

Årsrapport 2019

Indhold

Status for Danmarks Statistiks Forskningsservice 2019	3
Indledning.....	3
Vore kunder – de autoriserede institutioner	3
Projekter	3
Aktive forskere	4
Redegørelse for efterspørgsel på data.....	5
Leverede datasæt	5
Projektdatabaser	5
Brud på datasikkerhed eller -fortrolighed	5
Årets leveringstider	7
Igangværende projekter med en leveringstid på over to måneder.....	7
Status for Højkvalitetsdokumentation.....	8
Nye registre, forbedringer og større rettelser vedrørende data ..	9
Nyt register om akkumulerede indkomster (INDAKK)	9
Nyt register om Børnefamilieydelse og –tilskud (BOFAMYD).....	10
Nyt register om Boligstøtte (BOST).....	10
Nyt register om Personer i bestyrelse (PERSBEST)	10
Nye registre vedr. kontospecifikke data fra Skatteforvaltningen (IRTE, URTE, PANT)	11
Ny proces for opdatering af den Registerbaserede arbejdsstyrkestatistik (RAS) samt rettelser	11
Ny version tilgængelig af befolkningsregistret (BEF).....	11
Bygge- og boligregisteret samt Boligtællingen tilgængelig i en ny version (BBR, BOL)	12
Ny version af Landspatientregistret (LPR2).....	12
Forbedring af det Medicinske Fødselsregister (MFR).....	13
Rettelse i og forbedring af IDA arbejdssteder (IDAS)	13
Forbedret opgørelse af Iværksætterdatabasen (IVDB)	13
Rettelser i Børn og Unge registrene (BUAF, BUAS, BUAH)	13
Forbedring af Industriens salg af varer (VARSA)	13
Forbedret version af nøglen mellem IDA og Firmastatistik (FIDA)	13
Rettelse i Bystørrelser på arbejdssteder (BYSTRA).....	14
Rettelse i det Komprimerede elevregister (KOTRE)	14
Forbedring af og rettelse i Samlet opgørelse af formue og gæld (FORMGELD).....	14
Rettelser i Børn og unge – Underretninger (BUU)	14
Forbedring af og rettelser i opgørelsen af Boformer (BOFORM_F)	14
Forbedring af Kursistregistret – kursister ved voksen- og efteruddannelse (VEUV).....	14

Forbedring af Takster for Somatik – Landspatientregistret (DRGSOMA) ..	15
Rettelse i §48E Indkomster (INUF)	15
Rettelse i og forbedring af Indkomstregistret (IND), Familieindkomst (FAIK) og Arbejdsklassifikationsmodulet (AKM)	15
Forbedringer og udviklingsaktiviteter 2019	16
Projekt om Big-data	16
Opdateret vejledning i hjemsendelse af resultater	16
Fortsat udvikling af hjemsendelsessystem	17
Bestillings- og Administrationssystem, GDPR og standardprogrammer	17
Udviklingsarbejdet med FSE-BOA fortsatte i 2019.	17
Persondataforordningen - GDPR.	18
Standardprogrammer i FSE.....	18
Økonomiske nøgletal/regnskab	18
Regnskab 2019.....	19

Status for Danmarks Statistiks Forskningservice 2019

Indledning

Forskningservice var ligesom i de foregående år også i 2019 præget af stor en aktivitet og interessen for at arbejde med Danmarks Statistiks unikke registre fortsætter med at stige.

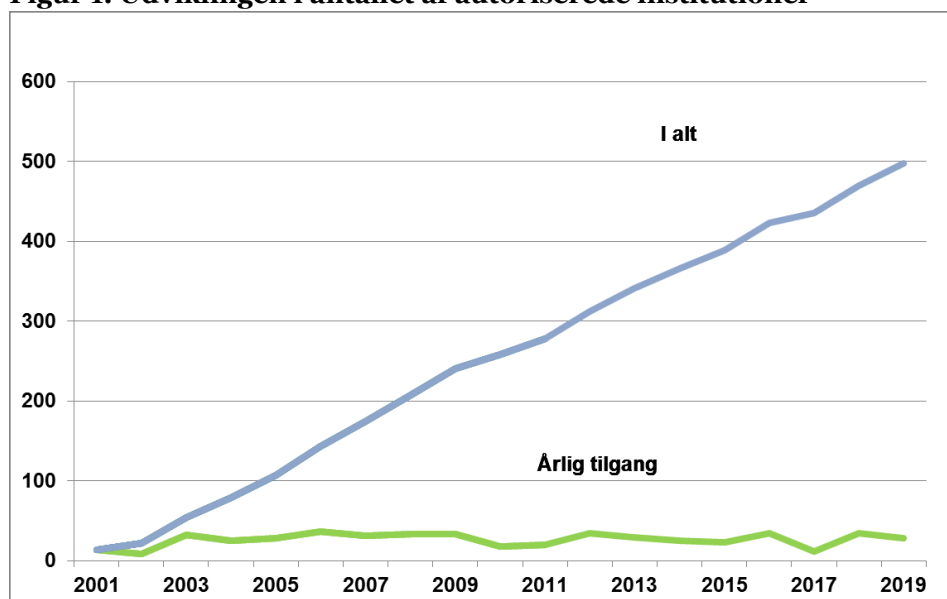
I årets rapport beskrives aktiviteten i Forskningservice nærmere, herunder væksten i vore kunder, projekter og forskere, udviklingen i datasikkerheden, projekternes leveringstider og det store vedligeholdelsesarbejde der er, for fortsat at sikre de bedst mulige grunddata til forskernes projekter. Herudover nævnes nogle af de udviklingsprojekter der fremadrettet skal sikre en fortsat effektiv forskningsservice samt fokus på GDPR.

Vore kunder – de autoriserede institutioner

Ved udgangen af 2019 havde Forskningservice **497 autoriserede institutioner**, hvoraf 28 nye institutioner blev autoriseret i 2019. Tilgangen af nye autoriserede institutioner lå i 2019 på et normalt niveau sammenlignet med tidligere år, jf. figur 1 nedenfor.

Størstedelen af institutionerne kommer fortsat fra offentlige miljøer. Andelen af autoriserede institutioner fordelte sig på godt 80 pct. offentlige institutioner, som er de institutioner, der får andel i bevillingen fra KOR, samt de resterende lidt under 20 pct. på private institutioner.

Figur 1. Udviklingen i antallet af autoriserede institutioner



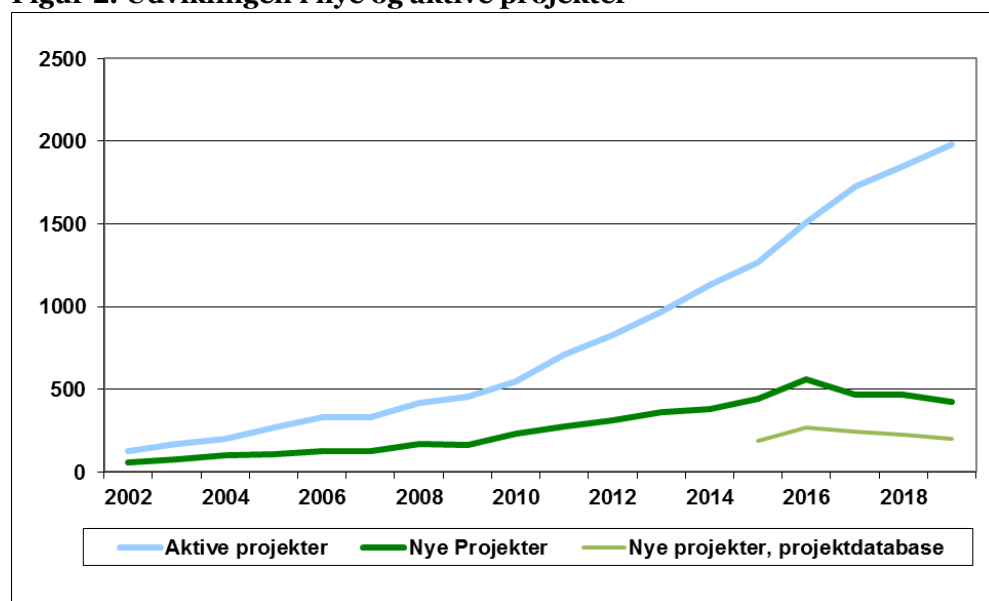
Projekter

Ved udgangen af 2019 arbejdede forskerne på over **1.980 aktive projekter**, dvs. projekter, som forskerne har været logget på i løbet af året. Disse projekter er dels placeret på Danmarks Statistiks forskningsservere og dels på forskernes egne servere - de såkaldte "hostede" servere. Det store antal aktive projekter afspejler et fortsat højt aktivitetsniveau, som Forskningservice har i forbindelse med levering af data til nye projekter, opdatering og udvidelse af data på eksisterende projekter samt overførsler af forskernes egne data.

Der blev i 2019 oprettet **423 nye projekter**, heraf 199 nye projekter fra de såkaldte projektdatabaser. Tilgangen af nye projekter er faldet i 2019 og fortsætter den svagt faldende tilgang af nye projekter, der har været siden 2016. Projektdatabaser giver større miljøer, via deres egen databaseadministrator, muligheden for selv at danne data til institutionens projekter, efter de er blevet godkendt af Danmarks Statistik. Lidt under halvdelen af de nye projekter kommer fra projekter dannet af en af miljøernes egne projektdatabaser i 2019.

Figur 2 nedenfor viser den årlige tilgang af nye projekter samt udviklingen i aktive projekter, herunder for de seneste år tilgangen af nye projekter fra projektdatabaser. I Appendiks A ses en liste med de projekter, hvor der har været aktivitet på i løbet af 2019.

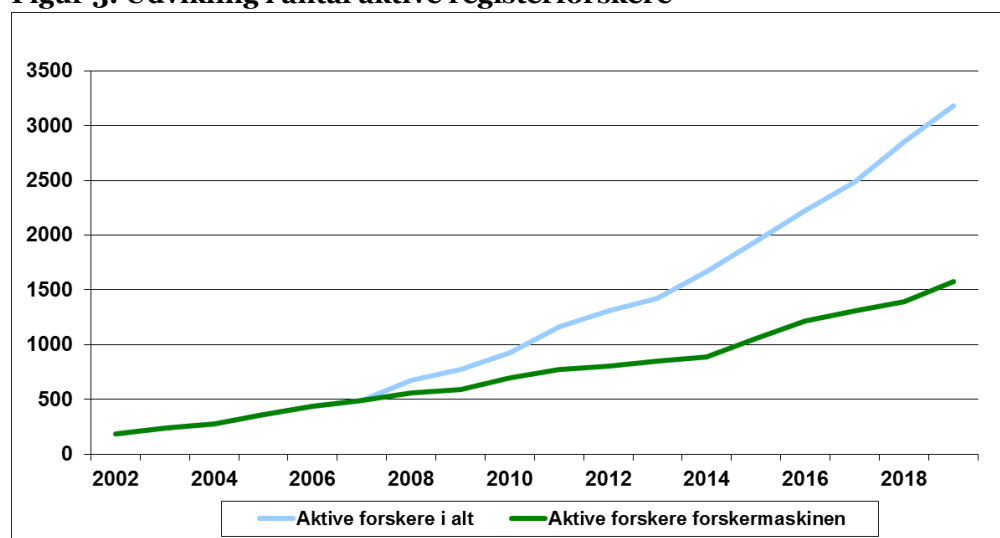
Figur 2. Udviklingen i nye og aktive projekter



Aktive forskere

I alt har ca. **3200 forskere i 2019 aktivt arbejdet på et eller flere projekter** på forskermaskinerne. Cirka halvdelen af forskerne arbejder på de hostede maskiner. Figur 3 nedenfor viser udviklingen i aktive forskere fra 2002 og frem.

Figur 3. Udvikling i antal aktive registerforskere



Note: Arealet mellem aktive forskere i alt og aktive forskere på forskermaskinen viser udviklingen i forskere, der arbejdede med data på hostede maskiner.

Redegørelse for efterspørgsel på data

Leverede datasæt

Ultimo året 2019 rådede Forskningsservice' grunddatabank over 335 registre.

Der blev i alt leveret 159.540 filer, heraf blev 104.340 leveret som SAS-datasæt. De resterende blev først og fremmest leveret som STATA-datasæt eller SAS-view.

Nedenfor ses top 10 for leverede grunddatasæt efter register i 2019:

1	BEF	Befolkningen
2	IND	Indkomst
3	UDDA	Uddannelser (BUE)
4	LPR	Landspatientregistret
5	AKM	Arbejdsklassifikationsmodulet
6	IDAP	IDA-personer
7	LMDB	Lægemiddelstatistikregisteret
8	IDAN	IDA-ansættelser
9	RAS	Registerbaserede arbejdsstyrkestatistik
10	KRAF	Kriminalstatistik afgørelser

Projektdatabaser

Forskningservice har gennem en årrække tilbudt projektdatabaseordningen til forskningsmiljøer med særlige behov. En projektdatabase indeholder en samling af mange forskellige mikrodata, som over tid anvendes til flere forskellige forskningsprojekter i det samme forskningsmiljø. Målgruppen for ordningen med projektdatabaser under forskerordningen er autoriserede institutioner med et omfangsrigt forskningsmiljø, som gennemfører et stort antal forskningsprojekter med væsentligt overlap i deres datagrundlag. Det er et krav, at den autoriserede institution har sin egen databaseadministrator.

Efterspørgslen efter og interessen for projektdatabaser er fortsat eksisterende. I 2019 blev der både oprettet og nedlagt projektdatabaser således, at der ved udgangen af 2019 var i alt 37 projektdatabaser, ligesom ved udgangen af 2018. Af de 37 projektdatabaser er hovedparten offentlige kunder (28), suppleret af et stigende antal private kunder fordelt på henholdsvis interesseorganisationer (7) og konsulentvirksomheder (2).

Informationsmateriale om projektdatabaseordningen er tilgængeligt på Forskningservice' hjemmeside:

<http://www.dst.dk/da/TilSalg/Forskningservice/Dataadgang/Projektoprettelse.aspx>

Brud på datasikkerhed eller -fortrolighed

Danmarks Statistik har siden 2014 skærpet kontrollen med og sanktionerne ved overtrædelser af Danmarks Statistiks datafortrolighedsregler.

Skærpelsen af vores fokus på disse forhold har betydet at forskningsmiljøer, som i strid med autorisations- og forskeraftaler hjemsender mikrodata fra vores forskerordning, får lukket deres adgang til data. Når Forskningservice efterfølgende får

en redegørelse fra forskningsmiljøet om, hvad der er sket, samt en plan for hvordan miljøet fremover vil sikre, at nye brud på datafortroligheden undgås, tages sagen op i Danmarks Statistiks direktion, der beslutter hvilken sanktion miljøet skal have. Det vil typisk være en måneds udelukkelse for en institution ved førstegangsovertrædelser.

Der blev kun lukket for adgangen for 5 forskningsmiljøer i 2019. Nedlukningerne var jævnt fordelt over de 4 kvartaler i 2019. Tabel 1 nedenfor viser antallet af nedlukninger og lukkede institutioner.

Tabel 1. Nedlukkede forskningsmiljøer som følge af brud på datafortroligheden

År	2017	2018	2019
Nedlukninger	17	9	5
Nedlukkede forskningsmiljøer	16	9	5

Note. Hvis antallet af nedlukninger er større end nedlukkede forskningsmiljøer, så skyldes det at et forskningsmiljø er blevet nedlukket flere gange i løbet af året.

Generelt har antallet af nedlukninger været faldende i den periode hvor Danmarks Statistik har strammet sanktionerne ved institutionernes hjemsendelse af mikrodata under forskerordningen.

I 2014 var der 24 nedlukninger af institutioner, mod 5 nedlukninger i 2019 jf. tabellen nedenfor.

Tabel 2. Udvikling i antallet af nedlukkede institutioner og tiltag for at nedbringe databrud

Periode	Antal brud	Tiltag for at reducere brud
2014	24	
2015	19	Informationsmøder om datasikkerhed med institutionerne
2016	20	Seminar om datasikkerhed i DST
2017	17	Seminar om datasikkerhed i DST Nyt hjemsendelsessystem
2018	9	Seminar om datasikkerhed i DST
2019	5	Seminar om datasikkerhed i DST Hjemsendelsessystem version 2

I perioden er der løbende igangsat nye tiltag for at reducere miljøernes databrud fra afholdelse af møder med institutionerne om datasikkerhed, udvikling af et datasikkerhedskursus med fokus på regler for adgang til mikrodata under forskerordningerne, til - i samarbejde med KOR - udvikling af et hjemsendelsessystem, hvor brugerne bliver advaret, hvis systemet mistænker, at de resultater som brugerne vil hjemsende indeholder mikrodata.

Særligt fra andet halvår af 2017, hvor det nye hjemsendelsessystem blev sat i drift, har disse initiativer haft effekt på reduktionen af databrud, der medfører en midlertidig lukning af institutionen.

Årets leveringstider

Den gennemsnitlige effektive leveringstid på et projekt var lidt under 10 dage i 2019. Den effektive leveringstid er den tid der går, fra der er enighed om hvilke data og population et projekt skal omfatte, indtil data er leveret.

Den effektive leveringstid, der er opgjort i hverdage, ligger på stort set samme niveau som i 2018.

Tabel 2 nedenfor viser andel projekter efter leveringstider i 2019.

Tabel 2. **Leveringstider 2019**

Leveringstid	Andel projekter	Antal projekter
1 uge	72,0	1065
1-2 uger	7,8	115
3-4 uger	9,9	147
1-2 måneder	7,4	110
Over to måneder	2,9	42
I alt	100	1479

Note: I opgørelsen af leveringstider indgår i Antal projekter antallet af projekter der i løbet af 2019 har fået leverancer. Enten i form af nye projektdata, opdateringer eller overførsler af eksterne data. Dvs. at dette tal godt kan afvige fra det samlede antal aktive projekter, da ikke alle aktive projekter har haft behov for nye data i løbet af 2019.

Lidt under 90 procent af projekterne blev i 2019 afleveret indenfor en måned. De resterende projekter havde en effektiv leveringstid, der lå over en måned. Disse projekter fordelte sig med ca. 7 procent hvor leveringstiden var på mellem 1 og 2 måneder og for de resterende ca. 3 procent af projekterne, lå leveringstiden på over 2 måneder.

De projekter, der indgår i opgørelsen er alle projekter, hvor data er leveret i 2019, dvs. at der i opgørelsen også indgår projekter, der er oprettet i tidligere år.

Projektdata indgår ikke i opgørelsen, da kontraktaftaler for projektdata typisk indgås for et kalenderår. Leveringen foregår her løbende over året i takt med at relevante grunddata bliver opdateret.

Igangværende projekter med en leveringstid på over to måneder

I forbindelse med KOR-aftalen for 2019 må heller ikke igangværende projekter have en nettoliveringstid på mere end to måneder. I givet fald skal Forskningservice give en begrundelse på sagsniveau med en eventuel redegørelse for hvad Forskningservice har gjort for at få sagen afsluttet.

Forskningservice har i forbindelse med kvartalsrapporten fra 2. kvartal 2019 og frem opgjort antallet af aktive projekter med en leveringstid på over to måneder, samt begrundelserne for den lange leveringstid.

I 2. kvartal 2019 blev der identificeret 59 aktive projekter med en leveringstid på over to måneder. I 4. kvartal 2019 var antallet faldet til 39.

I langt de fleste tilfælde skyldes den lange leveringstid at Forskningservice mangler en afklaring fra forsker for at kunne fortsætte projektet. I de tre kvartaler lå denne andel af projekterne tæt på 60 procent. I størstedelen af de øvrige projekter blev data enten leveret i opgørelsesperioden, eller projektet blev aktivt.

Det er forventningen at en stor del af de projekter hvor forskeren ikke er vendt tilbage på trods af henvendelser fra Forskningservice, ikke bliver til noget. Det er

derfor aftale med KOR, at projekter hvor Forskningservice ikke har fået en tilbagemelding indenfor 6 måneder, lukkes efter at forsker er blevet adviseret herom. Denne procedure starter i 2020.

Status for Højkvalitetsdokumentation

I højkvalitetsdokumentationen på Forskningservice´ hjemmeside er flere statistikområder i 2019 indenfor emneområderne ” Sociale forhold, sundhed og retsvæsen”, ”Beskæftigelse”, ”Befolkning”, ”Uddannelse” og ”Privatøkonomi” opdateret med ajourførte grafer og tabeller. Ligeledes er indholdet i dokumentationen blevet opdateret.

Indenfor emneområdet ”Sociale forhold, sundhed og retsvæsen” vedrører det kriminalitets-statistikområderne. Indenfor emneområdet ”Beskæftigelse” vedrører det statistikområderne ”Befolkningens tilknytning til arbejdsmarkedet (RAS)” og ”Beskæftigede personer – RAS”, som beskrevet i nedenstående tabel.

HKD Emneområde	HKD Statistikområde
Sociale forhold, sundhed og retsvæsen	Kriminalitet - afgørelser
Sociale forhold, sundhed og retsvæsen	Kriminalitet - anmeldelser
Sociale forhold, sundhed og retsvæsen	Kriminalitet - indsættelser
Sociale forhold, sundhed og retsvæsen	Kriminalitet – konfererede sager
Sociale forhold, sundhed og retsvæsen	Kriminalitet – ofre for straffelovsforbrydelser
Sociale forhold, sundhed og retsvæsen	Kriminalitet – sigtelser
Beskæftigelse	Befolkningens tilknytning til arbejdsmarkedet (RAS)
Beskæftigelse	Beskæftigede personer – RAS

Indenfor emneområdet "Uddannelse" vedrører det statistikområderne "Højst fuldførte uddannelse" og "Elevregister 3". Indenfor emneområdet "Befolkning" vedrører det statistikområderne "Folketal", "Opholdsgrundlag" samt "Familie".

HKD Emneområde	HKD Statistikområde
Uddannelse	Højst fuldførte uddannelse
Uddannelse	Elevregister 3
Befolkning	Folketal
Befolkning	Opholdsgrundlag
Befolkning	Familie

Høj kvalitetsdokumentationen kan ses på følgende link;

<http://www.dst.dk/da/TilSalg/Forskningsservice/Dokumentation/hoejkvalitetsvariable>

Nye registre, forbedringer og større rettelser vedrørende data

I løbet af 2019 er der arbejdet ekstra meget med vedligeholdelse af grunddata i Forskningsservice. Mange nye registre er oprettet og et omfattende antal registre er kørt om grundet kvalitetsforbedringer og rettelser i data. I det følgende fremhæves de vigtigste tiltag og ændringer på området.

Nyt register om akkumulerede indkomster (INDAKK)

I november 2019 blev et nyt indkomstregister, benævnt INDAKK – Akkumulerede indkomster, tilgængeligt i Forskningsservice. Registret indeholder akkumulerede indkomster, dvs. summen af personers indkomster over flere år. Registeret er et årligt register og dækkede i november 2019 perioden 1980 til 2017.

Den væsentligste variabel i registret er den akkumulerede disponible indkomst – altså summen af den disponible indkomst for alle år, som i november 2019 var 38 år. Ud over den disponible indkomst indgår en række øvrige variable, herunder fordelingen af bruttoindkomsten på løn, overskud af egen virksomhed, formueindkomst, overførsler mv. Fra 2014 og frem indgår også oplysninger om den samlede pensionsformue.

Beløbene er efterfølgende omregnet til 2017-prisniveau af hensyn til sammenligneligheden. Registret vil årligt blive opdateret. Ved opdateringen pristalsreguleres

registret med forbrugerpristallet, hvilket betyder at indkomstdataene hvert år opdateres til et nyt prisniveau.

Populationen i registret er i hvert år alle personer, der fandtes i indkomstregistret det pågældende år.

Nyt register om Børnefamilieydelse og –tilskud (BOFAMYD)

Det nye register om Børnefamilieydelse og tilskud erstatter de to registre BOBO (Børnetilskud barn) og BOTI (Børnetilskudsmodtager).

Registret indeholder kvartalsvise data fra 2017 til 2018 med referencetid sidste dag i kvartalet. I det nye register er der en række for hver kombination af modtagere og børn.

Nyt register om Boligstøtte (BOST)

Det nye register om Boligstøtte erstatter fra og med 2019 registrene Boligsikring (BOSA) og Boligydelse (BOYA). Alle registrene er opdateret i sommeren 2019 med data for årene 2007-2018 men fra 2019 vil kun Boligstøtte (BOST) opdateres. Registerne BOSA og BOYA vil stadig være tilgængelige, men kun med data til og med 2018.

BOST indeholder opgørelser for boligstøtten – både boligsikring og ydelse – for hver måned i stedet for kun december måned tilbage til 2007. Det er således muligt at følge den årlige variation af boligstøtte.

Derudover er der følgende ændringer i variablene:

- BOST_CBOTSAG viser koden for boligstøttetype.
- SAGTYPR i BOSA/BOYA hedder BOST_CBOTSAG i BOST.
- Grupperingen i BOST_SAGGRUP er ændret.
- BELOEB i BOSA/BOYA hedder BOST_UDB i BOST.
- TILSK, som var den rene boligstøtte, udgår.
- BOST_SATS introduceres som summen af boligstøtten og evt. lån, hvilket i langt de fleste tilfælde svarer til TILSK.
- BOST_HUSLEJE ændres, så beløbet er den reelle månedlige husleje i BOST, hvilket svarer til FAKLEJE i BOSA/BOYA.
- FAKLEJE udgår.

Nyt register om Personer i bestyrelse (PERSBEST)

Det nye register, Personer i bestyrelse (PERSBEST), baseres på statikken om medlemmer i bestyrelser og direktioner.

Statistikken er baseret på Erhvervsstyrelsens deltagerregister, hvor selskaber er forpligtede til selv at rapportere informationer om personer mv., der har tilknytning til selskabet, hvorfor der bør være fuld dækning af alle selskabers ledelsesorganer. Statistikken er afgrænset til private selskaber af typen anparts-, aktie- og iværksætterselskaber over Danmarks Statistiks bagatelgrænse.

Personer kan indgå flere gange i statistikken. Såfremt en person indgår både i en virksomheds direktion, samt i dennes bestyrelse, vil vedkommende tælle i begge roller. Endvidere tæller en person flere gange, hvis vedkommende f.eks. er medlem af bestyrelsen i flere selskaber. Der er en række personer, hvor det ikke er muligt at identificere baggrundsoplysninger som køn, uddannelse mv., f.eks. fordi bestyrelsesmedlemmet ikke er en del af den danske befolkning.

Statistikken udkom første gang i 2019 og udgives herefter årligt.

Nye registre vedr. kontospecifikke data fra Skatteforvaltningen (IRTE, URTE, PANT)

Danmarks Statistik har indgået en aftale med Skatteforvaltningen om adgang til kontospecifikke data IRTE, URTE, PANT for personer og virksomheder.

Denne aftale er en del af forskningsinfrastrukturprojektet DRDS, der giver forskerne adgang til en række nyere data.

Data indeholder informationer om

- Indlånsrenter (IRTE)
- Udlånsrenter (URTE)
- Pantebreve (PANT)

Følgende registre er udgivet i november/december 2019

- IRTEPERS 2003-2018
- IRTEVIRK 2003-2018

- URTEPERS 2003-2018
- URTEVIRK 2003-2018

- PANTPERS 2003-2018
- PANTVIRK 2003-2018

Data er opgjort ultimo hvert år og opdateres fremover årligt.

Ny proces for opdatering af den Registerbaserede arbejdsstyrkestatistik (RAS) samt rettelser

Den Registerbaserede arbejdsstyrkestatistik blev fra og med offentliggørelsen af RAS2018 i november 2019 opdateret ad to omgange i Forskningsservices grunddatabank. Statistikken bliver hermed tilgængelig flere måneder tidligere end før.

Ved offentliggørelsen i den første version mangler registret den underopdelte klassifikation af personer uden for arbejdsstyrken. Samtlige personer uden for arbejdsstyrken har i den frigivne version-1 en "SOC-STATUS-KODE" = 300. Desuden mangler registret variabelen "AFSTAND".

I den efterfølgende version 2, som fremover vil offentliggøres i februar det efterfølgende år, er variabelen "AFSTAND" tilføjet og klassifikationen af personer uden for arbejdsstyrken er opdateret.

For RAS2016-2018 er variabelen "AFSTAND" tilføjet i oktober og for RAS2017 er der yderligere tilføjet værdier i "ARB_DST_ADGANGSADRESSE_ID" og "ARB_DST_ENHEDSADRESSE_ID", dvs. antallet af uoplyste for disse to variable er reduceret i november 2019.

Ny version tilgængelig af befolkningsregistret (BEF)

I marts 2019 blev registrene for Befolkningen (år) (BEF), Befolkningen (kvt.) (BEFK), Familieforhold (FAM) og Husstandsoplysninger (HUST) revideret og samlet i registret Befolkningen (BEF).

Registret vil være navngivet efter referencetidspunktet for det data som et datasæt indeholder. For BEF er dette en status opgjort på sidste dag ultimo kvartalet. Denne oplysning kan også findes i variabelen REFERENCETID. Registret opdateres kvartalsvis fremadrettet.

Variablene BOPIKOM, OPGIKOM, KOM og REG kan ikke længere findes i BEF. I stedet vil variabelen ADRESSE_ID kunne bruges til at koble med registret BEFADR. I BEFADR findes variablene BOPIKOM, OPGIKOM og KOM. Der er periodeangivelser, således at man kan finde en kommunekode ud fra det tidsmæssige perspektiv, man måtte ønske.

Omlægningen er sket, for at få BEF til at leve op til FSE-grunddatapolitikken om at data, der stilles til rådighed via Forskningsservice, afspejler det datagrundlag, som er dannet i statistikkontorene i Danmarks Statistik.

Bygge- og boligregisteret samt Boligtællingen tilgængelig i en ny version (BBR, BOL)

En ny version af Bygge- og boligregistre samt Boligtællingen blev i 2019 tilgængelig i Forskningsservice.

Fire nye BBR-registre har afløst de hidtidige BBR-registre. Indholdet i de fire nye registre BBRBYGNING, BBREJENDOM, BBRENHED og BBRJORD svarer stort set til indholdet i de tidligere registre. Det er samme observationer, der optræder, og med fordelingen på de fire registre afspejles strukturen fra det register, som datasættet dannes af.

I navngivningen af registre indgår år og måned som en del af navnet. De nye registre dækker perioden 200912 - 201712. De tidligere registre BBRB og BBRE vil fortsat være tilgængelige for perioden 200412 - 200812 (tidligere 2005-2009). BBRB og BBRE for 200912 og frem vil herefter ikke være tilgængelige i grunddata.

Der er anvendt nye værdisæt for kodetekstvariable, og det er derfor ikke muligt at lave en simpel videreførsel fra de tidligere BBR-registre til de nye. Det er muligt at videreføre en tidsserie, hvis man korrigerer for ændringer i kodetekstværdier. Til hvert af de nye registre, undtagen BBRJORD, er der tilknyttet et nøgledatasæt, som kan være nødvendigt i forbindelse med sammenkøring med andre registre. Tidligere har nøglerne været lagt ind i BBR-registrene. Baggrunden for, at der anvendes separate nøgledatasæt er, at der fremover kun vil findes et nøgledatasæt, der gælder for alle år.

Omlægningen af BBR-registre betyder, at det fremover vil være i BBR man skal finde mange af de variable, der hidtil har været tilgængelige i registeret BOL.

Omlægningen af Forskningsservices BBR-registre sker som følge af den omlægning, der er foregået i SKATs BBR til version 1.8, og som har stået på gennem de to sidste år. På grund af en række problemer i forbindelse med overgangen, har Forskningsservice ventet med at implementere omlægningen, indtil der var melding om, at registrene var tilstrækkeligt anvendelige.

Ny version af Landspatientregistret (LPR2)

LPR2 - Landspatientregistret blev i 2019 opdateret til og med 2018 samtidig med at nye versioner af de fleste LPR2 tabeller for perioden 2005-2017 blev opdateret i grunddata i Forskningsservice.

Genkørslen er foretaget for alle LPR_ (Afsluttede kontakter), UAF_ (Uafsluttede kontakter) og PRIV_ (Private hospitaler) tabeller, i tidsperioden 2005 til 2017.

Landspatientregistret opdateres løbende af Sundhedsdatastyrelsen, både med nye kontakter og med rettelser i gamle kontakter, tilbage i tid. Den nye version indeholder alle ændringer foretaget af Sundhedsdatastyrelsen siden sidste genkørsel.

Forbedring af det Medicinske Fødselsregister (MFR)

Det Medicinske Fødselsregister - MFR blev i 2019 opdateret i grunddata i Forskningsservice for årene 1997-2017. Registeret har undergået nogle kvalitetsforbedringer og rettelser tilbage i tid.

Registreringen af dødsdato og dødsalder for barn er ændret således, at det nu er muligt at opgøre børnedødelighed i første leveår. Registreringen af SKS-koder i variabelen "epiduralblokade" er nu afgrænset således, at i tilfælde, hvor der for samme fødsel er indberettet flere SKS-koder af typen "NAADo*" kan der ikke prioriteres imellem alle registreringer af den konkrete kode.

Yderligere er der rettet op på fejlbehæftede/manglende beregning af antal tidligere fødsler og kejsersnit i Danmark. Rettelsen giver sig primært udslag i årene fra 2013 og frem.

Rettelse i og forbedring af IDA arbejdssteder (IDAS)

Ida arbejdssteder er for året 2008 blevet rettet, idet ukorrekte baggrundsoplysninger for ca. 30.000 arbejdssteder var medtaget i registret. Samtidig blev registret opdateret for årene 2008-2016 med følgende ændringer:

- tilføjelse af nye variabler "KOM" og "OPGIKOM"
- variabelen BRANCHE er omdøbt til ARB_HOVED_BRA_DBo7 og dermed navngivet på samme måde som i IDAN
- variablerne LONGNS og LONAAR er slettet. Informationen om den sumerede løn ligger i stedet i variablerne SMALT_LOEN_BELOEB og BREDT_LOEN_BELOEB

Forbedret opgørelse af Iværksætterdatabasen (IVDB)

Iværksætterdatabasen (IVDB) blev i januar 2019 opdateret frem til 2017 og ensrettet tilbage til 2001. I modsætning til tidligere versioner af databasen fungerer den opdaterede database, som en nøgle indeholdende CVR-nummer og PNR. Det er således muligt for brugerne selv at koble til relevante erhvervs- og personstatistikker og tilgå de ønskede variable.

Rettelser i Børn og Unge registrene (BUAF, BUAS, BUAH)

Børn og Unge registrene er for årene 2014-2017 blevet rettet, idet der var fejl i de indberettede data til Danmarks Statistik. Fejlen var indeholdt i 150 af de samtykke-koder, som var leveret til Danmarks Statistik af systemleverandøren. Årsagen hertil var den, at systemleverandøren tilbage i 2014 beklageligvis ikke har fået implementeret de ændringer i samtykkevariabelen, som Ankestyrelsen havde meldt ud.

Forbedring af Industriens salg af varer (VARS)

Industriens salg af varer (diskretioneret) (VARS) er opdateret for perioden 2014-2017 i grunddata i Forskningsservice. Mindre kvalitetsforbedringer er foretaget i forbindelse med statistikkontorets revision, som løbende foretages med tilbagevirkende kraft for 3 år.

Forbedret version af nøglen mellem IDA og Firmastatistik (FIDA)

I maj 2019 blev FIDA opdateret for årene 2008-2017. Opdateringen er baseret på e-indkomst data i IDA og fungerer, i modsætning til tidligere versioner, som en nøgle indeholdende ARBGNR, ARBNR, CVRNR, OK_NR og PNR. Det er således muligt for brugerne selv at koble til relevante erhvervs- og personstatistikker og tilgå de ønskede variable.

Fra- og til datoerne i det nye FIDA vedrører dateringerne på de enkelte job.

Rettelse i Bystørrelser på arbejdssteder (BYSTRA)

Registret Bystørrelser på arbejdssteder blev i 2019 opdateret for året 2014, idet variabelen DSTKODE indeholdt godt 5000 forkerte koder.

Rettelse i det Komprimerede elevregister (KOTRE)

Det komprimerede elevregister blev i 2019 tilrettet for årene 2017 og 2018 i KOTRE2018. Rettelsen bestod af en tilføjelse af 1400 nye elever på grundskoleområdet og 23 elever på gymnasieområdet, som ikke tidligere havde været med i registret.

Forbedring af og rettelse i Samlet opgørelse af formue og gæld (FORMGELD)

I foråret 2019 blev der for årene 2014-2017 foretaget en kvalitetsforbedring af beregningen af bilformuen. Nettopriser, som udgør de priser, man kan forvente at få ved et salg, benyttes nu fremfor bruttopriser, hvor prisen også indeholdt en beregnet forhandleravance.

Senere på året i 2019 viste det sig, at der ved offentliggørelsen af FORMGELD2014, var sket en teknisk fejl i udarbejdelsen af data. Dette bevirkede, at der ikke var nogen observationer med værdien A.1 (Ejerbolig) under variabelen "Formueart". Der til indeholdt datasættet ikke tallene for den nyeste kvalitetsforbedring omkring ændringerne i opgørelsen af bilformuen fra brutto- til nettopriser.

Fejlene var afgrænset til datasæt FORMGELD2014, som derfor blev opdateret i efteråret 2019.

Rettelser i Børn og unge – Underretninger (BUU)

I efteråret 2019 blev årene 2015-2017 opdateret for registret omhandlende underretninger for børn og unge. Registret opdateres årligt med et nyt år og for årene tilbage i tid, idet tidligere år opdateres i forbindelse med udgivelsen af det nyeste år. For årene 2015-2017 var de seneste rettelser desværre ikke sket rettidigt og derfor ikke med ved offentliggørelsen af BUU2018.

Forbedring af og rettelser i opgørelsen af Boformer (BOFORM_F)

Ved udgivelsen af BOFORM_F2018 var data baseret på en forbedret metode til databehandling af overlappende ophold. Samtidig indgik rettelser tilbage i tid i registret. Registret indeholder således væsentlige ændringer i data i forhold til 2017-data. Rettelserne tilbage i tid er fortaget ud fra indberettet data fra bostederne i 2018.

Forbedring af Kursistregistret – kursister ved voksen- og efteruddannelse (VEUV)

Fra og med 2018 er variabelen Organisationstilhørsforhold (ORGTILH) ikke længere indeholdt i Kursistregistret men erstattet af variabelen områdekode (OMR_KODE). Den nye variabel tager udgangspunkt i hvilken type institution, der har indberettet data til Danmarks Statistik baseret på 7 indberetningskilder.

Forbedring af Takster for Somatik – Landspatientregistret (DRGSOMA)

I 2019 blev en ny version af Takster for Somatik opdateret i grunddata i Forskningsservice, idet DRGSOMA fra 2018 er blevet omstruktureret. Den nye struktur afspejler, at borgernes kontakt med sundhedssystemet skal ses som forløb og ikke blot enkelte kontakter.

Det nye DRGSOMA samler ambulante og stationære kontakter i tre nye registre fra og med 2018;

- *DRGSOMA_KONTAKT* (Indeholder oplysninger om hver enkelt kontakt i DRG, som enten stammer fra LPR eller minipas)
- *DRGSOMA_SGHFORLOB* (DRG-kontakter samles i DRG-sygehusforløb ud fra en lokalitets- og tidsmæssig sammenhæng)
- *DRGSOMA_PROCEDURE* (Indeholder alle de procedure, der er tilknyttet til en DRG-kontakt)

Registrene DRGSOMA_HEL og DRGSOMA_AMB findes fortsat i den gamle form frem til og med 2017 men opdateres ikke længere.

DRG-registrene for de psykiatriske hospitalskontakter for 2018 er ikke omstruktureret, og findes fortsat i den gamle form.

Rettelse i §48E Indkomster (INUF)

I september 2019 blev årene 1999-2003 rettet for registret indeholdende §48E-indkomster. Årsagen hertil var en forkert formatering af variablerne F_SLUTDATO og F_STARTDATO, som ved opdateringen blev rettet til tekststreng med formatet: DDMM.

Rettelse i og forbedring af Indkomstregistret (IND), Familieindkomst (FAIK) og Arbejdsklassifikationsmodulet (AKM)

I foråret 2019 blev indkomstregistret kørt om for årene 2014-2017 grundet en rettelse i variablerne Beskst13 og pre_socio. Afgrænsningen af selvstændigt erhvervsdrivende var dannet forkert, idet en lille gruppe personer, ved en fejl var blevet klassificeret som lønmodtagere eller studerende, hvor de skulle have været klassificeret som reelt selvstændige.

I familieindkomstregistret (FAIK) medførte fejlen i afgrænsningen af selvstændigt erhvervsdrivende en hidtidig undervurdering af antallet af ”selvstændige” familier. Årene 2014-2017 er således også kørt om for dette register, hvor rettelsen er sket i variabelen Famsociogrup_13.

I arbejdsklassifikationsmodulet (AKM) medførte fejlen i afgrænsningen af selvstændigt erhvervsdrivende en rettelse af variabelen BEKST13 for årene 2014-2017.

I indkomstregistret forekommer en række variabler, som vedrører 2013-indkomstdefinitionerne, som er den nyeste indkomstdefinition. I foråret 2019 blev årene 1980-1986 suppleret med en tilbageførsel af disse variabler, som allerede befandt sig i registret for årene 1986 og frem:

- FORMUEINDK_BRUTTO
- PEROEVRIGFORMUE_13
- LEJEV_EGEN_BOLIG
- LOENMV_13
- NETOVSKUD_13

- ERHVERVSINDK_13
- FOLKEFORTID_13
- OFFPENS_EFTLON_13
- OVRIG_KONTANTHJALP_13
- OVRIG_DAGPENGE_AKAS_13
- SYG_BARSEL_13
- DAGPENGE_KONTANT_13
- OVRIG_OVERFORSEL_13
- OFF_OVERFORSEL_13
- RESUINK_13
- PERINDKIALT_13
- SKATTOT_13
- SKATMVIALT_13
- DISPON_13
- BESKST13
- PRE_SOCIO

Familieindkomstregisteret blev efterfølgende rettet for året 2017 på grund af, at variabelen FAMBOLIGTYPE ved en fejl var tom ved produktionen af FAIK2017.

Forbedringer og udviklingsaktiviteter 2019

Forskningservice forsøger hvert år at forbedre den service og de produkter vi leverer til forsknings- og analyseformål, og vi tilstræber løbende at kommunikere disse tiltag ud til forskerne. De primære virkemidler for kommunikation til forskningsmiljøet er dels udsendelse af mails til alle aktive forskere eller de autorisationsansvarlige og dels udsendelse af vores Nyhedsbrev, der informerer om nye tiltag mht. data, servere, programmel, vejledninger m.v.

Herudover har Forskningservice en hjemmeside, der i 2019 blev omorganiseret. Her kan bl.a. den løbende opdatering og status for Forskningservice' grunddata-bank til enhver tid følges sammen med de forventede udgivelsestider.

Projekt om Big-data

Flere forskere vil fremadrettet have behovet for at kunne kombinere registerdata fra Danmarks Statistik med Big Data, fx gendata.

I 2017 afsatte KOR derfor midler til, at Danmarks Statistik under forskerordningen kunne undersøge mulighederne for at inkludere Big Data i konkrete forskningsprojekter.

Det blev besluttet af Danmarks Statistiks direktion at gå videre med et testprojekt baseret på data fra iPSYCH, hvor Danmarks Statistiks forskningsservices IT-infrastruktur udvides til også at omfatte servere placeret hos Computerome, der har datakraften til at kunne håndtere beregninger baseret på Big Data.

Testprojektet fortsatte i 2019, og der arbejdes i samarbejde med KOR videre med en generel model for at give projekter adgang til data på servere placeret hos Nationalt Genom Center. Dette arbejde fortsætter i 2020.

Opdateret vejledning i hjemsendelse af resultater

Danmarks Statistik har udarbejdet en vejledning i hjemsendelse af resultater fra forskermaskinen. KOR har finansieret en udbygning og opdatering af vejledningen, så den indeholder konkrete eksempler på databrud med henblik på yderligere at

forebygge disse. Vejledningen opdateres løbende og anvendes ligeledes også i forbindelse med vore kurser om datasikkerhed. Vejledningen ligger tilgængelig på Forskningsservices hjemmeside under linket:

<https://www.dst.dk/da/TilSalg/Forskningservice/hjemtagelse-af-analyseresultater>

Fortsat udvikling af hjemsendelsessystem

Den 1. juni 2017 satte Forskningsservice et nyt hjemsendelsessystem i drift. Det nye hjemsendelsessystem skanner forskernes output for mikrodata og giver forskerne en advarsel, hvis der er mistanke om mikrodata i de resultater som forskerne vil hjemtage. Det nye system har siden det blev sat i drift været medvirkende til at reducere antallet af alvorlige databrud. I 2018 afsatte KOR midler til at forbedre systemet yderligere, så der dannes en log over de steder, hvor systemet mistænker en fil, der påtænkes hjemsendt, for at indeholde mikrodata. Derved gives brugerne et bedre grundlag for at beslutte om filen skal sendes hjem. Denne forbedring af systemet blev sat i drift i 2019 og har medvirket til at reducere antallet af databrud yderligere.

Bestillings- og Administrationssystem, GDPR og standardprogrammer

Det omfattende IT-udviklingsprojekt i Forskningsservice med at opbygge et overskueligt og brugervenligt bestillings – og administrationssystem (BOA), så forskerne direkte via vores hjemmeside kan bestille data til deres projekter, fortsatte i 2019. Dette bestillingssystem vil systematisere og sammenkæde dokumentation, databestilling, projektoprettelse mv. Dette vil på sigt gøre det langt nemmere og mere overskueligt at bestille data, og det vil også øge kvaliteten i databestillinger betragteligt. Samtidigt vil det give værdifulde oplysninger om, hvilke data der bruges, hvordan forskere, projekter og institutioner hænger sammen på kryds og tværs, og sagsbehandlingstider på forskellige typer projekter og kunder vil kunne beregnes meget præcist.

Derudover har vi ligeledes i 2019 forsat den omfattende implementering af persondataforordningen (GDPR) så alle autoriserede enheder med adgang til data nu har et forbedret aftalegrundlag på plads. Udviklingen af standardprogrammer –for både at effektivisere opgaveløsningen og sikre en så ensartet kvalitet som muligt er ligeledes forsat, og vi har herudover også forsat arbejdet med at standardisere vores opgavepriser og måden vi beregner priser på. Formålet med denne standardisering er at sikre kunderne ensartede priser for sammenlignelige opgaver og samtidig at få dækket alle Forskningsservices omkostninger.

Nedenfor en kort status på de tre nævnte projekter:

Udviklingsarbejdet med FSE-BOA fortsatte i 2019.

Udviklingen af FSE-BOA fortsatte i 2019. Der blev nået en stor milepæl ved årsskiftet 2019/2020, hvor første version af administrationssystemet blev lagt ud til de første eksterne brugere. FSE-BOA giver eksterne brugere mulighed for selv at administrere deres brugeradgange, herunder at tilknytte eksisterende brugere til projekter, ændre brugernes rettigheder, fx hjemsendelsesrettigheder samt oprette og nedlægge brugeres projektadgange. Institutionerne har derved fået et effektivt værktøj til administration af brugere og projekter, samtidig med at de har fået et langt større overblik.

Udviklingen fortsætter i 2020, hvor administrationsdelen af FSE-BOA videreudvikles og lægges ud til alle brugere. Samtidig arbejdes der videre på bestilingsdelen, hvor brugerne får mulighed for at bestille data online.

Persondataforordningen - GDPR.

Danmarks Statistik har i løbet af 2019 lavet et omfattende arbejde med at re-dokumentere, at bl.a. IDV området, herunder at Forskerordningen og Ministerieordningen lever op til persondataforordningen (GDPR). En del af implementeringsarbejdet har bl.a. handlet om at sikre databehandleraftaler med alle enheder der anvender mikrodataordningerne, og at få udarbejdet den årlige uvil-dig revisionserklæring (ISAE3000-erklæring) til imødekommelse af Dataansvarlige kunders tilsynspligt med Danmarks Statistiks behandling af deres data under mikrodataordningerne.

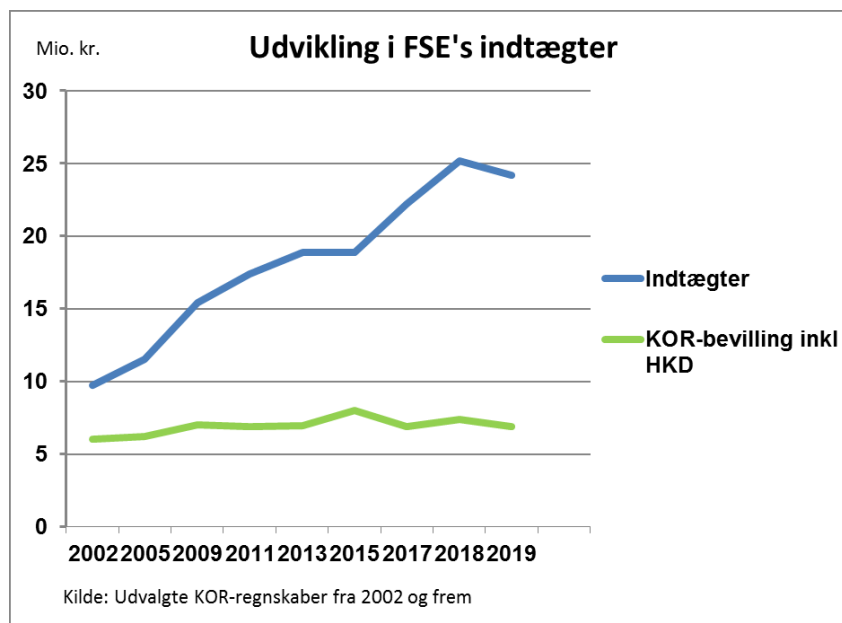
Standardprogrammer i FSE

Som et led i at kvalitetssikre og standardisere FSE's arbejdsgange udvikles der løbende standardprogrammer, der så vidt muligt sikrer en ensartet opgaveløsning på tværs af medarbejdere i FSE. I 2019 er der arbejdet videre på automatisering af afidentificering af såvel grunddata som eksterne data. Derudover er der udviklet udtræksskabeloner, som sikrer enkle og effektive udtræk af FSE's grunddata. Dette arbejde med yderligere standardisering fortsætter i 2020.

Økonomiske nøgletal/regnskab

Siden den første bevilling til Forskningsservice i 2002, har indtægterne i Forskningsservice vokset i takt med den øgede interesse for registerforskning.

Figuren nedenfor viser udviklingen i Forskningsservice' samlede indtægter og bevillingen fra KOR i udvalgte år fra 2002 og frem. I bevillingen indgår ligeledes i de særlige bevillinger til udarbejdelse af højkvalitetsdokumentation.



Note: I KOR-bevillingen indgår de årlige bevillinger til HKD-projektet, hjemsendelsessystem samt Computerome.

I 2002 udgjorde bevillingens andel af de samlede indtægter lidt over 60 pct. I 2019 var denne andel faldet til under 30 pct.

Regnskab 2019

Forskningservice havde i alt haft et underskud på lidt under. kr. 1 mio. i 2019. Nedenfor præsenteres regnskabet for 2019.

Tabel 3. Regnskab 2019

I regnskabet indgår hele årets KOR bevilling for 2019

Driftsregnskab:

Udgifter:	Regnskab 2019 kr.	Budget KOR tilskud kr.	Afvigelse kr.
Drift og udvikling i FSE	31.485.464	6.400.000	-25.085.464
Computerome	559.521	500.000	-59.521
Udgifter i alt	32.044.985	6.900.000	-25.144.985
Indtægter:			
Drift og udvikling i FSE	24.163.507	0	-24.163.507
Basisbevilling til medfinansiering af drift og udvikling i	6.400.000	6.400.000	0
Computerome	500.000	500.000	0
Indtægter i alt	31.063.507	6.900.000	-24.163.507
I alt driftregnskab (indt. - udg.)	-981.478	0	981.478
Årets resultat i alt (indt. - udg.)	-981.478	0	981.478

Regnskab 2019 er udarbejdet som skitseret i KOR-aftalen