valide adresseoplysninger, match-niveau og entydige virksomhedskoblinger og 2) flertydige virksomhedskoblinger. DST kan identificere og validere 98,4 % af adresserne i datasættet, som er et minimumskrav for videre kobling. Der matches entydigt 128.822 cvr numre til elmåler-adresser. Entydige koblinger betyder, at der kan laves en kobling mellem 1 elmåler og 1 virksomhed.

DST's opgave har alene været, at gøre data tilgængeligt i Forskningservice. For at øge muligheden for at koble til DST's øvrige registre har DST gennemført en adressevask for, at validere adresseangivelserne¹. De resterende stam- og aflæsningsdata er ikke fejlsøgt.

Notatet er gjort tilgængeligt for forskere og er godkendt af Energinet. Projektet er igangsat efter en stigende efterspørgsel på data fra DataHub. Bagerst i notatet findes en liste med relevante rapporter og links fra Energinet. Energinet påtager sig intet ansvar for eventuelle fejl i datagrundlaget, eller beslutninger og tolkninger foretaget på baggrund af data.

De modtagne data

DST har modtaget et total udtræk fra Energinets DataHub for 2013. DataHub blev etableret i marts 2013 og data er systematisk registreret i perioden fra marts-december 2013. I databasen er der betydelige mængder aflæsningsdata fra perioden før, da disse er blevet indlæst i DataHub efterfølgende.

Udtrækket er opbygget som en relations database, hvor relationen mellem tabellerne er elmåler-id. Relationen videre til DST's registre er baseret på elmåler-adresser. Således bør der i resten af notatet skelnes mellem elmåler-id og elmåler-adresser. En elmåler har et unikt elmåler-id, mens en elmåler-adresse kan have tilknyttet flere unikke elmåler-id. Der vil være 5 tabeller tilknyttet elforbruget.

- 1) Elmåler stamdata (meter_mp2), DataHub
- 2) Elforbrug, skabelonafregnede aflæsninger (meter_readings), DataHub
- 3) Elforbrug, timeafregnede aflæsninger (meter_values), DataHub
- 4) Valide elmåler-adresser, cvr nummer, arbejdssteds nummer og match-niveau, (meter_entydig) DST
- 5) Flertydige virksomhedskoblinger, (meter_flertydig) DST

¹ Behandlede data fra DST vil være tilgængelige i separate tabeller. Koblinger til virksomheder og BBR er baseret på særudtræk og match-procenter kan afvige fra koblinger i DST Forskningservice.

Elforbruget i DataHub indeholder både forbrug og produktion. Desuden er summen af det samlede elforbrug inklusiv nettab².

Elmåler stamdata (meter_mp2)

Tabellen *meter_mp2* indeholder stamdata om ca. 4 mio. elmålere i Danmark. Tabellen indeholder relevante oplysninger på elmålerniveau bl.a. elmåler-id, elmåler-adresse, gyldighedsdato, måler-type, afregningsform samt kunde- og ejer oplysninger. Tabellen indeholder 73 variable og beskrivelse heraf findes i tilhørende variabel beskrivelse. Det bør bemærkes, at stamdata på elmåler-niveau indeholder oplysninger fra før og efter etableringen af DataHub³.

Elforbrug/aflæsninger (meter_readings, meter_values)

Elmålingerne er delt op i to tabeller. *Meter_readings* indeholder ca. 10,2 mio. aflæsninger for ca. 3,2 mio. elmålere, der er skabelonafregnet eller manuelt aflæst. Tabellen dækker over måneds-, kvartals-, årsaflæsninger eller slutopgørelse. *Meter_values* indeholder ca. 633,9 mio. aflæsninger, der er timeafregnede. Tabellen indeholder ca. 58.000 elmålere med et forbrug på mindst 100.000 kWh/år. En måler kan indgå i begge tabeller ved at være timemålt, men periodeafregnet. Det bør bemærkes, at der findes 437 elmålere med registrerede afmålinger, som ikke findes i stamdata⁴.

Oversigt - DataHub tabeller

	Meter_mp2	Meter_readings	Meter_values
Records (mio.)	59,3	10,2	633,9
Elmåler-id	3.987.105	3.183.902	58.704
Antal variable	73	5	4

DST tabeller (meter_entydig, meter_flertydig)

DST har udarbejdet to tabeller, der indeholder valide adresseangivelser. Data har været igennem en adressevask, der indebærer at adresseangivelserne er blevet ændret til DST's adresseformat. Tabellerne indeholder ligeledes virksomhedsoplysninger på arbejdsstedsniveau baseret på de valide adresseangivelser. DST har koblet 83 % af de aktive virksomheder i 2013 til elmåler-data. Ca. halvdelen af virksomhedskoblingerne er entydige, hvilket betyder, at der kan laves en kobling mellem 1 elmåler og 1 virksomhed. De resterende virksomhedskoblinger er flertydige ex 1 elmåler kobles til 2 virksomheder. DST har oprettet en variabel *match*, der grupperer elmålerne på match-niveau⁵.

Adressevask

I DST Forskningservice kan elmåler-data kobles til DST's øvrige registre på adresseniveau. Match-procenter afhænger af adressekvaliteten i både elmåler-data og DST's øvrige registre. For at opnå de størst mulige match-procenter, valideres elmåler-adresserne fra DataHub til formatet der bruges i DST Forskningservice. Elmåler-adresserne sammenholdes med CPR's vejregister for, at indhente kommunekode og vejnr. Valideringen sikrer, at 98,4 % af elmåler-adresserne er på samme format som i de nuværende DST registre; kommunekode, vejnr., husnr., etage, sidedør.

² Nettabet er den mængde af energi der går tabt i forbindelse med transport af energien. Nettabet registreres på skabelonafregnede elmålere og kan dermed ramme specifikke områder.

³ Der er ca. 3,3 mio. elmålere i 2013. De ca. 4 mio. elmålere er inkl. inaktive og nye (efter 2013) elmålere.

⁴ Stamdata med valide adresser.

⁵ Se tabel DST gruppering af elmåler-adresser

Validering af stamdata

	Meter_mp2*	Valideret	Ikke valideret
Elmåler-adresser	4.028.343	3.977.554	50.789
Elmåler-id	3.861.278	3.841.540	19.738

Note: Stamdata findes i meter_mp2

* Blanke adresser og dubletter er fjernet⁶

Relevante elmålere

Elmåler stamdata indeholder oplysninger om elmålere frem til 1. april 2016. DST definerer relevante elmålere som de målere, der har haft aflæsninger i 2013. Afgrænses der til kun de relevante elmålere kan 99,5 % af elmålerne identificeres med valide adresser⁷.

Validering af aflæsningsdata

	Relevante elmålere 2013	Valideret	Ikke valideret
Elmåler-id	3.240.709	3.224.366	16.343
Andel (pct.)	100	99,5	0,5

Note: Aflæsningsdata findes i meter_readings og meter_values

De relevante og validerede elmålere står for 99,1 % af elforbruget i 2013.

Elforbrug 2013 fordelt på validerede elmålere

	Elforbrug 2013 (kWh)	Andel (pct.)
Valideret	31.143.640.583	99,1
Ikke valideret	286.430.228	0,9
I forbrug men ikke stamdata*	7.443.927	0,0
I alt	31.437.514.738	100

Note: Forbruget (type_of_mp=E17) fordelt på validerede og ikke-validerede adresseangivelser

* Svarer til forbruget for de 437 elmålere der ikke findes i stamdata

Match-niveau

DST gruppering af elmåler-adresser

Niveau ⁸	Måler adresser	Beskrivelse
90	437	I forbrug men ikke i stamdata med adresseangivelse
91	633.032	I stamdata men ikke i forbrug – valideret
92	5.504	I stamdata men ikke i forbrug – ikke valideret
93	45.285	I stamdata og forbrug – ikke valideret
1.1	2.855.218	I stamdata og forbrug – valideret (1 måler 1 adresse)
1.2	254.172	I stamdata og forbrug – valideret (x målere 1 adresse)
1.3	160.497	I stamdata og forbrug – valideret (1 måler x adresse)
1.4	74.635	I stamdata og forbrug – valideret (x målere og x adresse)
Total	4.028.780	Elmåler-adresser i alt

Note: "I forbrug" svarer til, at elmåler-id kan findes i meter_readings eller meter_values. "I stamdata" svarer til, at elmåler-id kan findes i meter_mp2.

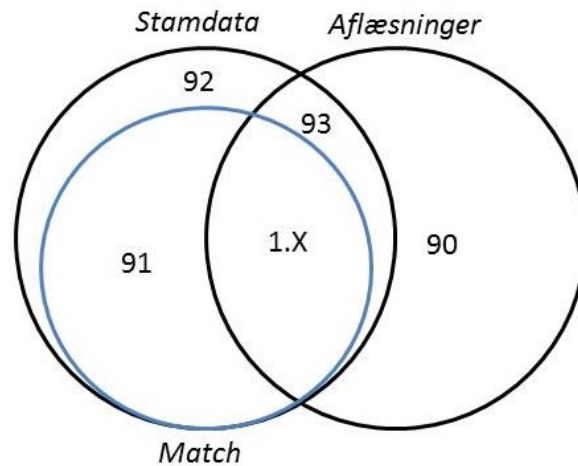
⁶ Elmålere med blanke adresseangivelser og dubletter fjernes inden adressevask. Efter adressevask opstår nye dubletter som ligeledes fjernes.

⁷ Elmålere kan være registreret "aktiv" selvom der ikke er aflæsninger på måleren, se variabel *Physical_Metering_Point (PhysicalStatusOfMeteringPoint)*

⁸ DST definition specifik for elmåler-data. Match niveau 90 + 93 svarer til de 16.343 unikke ikke validerede elmålere i tabellen **Validering af aflæsningsdata** når dubletter (elmåler-id) er fjernet. En stor del af elmålerne i gruppe 93 kan findes i grupperne 1.x med en valid adresse.

Der er gode match-muligheder på adresseniveau til DST's øvrige registre, da 99,5 % af de relevante elmålere har valide adresseangivelser. En valideret adresse der findes både i elmåler-data og i DST's register betyder imidlertid ikke, at koblingen er entydig. Kombinationen af elmåler-id og elmåler-adresse(r) kan skabe en match således, at et forbrug ikke entydigt kan kobles til en enhedsadresse⁹. Samtlige elmåler-adresser har fået et match-niveau jf. ovenstående tabel. Det bør bemærkes, at tabellen indeholder elmåler-adresser, hvor et elmåler-id kan optræde flere gange hvis den er registreret med flere adresser.

Venn-diagram



Note: *aflæsninger* omfatter det samlede elforbrug og *match* omfatter de målere, der har en valid adresseangivelse

Note: størrelsen på områderne afspejler ikke antallet af elmålere

Jf. Venn-diagrammet kan gruppe 1.x og 91 kobles til DST's øvrige registre. Delmængden 1.x kan deles op på følgende grupper: *gruppe 1.1* indeholder målere, der har 1 adresse og hvor adressen ikke har tilknyttet andre målere. *Gruppe 1.2* indeholder målere, der har 1 adresse og hvor adressen har tilknyttet andre målere. *Gruppe 1.3* indeholder målere, der er tilknyttet flere adresser og hvor adressen kan have ingen eller flere målere tilknyttet. *Gruppe 1.4* indeholder målere, der kan være tilknyttet flere adresser og hvor der kan være flere målere tilknyttet hver adresse.

Match og kobling til registre

Persondata

Videre match til registre bag personstatistikken tager udgangspunkt i de valide adresseangivelser. Jf. forrige afsnit er alle koblinger ikke nødvendigvis relevante.

Gruppe 1.1 indeholder elmålere, hvor 1 måler entydigt kan findes på 1 adresse. Dvs. forbruget på elmålere i gruppe 1.1 vil kunne kobles direkte til de personer der er registreret på adressen. Der findes dog adresser, hvor der ikke er registreret personer samt tilfælde hvor en adresse kan være flertydig. Ex kan en elmåler optræde med flere adresser med forskellige detaljeringsgrader. En elmåler kan optræde med en enheds- og en adgangsadresse. Kobling via adgangsadressen vil således være misvisende.

Koblinger til virksomheder¹⁰ (arbejdsstedsniveau)

DataHub indeholder virksomhedsnumre (cvr nummer på arbejdsstedsniveau), men både Energinet og DST vurderer, at kvaliteten og dækningen er lav. Der findes 219.173 cvr numre i stamdata. Afgrænses til 2013 findes der 126.886 cvr numre,

⁹ En adgangsadresse betegner indgangen til bygningen, hvor en enhedsadresse er døren til ex en lejlighed.

¹⁰ Virksomheder, juridiske personer og foreninger med cvr numre.

hvor 40.404 kan matches til DST's registre¹¹. Det svarer til 13,4 % af de aktive virksomheder i 2013. Grundet den lave dækning af virksomheder i elmåler-data tages udgangspunkt i DST's egen virksomhedskobling via adresseangivelserne.

I 2013 identificeres 298.557 aktive virksomheder i DST's registre. Af disse fremgår 298.001 med adresseangivelser i DST's register. For at sikre en entydig kobling mellem 1 måler og 1 virksomhed kobles kun på gruppe 1.1 med virksomheder med unik adresse (1 virksomhed per adresse).

Der kobles entydigt i alt 128.822 virksomheder (cvr numre) på elmåler-adresser svarende til 43 % af alle aktive virksomheder i 2013. Det er muligt at koble yderligere 118.371 virksomheder flertydigt til elmålere. Således kan samlet 247.193 af de aktive virksomheder i 2013 kobles til elmåler-data.

Virksomhedsmatch

	Antal virksomheder	Andel (pct.)
Entydig match	128.822	43
Flertydig match*	118.371	38
No match	50.808	17
Alle virksomheder*	298.001	100

*Der er i alt 138.665 virksomheder der kobles flertydigt. 20.294 af disse virksomheder har også en entydig kobling. Dette skyldes, at virksomheder kan være registreret på flere adresser

Flertydig og no match

En flertydig kobling betyder, at DST kan koble en elmåler og en adresse/virksomhed, men koblingen er *sløret*. I tabel *meter_flertydig* findes de flertydige kombinationer til virksomheder. Følgende er eksempler hvor en kobling kan være flertydig:

1 eller flere målere på 1 adresse (min. 2 virksomheder registreret på adressen)

- 1) Flere virksomheder registreret på samme adresse, som er i samme koncern
- 2) Flere virksomheder registreret på samme adresse, som ikke er i samme koncern
- 3) Flere virksomheder registreret på samme adresse i perioden pga. ejerskifte

1 måler på flere adresser

Der er som udgangspunkt 2 tilfælde, hvor en måler vil have flere adresser.

- 1) Den samme adresse fremgår på flere måder (adgangsadresse, enhedsadresse, stavfejl og lignende)
- 2) En måler har 2 forskellige valide adresser

Flere målere på 1 eller flere adresser

Flere målere registreret til 1 eller flere adresser, hvor minimum 1 måler er registreret til flere adresser.

- 1) 2 eller flere målere, som er registreret til de samme adresser
- 2) 2 eller flere målere, som er registreret til den samme adresse, men minimum 1 måler er registreret til mindst én yderlig adresse

Fordeling af arbejdssteder på kommuner

Fordeles virksomhederne på kommunekode er København den største kommune med 11,1 % af samtlige virksomhedsadresser i Danmark. I *no match* gruppen er

¹¹ Match kræver at adresse og cvr nummer er ens i elmåler-data og DST's register.

andelen af virksomheder i København 26,2 %. Der er således et relativt lavere match i København.

Virksomheder i Aarhus underrepræsenteres også, men ikke på niveau med København. Andelen i Aalborg og Odense er ens med niveauet for populationen.

Virksomheder fordelt på kommuner

	No match	Entydig match	Flertydig match	Population
København (101)	26,2	6,9	10,6	11,1
Aarhus (751)	6,1	4,6	5,7	5,2
Aalborg (851)	2,5	3,1	3,5	3,2
Odense (461)	2,4	2,7	3,0	2,8

Note: virksomheder er opgjort på arbejdsstedsniveau

Fordeling på brancher

Følgende tabel viser fordelingen af virksomheder på brancher. Sammenlignes de entydige match med fordelingen på populationen, overrepræsenteres bl.a. *landbrug, jagt, skovbrug og fiskeri* og *bygge og anlæg*. Virksomhederne underrepræsenteres derimod i brancherne *handel og transport, ejendomshandel og udlejning* og *andre serviceydelser*. For *no match* gruppen ses der at bl.a. brancherne *social- og sundhedsvæsen* og *videnservice* er overrepræsenteret.

Virksomheder fordelt på brancher

	No match	Entydig match	Flertydig match	Population
Videnservice	14,3	11,2	9,0	10,8
Social og sundhedsvæsen	9,8	6,1	5,6	6,5
Råstofindvinding	0,0	0,1	0,1	0,1
Rejsebureauer, rengøring og anden operationel service	5,6	6,3	4,6	5,5
Offentlig administration, forsvar, politi og undervisning	2,0	2,2	1,6	1,9
Landbrug, jagt, skovbrug og fiskeri	2,7	14,6	8,8	10,3
Kultur og fritid	2,6	2,0	2,3	2,2
Ingen branche	0,0	0,0	0,0	0,0
Information og kommunikation	7,2	5,2	4,3	5,2
Industri	3,9	4,9	5,7	5,0
Handel, transport, hoteller og restauranter	24,0	19,7	25,3	22,7
Finansiering og forsikringer	3,1	2,7	3,6	3,1
Energi- og vandforsyning, og renovation	0,7	1,4	1,9	1,5
Ejendomshandel og udlejning	9,3	5,4	13,4	9,2
Bygge og anlæg	7,4	13,6	7,9	10,3
Andre serviceydelser	7,3	4,7	5,9	5,6
Andre organisationer	0,0	0,0	0,0	0,0
I alt	100,0	100,0	100,0	100,0

Periodeafgrænsninger

Elforbruget er opgjort for regnskabsåret 2013. Etableringen i marts 2013 af DataHub garanterer kun data fra marts-december 2013, men grundet indlæsning af data fra januar-marts vil der være betydelig mængder af elmåler-data fra denne periode.

Eksempel på kobling

DST har testet kobling til BBR registret hvor 92,3 % af de validerede elmålere kan kobles til BBR¹².

Test-kobling til BBR

	Validerede	BBR match	No match
Elmåler-id	3.224.366	2.974.560	249.806
Andel (pct.)	100,0	92,3	7,7

Kontakt

Ønsker I mere information, har spørgsmål eller synspunkter til metodenotat eller data er I velkomne til at kontakte fuldmægtig Sebastian Villekjær, e-mail: skv@dst.dk, tlf. 39 17 30 33. Spørgsmål relateret til adgang og om Forskningservice henvises direkte til DST Forskningservice.

Litteratur

Relevant litteratur

- Markedsrapporter 2006-2011 ([link](#))
- DataHub Markedsrapport nr. 3 ([link](#))
- Detailmarkedsrapport nr. 2 ([link](#))
- Baggrundsdata til Detailmarkedsrapport ([link](#))
- EDI transaktioner for det danske el marked ([link](#))
- Yderlig information om DataHub Markedsrapport ([link](#))

¹² Test kobling lavet på et særudtræk af BBR registret og ikke direkte på versionen i Forskningservice. Afvigelser fra resultaterne i notatet for koblinger i Forskningservice kan forekomme.