

Årsrapport 2016

Indhold

Status for Danmarks Statistiks Forskningservice 2016	2
Vore kunder – de autoriserede institutioner	2
Projekter	2
Aktive forskere	3
Redegørelse for efterspørgsel på data.....	4
Leverede datasæt	4
Projektdatabaser	4
Brud på datasikkerhed eller -fortrolighed	4
Årets sagsbehandlingstider	5
Status for Højkvalitetsdokumentation.....	6
Forbedringer og udviklingsaktiviteter 2016	6
Forbedringer vedrørende data.....	6
Data om elforbrug.....	6
Beskæftigelse for lønmodtagere (BFL)	7
Forbedringer på forskningsmaskiner og opgraderinger	7
Styrket datafortrolighed	7
Effektivisering:.....	7
Nye særordninger	8
Økonomiske nøgletal/regnskab	8
Regnskab 2016.....	9

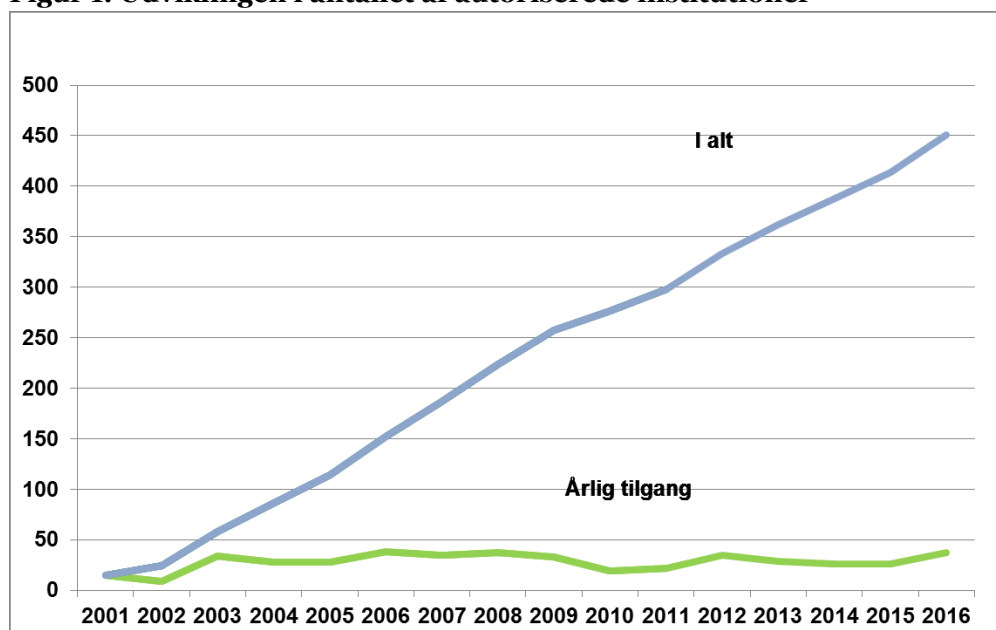
Status for Danmarks Statistiks Forskningservice 2016

Vore kunder – de autoriserede institutioner

Ved udgangen af 2016 havde Forskningservice **451 autoriserede institutioner**, hvoraf 37 nye institutioner blev autoriseret i 2016. Tilgangen af nye autoriserede institutioner har indtil videre vist sig at være meget stabil fra år til år, jf. figur 1 nedenfor.

Størstedelen af institutionerne kommer 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

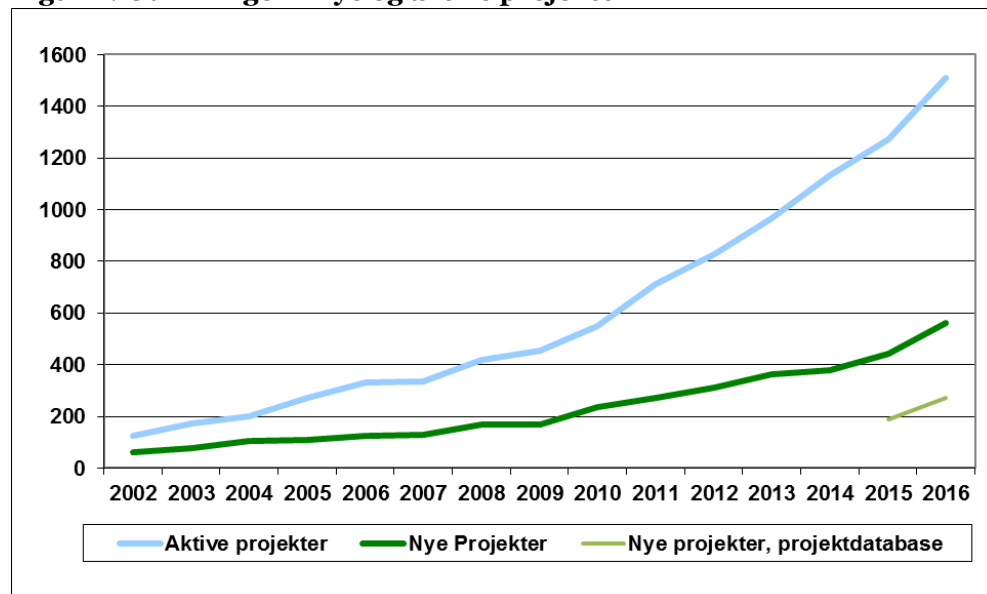


Projekter

Ved udgangen af 2016 arbejdede forskerne på over **1.500 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 det fortsat høje 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 2016 oprettet **561 nye projekter**, heraf 271 nye projekter fra de såkaldte projektdatabaser. 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. Figur 2 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 2016.

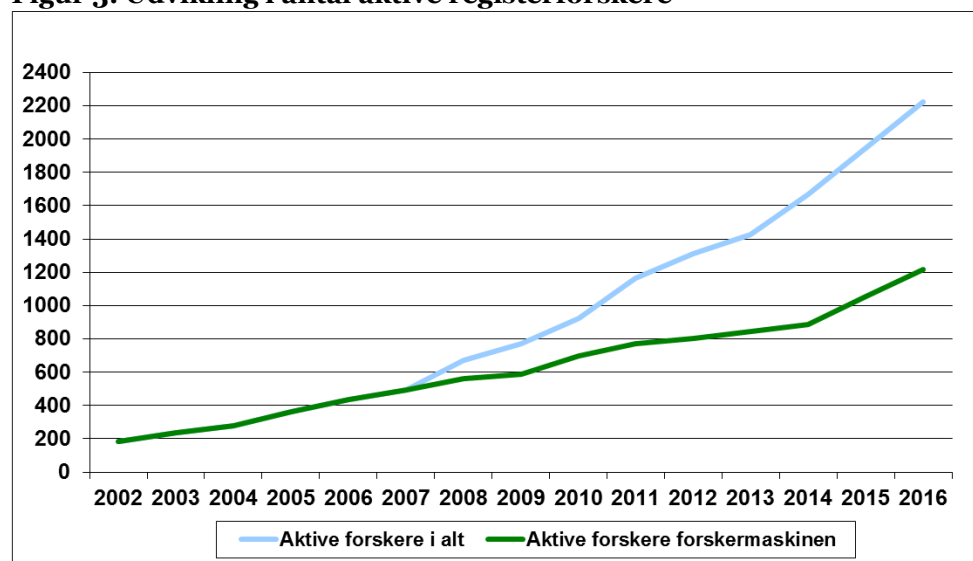
Figur 2. Udviklingen i nye og aktive projekter



Aktive forskere

I alt har ca. **2220 forskere i 2016 aktivt arbejdet på et eller flere projekter** på forskermaskinerne. Tæt på halvdelen af forskerne arbejder nu 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 2016 rådede Forskningservice' grunddatabank over 237 registre.

Der blev i alt leveret 129.351 filer, heraf blev 76.281 leveret som SAS-datasæt. De resterende blev først og fremmest leveret som STATA-datasæt eller SAS-view.

Nedenfor ses top 10 for leverede grunddatasæt i 2016 efter register:

1	BEF	Befolkningen
2	UDDA	Uddannelse
3	IND	Indkomst
4	LPRDIAG	Landspatientregisteret - diagnoser
5	AKM	Arbejdsklassifikationsmodulet
6	LPRPOP	Landspatientregistret -population
7	RAS	Registerbaserede arbejdsstyrkestatistik
8	INDH	Indkomst
9	IDAP	IDA-personer
10	LMDB	Lægemiddelstatistikregisteret

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 har været stadig stigende gennem årene. Der var ved udgangen af 2016 i alt 35 projektdatabaser mod 30 ved udgangen af 2015. Af de 35 projektdatabaser er hovedparten offentlige kunder (26), suppleret af et stigende antal private kunder fordelt på henholdsvis interesseorganisationer (8) og konsulentvirksomheder (1).

Informationsmateriale om projektdatabaseordningen er tilgængeligt på Forskningservice' hjemmeside:

<http://www.dst.dk/da/TilSalg/Forskningservice/Dataadgang/Projektrettelse.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 forsker-

ordning, får lukket deres adgang til data. Når Forskningsservice 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.

Det nye hjemsendelsessystem nåede ikke at blive taget i brug i 2016, og derfor blev adgangen til forskermaskinerne lukket for i alt 19 forskningsmiljøer, jf. tabellen nedenfor.

Tabel 1. Nedlukkede forskningsmiljøer som følge af brud på datafortroligheden

År	2015	2016
Nedlukninger	19	20
Nedlukkede forskningsmiljøer	16	19

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.

Lidt over halvdelen af udelukkelseerne blev på 1. måned. I de øvrige tilfælde har forskerne selv henvendt sig og gjort opmærksom på, at de ved en fejl er kommet til at hjemsende mikrodata. I disse tilfælde har forskningsinstitutionen kun været lukket ned, indtil sagen har været fuldt belyst. Det har typisk kun taget ganske få dage.

Årets sagsbehandlingstider

Den gennemsnitlige effektive sagsbehandlingstid på et projekt var ca. 20 dage i 2016. Den effektive sagsbehandlingstid er den tid der går, fra der er enighed om hvilke data og population et projekt skal omfatte, indtil data er leveret.

Den effektive sagsbehandlingstid, der er opgjort i hverdage, faldt lidt i 2016 i forhold til 2015, hvor den gennemsnitlige effektive sagsbehandlingstid var på 22 dage. Aktivitetsniveauet har i begge år været meget højt.

Tabel 2 nedenfor viser sagsbehandlingstiderne i 2016.

Tabel 2. Sagsbehandlingstider 2016

Leveringstid	Andel projekter	Antal projekter
1 uge	54,2	410
1-2 uger	10,9	82
3-4 uger	15,7	119
1-2 måneder	9,3	70
Over to måneder	9,9	75
I alt	100	756

Note: I opgørelsen af sagsbehandlingstider indgår i Antal projekter antallet af projekter der i løbet af 2016 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 2016.

80 procent af projekterne blev i 2016 afleveret indenfor en måned. De resterende projekter havde en effektiv sagsbehandlingstid, der lå over en måned. Disse projekter fordelte sig med 9,3 procent hvor leveringstiden var på mellem 1 og 2 måneder og for de resterende 9,9 procent af projekterne, lå leveringstiden på over 2 måneder.

De projekter, der indgår i opgørelsen er alle projekter, hvor data er leveret i 2016.

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

Status for Højkvalitetsdokumentation

Højkvalitetsdokumentationen på Forskningservice´ hjemmeside er i 2016 blevet udvidet med dokumentation af Indkomstregistret (nye definitioner), Opholdsgrundlag, Familiefordelt formue- og gældsstatistik og Firmaernes køb og salg.

Dokumentationen af Indkomstregistret og Familiefordelt formue- og gældsstatistik ligger i Emneområdet ”Privatøkonomi”, som beskrevet i nedenstående tabel. Indkomstregistret er dokumenteret under statistikområdet ”Personindkomster”.

HKD Emneområde	HKD Statistikområde
Privatøkonomi	Familiefordelt formue- og gældsstatistik
	Personindkomster

Dokumentationen af Opholdsgrundlag ligger i Emneområdet ”Befolkning”. Firmaernes køb og salg ligger i emneområdet ”Virksomheder, virksomhedsøkonomi og udenrigshandel.

HKD Emneområde	HKD Statistikområde
Befolkning	Opholdsgrundlag
Virksomheder, virksomhedsøkonomi og udenrigshandel	Firmaernes køb og salg

Højkvalitetsdokumentationen kan ses på følgende link;

<http://www.dst.dk/da/TilSalg/Forskningservice/Dokumentation/hoekvalitetsvariable>

Forbedringer og udviklingsaktiviteter 2016

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 og dels udsendelse af vores Nyhedsbrev, der informerer om nye tiltag mht. data, servere, programmer, vejledninger m.v.

Derudover kan den løbende opdatering og status for Forskningservice´ grunddatabank til enhver tid følges på vores hjemmeside, hvor også forventede udgivelsestider kan følges.

Forbedringer vedrørende data

Data om elforbrug

Fra og med oktober 2016 blev data om elforbrug i Danmark stillet til rådighed gennem Forskerordningen med afsæt i data fra Energinets Data Hub. Elforbruget er aflæst på elmålere og kan være enten time-, måneds-, kvartals- eller årsaflysninger.

I DST Forskningservice kan elmåler-data kobles til Danmarks Statistiks øvrige registre på adresseniveau.

Beskæftigelse for lønmodtagere (BFL)

Ved opdateringen i august 2016 af registrene for Beskæftigelse for lønmodtagere er alle registrene, undtagen BFLBRTIF (Aggregerede lønmodtagerdata på brancher), blevet kørt om tilbage til 2008 grundet forbedringer i datagrundlaget. Der er sket kvalitetsforbedringer med hensyn til opgørelse af arbejdstid for fleksjobbere, dagplejere og værnepligtige, forbedret arbejdsstedsplacering og brancheændringer fra beskyttede boliger til plejehjem.

Forbedringer på forskningsmaskiner og opgraderinger

Forskningservice monitorerer løbende behovet for programmel samt belastningen på forskningsmaskinerne. Formålet med monitoreringen er proaktivt at kunne imødekomme kapacitetsproblemer, før de opstår.

I 2016 blev udviklet en ny upload løsning, så forskerne kunne uploade deres egne data elektronisk til deres projekter, så de kan undgå at sende fortrolige data anbefalet på et medie eller aflevere data fysisk i Danmarks Statistiks reception.

Der blev ligeledes givet adgang til Statistikbanken.dk og domænet dst.dk direkte fra forskermaskinen/ministerieserveren/hostede maskiner hos Danmarks Statistik, så forskerne kan anvende Statistikbanken direkte i forbindelse med deres projekter.

Endelig blev der taget en ny "Diverse-server" i brug, primært for at imødekomme den stigende brug af programmet R.

Styrket datafortrolighed

Der blev i 2016 arbejdet en del med at forbedre datasikkerheden på Forskerordningen. Dels blev der arbejdet med at udvikle et *sikkerhedsprogrammel*, der skanner output for mikrodata og advare forskerne *før* de hjemsender output. KOR har støttet udviklingen af dette projekt med kr. 400.000 i 2016. Dette arbejde blev afsluttet medio 2017, hvor programmet blev implementeret på forskermaskinerne. Desuden blev der udviklet og udbudt et *forskerkursus i datasikkerhed*, som siden er blevet udbudt tre gange.

Endvidere har Danmarks Statistik igangsat et arbejde, der har til formål at *reducere medarbejdernes adgang til personnumre*. De registre der arbejdes med i Forskningservice må fremadrettet ikke indeholde personnumre. Personnumre skal i stedet erstattes af en neutral "ident". Dette arbejde er startet og vil fortsætte i hvert fald i de næste år. Det forventes ikke i første omgang at påvirke forskernes projekter (herunder opdateringer) nævneværdigt. Men der skal fremadrettet tages stilling til eksisterende projektdatabaser, som i fremtiden skal kunne opdateres på baggrund af den nye "ident". Dette arbejde forventes i 2018/19 at blive omfattende.

Endelig blev der som et led i arbejdet med at styrke datasikkerhed og – fortrolighed under Forskerordningen gennemført *en systematisk ekstern sikkerhedsvurdering* i 2016. Resultaterne af denne sikkerhedsvurdering fremlægges i 2017.

Effektivisering:

Det omfattende IT-udviklingsprojekt i Forskningservice med at opbygge et overskueligt og brugervenligt bestillingssystem, så forskerne direkte via vores hjemmeside kan bestille data til deres projekter, fortsatte i 2016. 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 og kombineres, hvordan forskere, projekter og institutioner hænger sammen på kryds og tværs, samt sagsbehandlingstider på forskellige typer projekter og kunder.

Som et led i dette program blev der i 2016 sat et forbedret administrationssystem i værk og arbejdet med at kravsificere bestillingssystemet fortsatte ligesom der blev foretaget forbedringer af de metadata, som systemet skal kunne trække på i fremtiden.

Udviklingsarbejdet med bestillingssystemet er planlagt til at fortsætte frem til og med 2018.

Nye særordninger

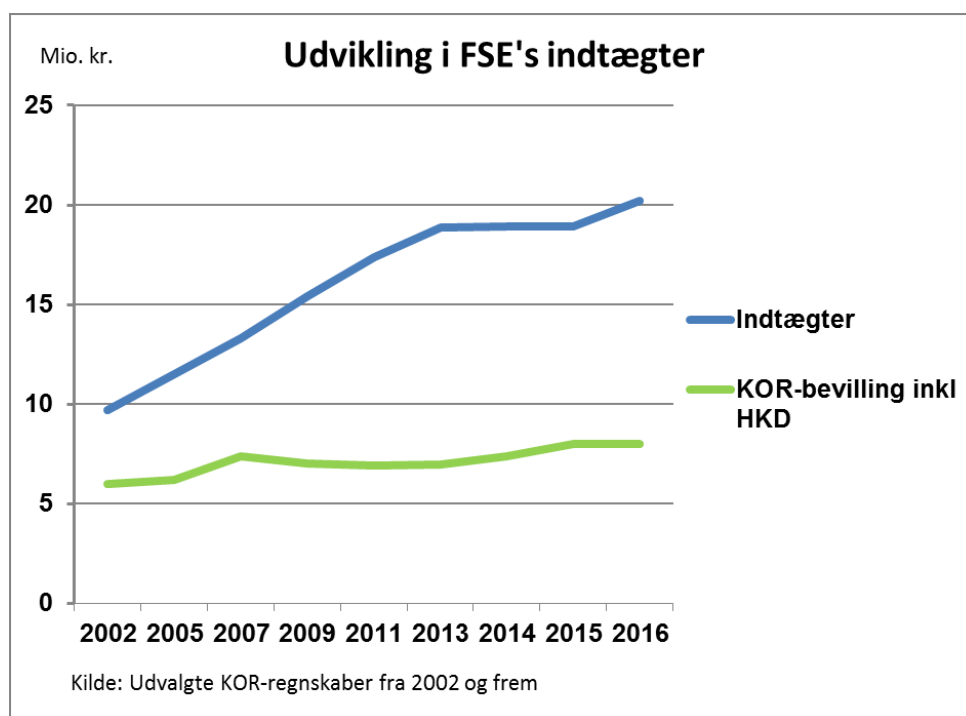
Sammen med de øvrige Nordiske Statistikker blev der i 2016 fortsat arbejdet i på at lette forskernes muligheder for at gennemføre komparativ nordisk registerforskning. Dette arbejde med at opbygge *en fælles Nordisk Forskerordning* omfattede i 2016 dels hjælp til en række konkrete nordiske forskningsprojekter, der ønskede adgang til tværnationale registerdata, og dels til opbygningen af sammenlignelige nordiske metadata. Sidstnævnte er finansieret af NORDFORSK. Dette arbejde fortsætter til udgangen af 2017.

Ministerieordningen er en ordning, der har til formål at stille data til rådighed for ministerier og styrelser. Ministerieordningen blev udvidet i 2016, hvor Erhvervsministeriet og Danmarks Nationalbank kom på ordningen. De særlige omkostninger, der er forbundet med denne ordning finansieres ikke af KOR, men af de involverede ministerier.

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 bevillingen indgår ligeledes i de seneste år den særlige bevilling til udarbejdelse af højkvalitetsdokumentation.



Note: I KOR-bevillingen indgår de årlige bevillinger til HKD-projektet.

I 2002 udgjorde bevillingens andel af de samlede indtægter lidt over 60 pct. I 2016 var denne andel faldet til ca. 40 pct.

Regnskab 2016

Forskningservice havde i alt et underskud på kr. 1,1 mio. i 2016. Underskuddet var planlagt, da Forskningservice i perioden 2009 – 2013 havde et akkumuleret overskud. Forskningservice skal over en periode have bragt regnskabet i balance. Indtil der er bragt balance vil Forskningservice fortsat køre med et årligt underskud.

Nedenfor præsenteres nogle nøgletal for Forskningsservice' økonomi for 2016.

Tabel 3. Nøgletalsregnskab 2016

Driftsregnskab	
Udgifter	Kr.
Lønudgifter i alt:	19.434.224
Heraf:	
• Basisbemanning	7.170.847
• Konstruktion af datasæt	11.021.465
• Udvikling af sikkerhedsprogrammel	568.992
• Håndtering af brud på datasikkerhed	68.345
• Bevilling lovmodel	604.576
Øvrige driftsudgifter (IT-udgifter)	2.490.267
Udgifter i alt:	21.924.491
Indtægter	
• Bevilling(1):	6.400.000
• Salg af datasæt:	10.572.090
• Bevilling Lovmodel	604.576
• Udvikling af sikkerhedsprogrammel	400.000
• Håndtering af brud på datasikkerhed	100.000
Diskeje og privates kørsler på forskermaskinen	2.189.585
Indtægter i alt:	20.266.251
I alt driftsregnskab	-1.658.240
Regnskab HKD-projekt	
Lønudgifter i alt:	555.824
Heraf:	
• Ekstern rådgivning	137.624
Indtægter	
HKD-bevilling(2)	1.126.926
I alt regnskab HKD-projekt	571.102
Årets resultat i alt	-1.087.102

Noter

(1) I bevillingen indgår både 85 procent af årets basisbevilling, samt sidste års HKD-bevilling, der kommer til udbetaling året efter, når hele HKD-projektet er godkendt.

(2) HKD-bevillingen er indtægter fra HKD-projektet. Bevillingen faktureres typisk året efter projektet er gennemført. Der kan derfor godt være en relativt stor forskel på indtægter og udgifter. Særligt hvis HKD-bevillingen har været forskellig fra år til år.