



# Registerdataanalyse til State of the Nation 2024 Metodenotat



**TEKNOLOGISK  
INSTITUT**



**TEKNOLOGISK  
INSTITUT**

# Registerdataanalyse til State of the Nation 2024: Metodenotat

**Udarbejdet for:**  
ATV

**Udarbejdet af:**  
Teknologisk Institut  
Analyse og Erhvervsfremme

2024

Forfattere: Kristian Kriegbaum Jensen, Arne Hørlück Høeg, Stig Yding Sørensen  
Forside billedet AI-generet af MidJourney



## Indhold

1. Indledning .....	4
2. Identifikation af grundpopulationen af vidensintensive virksomheder .....	6
3. Indkredsning af S&E-virksomheder blandt vidensintensive virksomheder .....	8
Bilag 1: Afgræsningskriterier for S&E-virksomheder .....	11
Bilag 2: Projektindstilling til DST .....	13
Bilag 3: Kildekode til databehandling og beregninger .....	13



## 1. Indledning

Science & Engineering-virksomhederne (S&E-virksomhederne) er i årrække blevet beskrevet statistisk i ATV's 'State of the Nation'-rapporter. S&E-virksomhederne er defineret på baggrund af virksomhedsdata fra Danmarks Statistiks virksomhedsregister (FIRM) og registeret over forskning og udviklingsaktiviteter (FUI). Derudover anvendes persondata fra registeret over arbejdsstyrken (RAS) og uddannelsesregisteret (UDDA). Desuden indgår oplysninger fra Uddannelses- og Forskningsstyrelsen samt Innovationsfonden om virksomhedernes deltagelse i offentligt støttede projekter.

Teknologisk Institut har i Danmarks Statistiks registre indkredset populationen af S&E-virksomheder i Danmark til brug for 'State of the Nation'-rapporten 2024 (SN24). ATV definerer S&E-virksomheder som virksomheder, der baserer deres forretning på et stærkt vidensfundament inden for ingeniørvidenskab og STEM-kompetencer (Science, Technology, Engineering, Mathematics).

Formålet med dette metodenotat er at beskrive udvælgelsen af S&E-virksomheder til statistisk analyse i Danmarks Statistiks registre.

Definitionen der hidtil er anvendt, har ATV lavet i samarbejde med DAMVAD Analytics med inspiration fra OECD og UNESCO's definitioner af STEM-uddannelser, og metoden er overordnet beskrevet i de tidligere 'State of the Nation'-rapporter. S&E-virksomheder defineres, både nu og tidligere, ud fra tre spor:

- Virksomhedens vidensintensitet
- Virksomhedens andel af humankapital inden for S&E
- Virksomhedens investeringer i forskning og udvikling

FIRM anvendes til at identificere nøgletal for virksomheder omkring årsværk, omsætning, eksport, produktivitet, værditilvækst og geografisk spredning. FUI anvendes til at identificere virksomheders udgifter til forskning og udvikling. RAS anvendes til at identificere virksomhedernes medarbejdere og årsværk leveret af medarbejdere, mens UDDA anvendes til at identificere den enkelte medarbejders højst fuldførte uddannelse. På denne baggrund kan vi beregne, hvor stor en andel af en virksomheds årsværk, der leveres af S&E-medarbejdere. Data omkring offentlig støtte anvendes til at identificere om virksomhederne har deltaget i offentlige forsknings- og udviklingsprojekter.

Vedlagt dette metodenotat til ATV er indstillingen til Danmarks Statistik (DST), som lister de anvendte data samt kildekoden til databehandlingen og beregningerne. Kildekoden er skrevet i programmeringssproget R til brug for fremtidige analyser. Ved læsning af kildekoden kan man se alle trin taget i behandlingen af data og udregningen af nøgletal.

Metoden følger i hovedtræk principperne i tidligere års analyser, men der er også afvigelser, hvor den hidtidige metode ikke længere er tilstrækkelig eller tilstrækkeligt belyst i den foreliggende dokumentation fra tidligere analyser. Hensigten med dette dokument er at gøre rede for de valg, vi har truffet, så der opnås fuld transparens omkring udvælgelsen af S&E-virksomheder i fremtiden.



Herunder beskrives to trin i udvælgelsen af virksomheder:

1. Identifikation af grundpopulationen af vidensintensive virksomheder
2. Indkredsning af S&E-virksomheder blandt vidensintensive virksomheder

Samlet set er udvalget af S&E-virksomheder defineret, så det stemmer overens med UNESCOs nyeste definitioner af ingeniørvidenskab og STEM-kompetencer. Desuden indgår alle private virksomheder, med undtagelse af finansielle selskaber.



## 2. Identifikation af grundpopulationen af vidensintensive virksomheder

Beskrivelse af definition. State of the Nation 2022 (SN22)	I SN22 er metoden beskrevet således: "...grundpopulation på 9.463 <b>vidensintensive</b> virksomheder. Disse virksomheder er udvalgt, fordi de enten har deltaget i offentligt støttede forsknings- og udviklingsprojekter, optræder i FUI Danmarks Statistiks register for Forskning, Udvikling og Innovation i perioden 2012-2019 eller har branchekoden 'Forskning & Udvikling'."
Udfordringer	<p>Det er ikke veldefineret, hvordan virksomheder kategoriseres i denne sammenhæng - udover via branchekoden. Teknologisk Institut kan genfinde det samme antal virksomheder som i 2019 (9.463) ved at anvende følgende tilgang:</p> <ol style="list-style-type: none"><li>1. Virksomheden optræder som deltager i et af Innovationsfondens projekter 2015–2019. Her anvendes Forsknings- og Uddannelsesstyrelsens data fra 2012-2015 ikke. Desuden virker periodens længde vilkårlig. Hertil bevirker tilgangen, hvor man anvender projektdata fra alle tilgængelige år, upåagtet i hvilket år man ønsker at identificere grundpopulationen, at der mangler en konsistent tilgang til at identificere grundpopulation på tværs af årene. Der vil være færre datapunkter i tidligere år og flere i senere år.</li><li>2. <i>Eller</i> virksomheden har branchekoden 721000 (inkl. underkoder) 'Forskning og eksperimentel udvikling indenfor naturvidenskab og teknik' (jf. DB07_V3_2014 &amp; DB07_V2_2013)</li><li>3. <i>Eller</i> virksomheden optræder i FUI-registeret (2012-2021). Det er uklart, hvorvidt der er anvendt et kriterie, så der er en konsistent tilgang på tværs af alle år – fx at man ser på de seneste 10 år.</li><li>4. Og er en virksomhed med mindst én ansat (Årsværk &gt; 0; baseret på data fra variabelen GF_AARSV i registret FIRM).</li><li>5. Og virksomheden er en privat virksomhed af typen aktieselskab, anpartsselskab, fond og selvejende institution samt fond. Iværksætter-selskaber og andre former for privat virksomhedsorganisering er derimod ikke medregnet.</li></ol>
Metodiske valg	Teknologisk Institut har valgt følgende definition, som lader sig gentage:





	<ul style="list-style-type: none"><li>- Har deltaget i offentligt støttede forsknings- og udviklingsprojekter inden for de sidste 5 år (inkl. indeværende år). Det svarer til udvalget i 2019, men betyder også, at der vælges lidt for få virksomheder før 2016, da datakilden først blev tilgængelig i 2012.</li><li>- Eller har branchekoden 721000 (inkl. underkoder) 'Forskning og eksperimentel udvikling indenfor naturvidenskab og teknik' (jf. DB07_V3_2014 &amp; DB07_V2_2013)</li><li>- Eller optræder i FUI-registeret inden for de seneste 5 år (inkl. indeværende år). Lidt kortere periode, men konsistent med modtagelsen af offentlig støtte.</li><li>- Og er en virksomhed med mindst én ansat (Årsværk &gt; 0; baseret på data fra variabelen GF_AARSV i registret FIRM).</li><li>- Og er en privat ikke-finansiell virksomhed eller privat penge- eller realkreditinstitut (sektorkoder 18 og 28 på variabelen JUR_SEKTORKODE i registret FIRM).</li></ul>
Resultat	Antal 2016: 8699 Antal 2017: 8785 Antal 2018: 8953 Antal 2019: 9104 Antal 2020: 8960 Antal 2021: 8423
Vurdering	Det er en lille svaghed i metoden, at data for modtagelse af offentlig støtte ikke kan spores længere tilbage end 2012. Data findes ikke ifølge Forsknings- og Uddannelsesstyrelsen. Det betyder, at udvalget fra 2015 og tidligere vil være lidt underrepræsenteret.



### 3. Indkredsning af S&E-virksomheder blandt vidensintensive virksomheder

#### Beskrivelse af definition SN22

S&E-virksomheder er virksomheder med en humankapital inden for S&E og evt. investeringer i forskning og udvikling. Humankapitalen består af diplom- eller civilingeniører samt øvrige personer med længerevarende uddannelse inden for STEM-områderne.

I metodenotatet til SN22 defineres afgrænsningskriterierne for henholdsvis store (over 10 årsværk) og små SE-virksomheder (10 eller færre årsværk). Vi anvender samme afgrænsningskriterier (se bilag 1). Nettopopulationen af S&E-virksomheder identificeres ved brug af minimumskriterierne. I det følgende diskuteres operationaliseringen af kriterierne omkring identifikation af FoU-udgifter og identifikation af personer med ingeniøruddannelse eller anden STEM-uddannelse. Det er varianter af disse to typer kriterier, som inddeler S&E-virksomhederne i henholdsvis forskningsintensive, vidensstærke og driftsorienterede (se bilag 1).

#### Udfordringer og metodiske valg

Afgrænsninger og metodiske valg er ikke tydeligt defineret i metodebeskrivelsen, og det har ikke været muligt for Teknologisk Institut at genfinde samme population. Der er flere udfordringer:

1. Virksomhedernes investeringer i forskning og udvikling er ikke veldefineret. Teknologisk Institut forstår kriteriet således, at der kun skal være registreret et beløb over 0 kr. til FoU-udgifter, hvis virksomheden optræder i FUI-registret. Hvis virksomheden ikke optræder i FUI-registret, er dette ikke grundlag for at frasortere virksomheden. Det er uklart, hvilken variabel fra FUI, der er anvendt til at vurdere, om en virksomhed har FoU-udgifter. Variablen FUI-UDGIFT (samlede innovationsudgifter, ekskl. købte FoU-tjenester) dækker kun årene 2008, 2009 og 2018-2021. Det er dog ved brug af denne variabel, at vi kommer nærmest udvalget fra SN2022. Variablerne K\_TOTAL (købte FoU-tjenester) og U\_TOTAL (udgifter til egen FoU) dækker alle år fra 2007 og frem. Teknologisk Institut anvender summen af K\_TOTAL og U\_TOTAL. Endelig er det uklart, om en virksomhed skal have haft udgifter inden for en bestemt årrække eller ej. Her er det nødvendigt at sætte en meningsfuld grænse, fx inden for de seneste fem år.
2. Hvad angår identifikation af uddannelser, som er ingeniøruddannelser eller øvrige uddannelser inden for STEM, så beskriver metodenotatet til SN22, at udvalget er lavet med inspiration fra OECD's og UNESCO's definitioner af S&E-medarbejdere. Hvordan metoden præcist er "*inspireret*", er ikke dokumenteret. OECD har i deres





definition, som gengives i rapporten, anvendt den internationale standardklassifikation ISCED-97. ISCED klassificerer uddannelser i forhold til niveau (level) og felt (field), hvor OECD angiver felt 4 ('Science, comprising the life sciences, physical sciences, mathematics and statistics and computing') og 5 ('Engineering, manufacturing and construction) som STEM-områder. Vedligeholdelsen af ISCED-97 sluttede dog i 2011. ISCED vedligeholdes af UNESCO, som i 2011 udgav en revideret og opdateret udgave ISCED-11. Den nyeste udgave anvendes af Danmarks Statistik til at klassificere danske uddannelser. *Uddannelser, som er kommet til fra 2011 og frem, er kun klassificeret ud fra ISCED-11.* Uddannelser fra før 2011 er også klassificeret ud fra ISCED-11. ISCED-11 er blandt andet mere udbygget, hvad angår ICT-uddannelser. DST leverer ikke direkte uddannelsesdata med ISCED-koder. Disse skal man som bruger selv berige data med på forskermaskinen, og her skal man vælge mellem at anvende ISCED-97 eller ISCED-11. Det nærmeste vi kommer udvalget af S&E-virksomheder fra SN22, er ved at anvende ISCED-97 (vi rammer 2.347 virksomheder mod 2.372 i SN22). Vi har i øvrigt eksperimenteret med forskellige kombinationer, herunder manuel udvælgelse af uddannelserne (fremfor de standardiserede koder). Anvendes ISCED-97 er der dog en række ingeniør- og STEM-uddannelser, som udelades, da de er etableret fra 2011 og frem, og dermed ikke er klassificeret på ISCED-97.

3. Uddannelsesregisteret blev gennemgribende opdateret d. 19. april 2021, og det er uklart, om data i tidligere rapporter er lavet før eller efter denne opdatering.
4. Metodebeskrivelsen er uklar med hensyn til uddannelsestidspunkt. Regnes en ingeniøruddannelse for eksempel med, selvom personen senere har færdiggjort en ph.d. inden for fx humaniora? Og hvor langt tilbage i tid regnes der i så fald? Teknologisk Institut anvender højst fuldførte uddannelse i det pågældende år.
5. Det er uklart, hvorvidt enkeltmandsvirksomheder er frasorteret. Det er dog kun ved at frasortere disse, at vi genfinder de samme mønstre i alle tal som i SN22. Vi foreslår derfor, at disse frasorteres. Argumentet er, at enkeltmandsvirksomheder ikke kan forventes at have kapacitet eller infrastruktur til reelt at være vidensintensive. Enkeltmandsvirksomheder frasorteres også grupperne 'Industrien' og 'Danmark' i analysen, så tallene er sammenlignelige på tværs af grupperne.



<b>Metodiske valg</b>	<p>Vi foreslår følgende tilgang som en gentagelig, transparent og stringent operationalisering af kriterierne (bilag 1):</p> <ul style="list-style-type: none"><li>- Virksomheder har investeringer i forskning og udvikling, hvis deres interne FoU-udgifter (U_TOTAL i FUI) eller køb af eksterne FoU-ydelser (K_TOTAL i FUI) inden for de seneste fem år (inkl. året) tilsammen er over 0.</li><li>- Ingeniører defineres som personer, hvis højst fuldførte uddannelse er klassificeret inden for ISCED-11 felt 7: 'Engineering, manufacturing, and construction' og niveauerne 'Bachelor or equivalent' (diplomingeniør) eller 'Master or equivalent'/'Doctor or equivalent' (civilingeniør).</li><li>- Øvrige STEM med LVU defineres som uddannelser på kandidat- eller ph.d.-niveau ('Master or equivalent'/'Doctor or equivalent' ) inden for ISCED-11 felt 5: 'Natural sciences, mathematics and statistics' eller felt 6: 'Information and communication technologies'.</li><li>- Virksomheder med et eller færre årsværk frasorteres (variablen GF_AARSV i registret FIRM anvendes)</li></ul>
<b>Resultat</b>	<p>Antal 2016: 2454 Antal 2017: 2466 Antal 2018: 2537 Antal 2019: 2592 Antal 2020: 2595 Antal 2021: 2616</p>
<b>Vurdering</b>	<p>Teknologisk Institut vurderer, at ovenstående tilgang er præcis og transparent, og lader sig anvende konsistent på tværs af år. En mulighed for at øge præcisionen er en manuel gennemgang af alle uddannelser, der er klassificeret inden for de tre felter: 'Engineering, manufacturing and construction' (7), 'Natural sciences, mathematics and statistics' (5) samt 'Information and communication technologies' (6). Enkelte ingeniøruddannelser er blevet placeret under felt 5 og 6, mens enkelte andre STEM-uddannelser er klassificeret under felt 4. Det vil øge præcisionen, men med al sandsynlighed ikke ændre analysens overordnede resultater, da det ikke vil ændre det samlede antal S&amp;E-medarbejdere.</p>



## Bilag 1: Afgrænsningskriterier for S&E-virksomheder

Afgrænsningskriterier til identifikation af tre typer S&E-virksomheder med over 10 ansatte

Afgrænsningskriterier	Forskningsintensive S&E-virksomheder	Vidensstærke S&E-virksomheder	Driftsorienterede S&E-virksomheder
Andel ingeniører eller STEM-medarbejdere	20 % ingeniører eller 20 % STEM m. LVU	15 % ingeniører eller 15 % STEM m. LVU	10 % ingeniører eller 10 % STEM m. LVU
	eller	eller	eller
	samlet 35 % ingeniører og STEM m. LVU	samlet 25 % ingeniører og STEM m. LVU	samlet 15 % ingeniører og STEM m. LVU
og			eller
Vidensniveau	Udgifter til FoU såfremt de optræder i FUI-registeret	Udgifter til FoU såfremt de optræder i FUI-registeret	Udgifter til FoU såfremt de optræder i FUI-registeret
		og	og
		beskæftiger en gruppe ingeniører og STEM m. LVU på mindst 10 personer	beskæftiger en gruppe ingeniører og STEM m. LVU på mindst 5 personer
og			
Ph.d.	1 % ph.d.er såfremt de beskæftiger minimum 100 årsværk		



Afgrænsningskriterier til identifikation af tre typer S&E-virksomheder med 10 eller færre ansatte

Afgrænsningskriterier	Forskningsintensive S&E-virksomheder	Vidensstærke S&E-virksomheder	Driftsorienterede S&E-virksomheder
Andel ingeniører eller STEM-medarbejdere	40 % ingeniører eller 40 % STEM m. LVU	30 % ingeniører eller 30 % STEM m. LVU	20 % ingeniører eller 20 % STEM m. LVU
	eller	eller	eller
	samlet 50 % ingeniører og STEM m. LVU	samlet 40 % ingeniører og STEM m. LVU	samlet 30 % ingeniører og STEM m. LVU
	og	eller	eller
Vidensniveau	Udgifter til FoU såfremt de optræder i FUI-registeret eller har branchekoden 'Forskning og udvikling'	Udgifter til FoU såfremt de optræder i FUI-registeret og beskæftiger én ingeniør eller STEM m. LVU	



**TEKNOLOGISK  
INSTITUT**

## Bilag 2: Projektindstilling til DST

Leveret separat til ATV

## Bilag 3: Kildekode til databehandling og beregninger

Leveret separat til ATV



**TEKNOLOGISK**  
**INSTITUT**