

Dato: 27. november 2017

Afrapportering af KORs undersøgelse af datavarehuse

Formål med undersøgelsen

Uddannelses- og Forskningsministeriets Registerforskningsudvalg anbefalede i rapporten "Registerforskning – Nye muligheder, nye udfordringer" fra 2013, at KOR, efter at have etableret i en referencegruppe for registerforskning "påbegynder et arbejde for at sikre en tværgående koordinering i forhold til de datavarehuse, som pt. er under opbygning i forskellige ministerier. Der er de senere år kommet betydelig fokus på anvendelse af registerdata til styring og analyse, og der er på flere ministerområder igangsat projekter for opbygning af datavarehusløsninger i eget regi - bl.a. for at sikre større viden om effekter af indsatser. Etableringen af datavarehuse for monitorering og analyse inden for arbejdsmarked, sundhed, uddannelse og socialforhold i regi af ressortministerier rummer store perspektiver for et øget samarbejde mellem forskningen og analytikere, men også for et øget koordinationsbehov, blandt andet for at sikre ensartethed i begrebsanvendelse og i forhold til dokumentationsstandarder."

KOR har derfor gennemført en spørgeskemaundersøgelse, der skal skabe grundlag for et samarbejde om udvikling, standardisering, dokumentation og adgang til ministerielle datavarehuse.

KOR vil præsentere undersøgelsen for referencegruppe for registerforskning, for at bruge den som grundlag for at kunne vurdere behovet et uddybende samarbejde.

KOR vil drøfte resultaterne med relevante forskermiljøer, for at vurdere, om forskerne ser noget behov for at arbejde på at give forskere og analytikere bedre muligheder for at få adgang til data fra ministerielle datavarehuse-

KOR

Rigsarkivet

Rigsdagsgården 9

1218 København K

Telefon: 4171 7211

Mail: jkd@sa.dk

Web: registerforskning.dk

Sagsnr.: 16/00590

Undersøgelsen

Hensigten med spørgeskemaet har været at indsamle oplysninger om hvilke ministerier og styrelser, der har datavarehuse på velfærdsområdet, herunder

- 1) oplysninger om datavarehusløsning,
- 2) oplysning om centrale populationer,
- 3) muligheder for adgang til data og
- 4) datadokumentation.

KOR har udarbejdet spørgeskemaet i samarbejde med relevante personer i Rigsarkivet, Styrelsen for IT og Læring samt Økonomisk Institut ved Aarhus Universitet.

KOR sendte spørgeskemaet ud til medlemmer af referencegruppe for registerforskning, med opfordring om at videresende til rette person eller orientere KOR om andre, der burde deltage i undersøgelsen. Det er KORs forventning, at referencegruppen selv dækker eller i det mindste er bedst orienteret om, hvilke personer, der har indsigt i aktiviteter vedr. datavarehuse i ministerier og styrelser, hvorfor undersøgelsen bør være kommet relativt bredt ud til relevante organisationer.

KOR sendte spørgeskema ud til 28 personer fra referencegruppen. Heraf har 21 besvaret skemaet, 2 har besvaret delvist og 5 har ikke svaret. Det giver en svarprocent på 75 %. Under hensyntagen til hvilke organisationer, der ikke har svaret, så vurderes det, at undersøgelsen er relativt dækkende, hvad angår bestanden af datavarehuse i statslige organisationer indenfor velfærdsområdet.

Af de 21 organisationer, der besvarede undersøgelsen, har 14 svaret, at de har et datavarehus, heraf her én organisation dog ikke svaret på resten af spørgsmålene. De 7 organisationer, der har svaret, at de ikke har et datavarehus har heller ikke aktuelle planer om det.

Der indgår således oplysninger om 13 datavarehuse i undersøgelsen, hvorfor det vurderes, at det i hovedsagen kun giver mening at beskrive undersøgelsen på et overordnet niveau.

KOR tager til efterretning, at undersøgelsen således er lagt op til at være mere grundig end nødvendigt.

Undersøgelsens resultater

Overordnet set er anvendelsen af datavarehuse i ministerier og styrelser ikke så udbredt, som forudset af Registerforskningsudvalget i 2013. I praksis er der stadig de tre hovedaktører på området, som udvalget pegede på. Disse områder er blevet konsolideret siden. Det drejer sig om datavarehuse om 1) uddannelse (dog ikke videregående uddannelser), der er forankret i Styrelsen for IT og læring, Beskæftigelse, der er forankret i styrelsen for arbejdsmarked og rekruttering, samt sundhed, der er drevet af sundhedsdataprogrammet som et samarbejde mellem bl.a. Sundhedsdatastyrelsen og RKKP.

Undersøgelsen kan ses i sin helhed som bilag_complete

Delkonklusioner indenfor de fire hovedspørgsmål

Datavarehusene anvendes hovedsageligt til ledelsesinformation og statistik og i meget mindre grad til forskning, administration m.v. Derudover er der stor forskel på, hvilke tekniske løsninger der er valgt, samt hvordan data stilles til rådighed for offentligheden.

Den tilvejebragte oversigt over populationer i datavarehusene og deres karakteristika bør kunne give forskere og analytikere et overblik over hvilke oplysninger, organisationerne har registreret i deres datavarehuse. Det er bemærkelsesværdigt, at der ikke er nogen datavarehus initiativer med oplysninger om det sociale område.

Kun seks styrelser udleverer data til forskere på anmodning, herunder alle de ovenfor nævnte hovedaktører. Der er muligvis et uudnyttet potentiale for forskere og analytikere, hvis der træffes aftaler om udlevering af data fra datavarehuse i (mindre) organisationer, der ikke har nogen praksis for dette.

Overordnet set er der ikke det store fokus på dokumentationsstandarder eller harmonisering med øvrige organisationer, bortset fra på sundhedsområdet. Eventuel harmonisering er sket i forbindelse med anvendelse af Danmarks Statistiks ministerieordning.

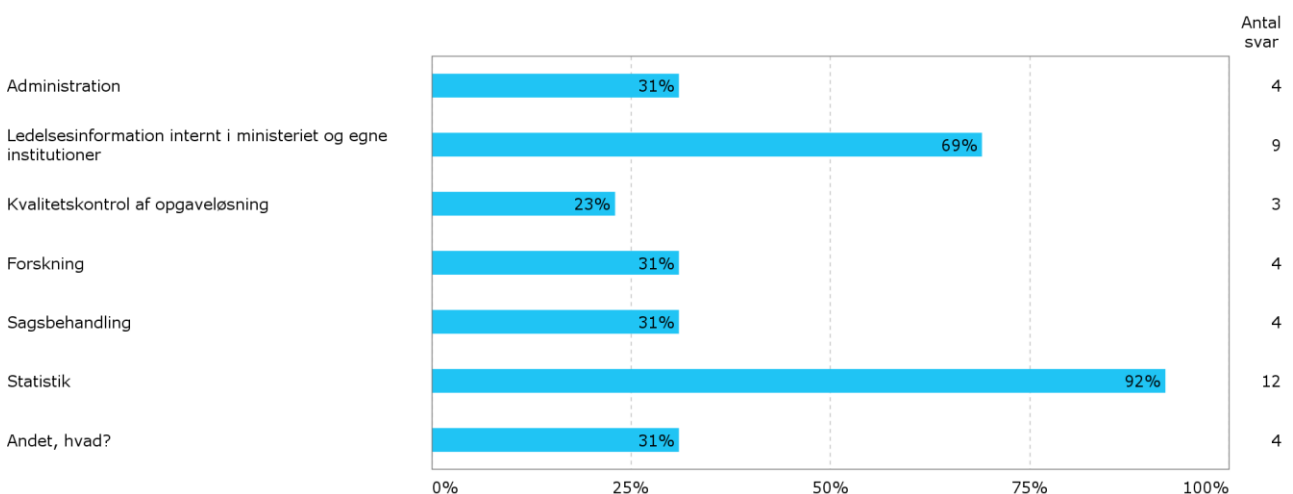
Hovedspørgsmål

Overordnet rejser undersøgelsen efter KORs mening nedenstående spørgsmål, som KOR vil arbejde videre med.

- Synes forskere og analytikere at der skal arbejdes henimod at gøre det lettere at få adgang til data fra datavarehuse i organisationer, der ikke har nogen praksis for dette?
- Ser de organisationer, der har et datavarehus, behov for koordination og harmonisering?
 - o Skal der etableres et samarbejdsforum blandt de organisationer, der har et datavarehus?
 - o Kan en eventuel harmonisering af variabler og algoritmer tage udgangspunkt i hvordan organisationerne anvender Danmarks Statistiks ministerieordning?

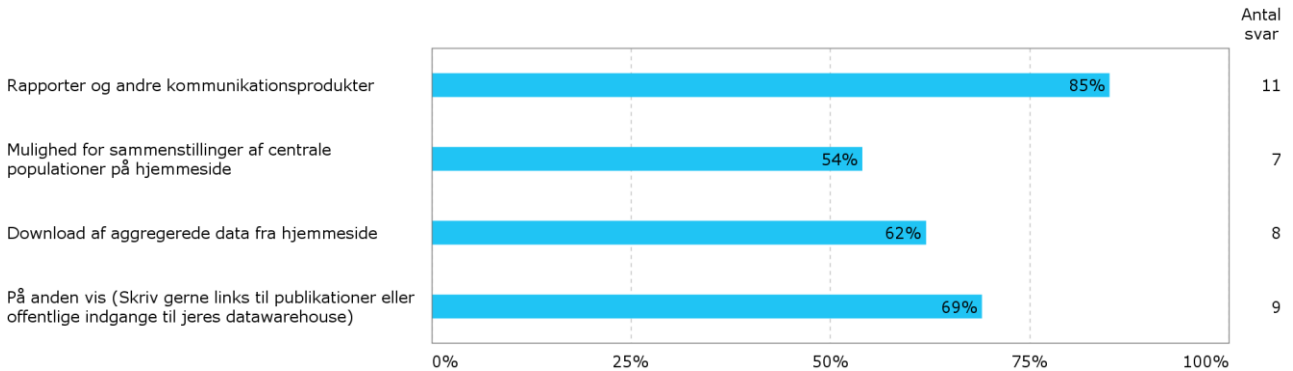
1. Oplysninger om datavarehusløsning

Hvad bruger I selv data jeres datavarehus til?



Datavarehusene bliver hovedsageligt anvendt til ledelsesinformation og statistik i mindre grad til forskning, administration etc.

Hvordan stilles data til rådighed for offentligheden?



Der stilles data til rådighed på flere forskellige måder, hovedsageligt som rapporter og kommunikationsprodukter.

Hvor kommer data fra?

Der er kun fem organisationer, der har svaret på dette spørgsmål, og de nævner alle Danmarks Statistiks ministerieordning, men i spørgsmålene om de enkelte populationer, har de fleste svaret data fra (underordnede) institutioner indenfor ressortområdet.

Hvordan er data struktureret?

Der synes ikke at være nogen tendens vedrørende de måder organisationerne har struktureret data, hvorfor det vurderes, at der ikke er en nemme, tekniske løsninger, der kan bidrage til standardisering på tværs af de forskellige datavarehusløsninger.

Se Bilag_grundoplysninger med åbne besvarelser om hvordan data er struktureret

Delkonklusion 1

Datavarehusene anvendes hovedsageligt til ledelsesinformation og statistik, men derudover er der stor forskel på, hvilke tekniske løsninger der er valgt, samt hvordan data stilles til rådighed for offentligheden.

2. Oplysning om centrale populationer



Svarene er meget uegale, hvorfor direkte sammenligninger, ikke vil være retvisende, de centrale populationer beskrives således:

Organisation	Population1	Population2	Population3
Arbejdstilsynet	Anmeldte arbejdsulykker	Anmeldte erhvervssygdomme	-
KL	Danske kommuner (borgere på tværs af 7 dataområder, Økonomi, Borgere, Personale, Skole, Ældre, Voksne handicappede og Udsatte børn og Unge)	-	-
Styrelsen for Forskning og Uddannelse	Personer med tilgang til en uddannelse på en institution	Dimittender	Befolkningens uddannelses- og arbejdsmarkedsstatus
Integrationsministeriet, Integrationsafdelingen	indvandre og efterkommere, data om beskæftigelse, uddannelse, bolig, ydelser, tilbud etc.	Den øvrige befolkning	-
Trafik-, Bygge- og Boligstyrelsen	Tidsserier ang. Bygninger og Boliger	Tidsserier ang. regnskaber for almene boligorganisationer	Data vedr. støttet byggeri mv.
Erhvervsstyrelsen	Virksomheder	-	-
Styrelsen for IT og læring	Grundskoleområdet, elever og institutioner	Gymnasieområdet, elever og institutioner	.
STAR - Styrelsen for Arbejdsmarked og Rekruttering	Modtagere af danske offentlige forsørgelsesydelse	Virksomheder i Danmark (Cvr, e-inkomst)	-
Danmarks Statistik	Totalpopulation af personer og virksomheder i Danmark	-	-
Sundhedsdatastyrelsen	Nationale sundhedsdata	-	-
RKKP	Sygdomsspecifikke registre - kliniske kvalitetsdatabaser	-	-
Rigsarkivet	Metadata om langt størstedelen af Rigsarkivets samlinger af både papirarkivalier og digitalt skabte data (registre) samt link til skannede papirarkivalier	Transkriberinger af papirarkivalier, ex lægdsruller, ansøgninger om erindringsmedaljer om de slesvigske krige mv.	-

Hvor kommer data fra

Mange organisationer anvender bl.a. data fra Danmarks Statistik i deres datavarehus, men alle modtager også data fra f.eks. sagsbehandling i egne institutioner.

Hvilke koder og hovedområder registreres der om populationerne

Navn på organisation, styrelse eller lignende	Koder registreret om populationerne	Hovedemneområder
Arbejdstilsynet	Regionsnummer Kommunenummer CPR nummer CVR nummer Geografisk placering	Arbejde og beskæftigelse Sundhed
KL	Kommunenummer	Uddannelse Økonomi
Styrelsen for IT og læring	Regionsnummer Kommunenummer CPR nummer Geografisk placering CVR nummer	Uddannelse
Integrationsministeriet, Integrationsafdelingen	Kommunenummer Geografisk placering Bestemte boligområder	Arbejde og beskæftigelse Bolig- og lokalplanlægning Demografi Retsvæsen, kriminalitet og retspleje Samfund og kultur Socialpolitik og velfærdssystemer Uddannelse Økonomi

Navn på organisation, styrelse eller lignende	Koder registreret om populationerne	Hovedemneområder
Rigsarkivet	Ingen	Arbejde og beskæftigelse Bolig- og lokalplanlægning Demografi Handel, industri og marked Historie Information og kommunikation Natur og miljø Politik Psykologi Retsvæsen, kriminalitet og retspleje Samfund og kultur Social stratifikation Socialpolitik og velfærdssystemer Sundhed Transport, rejse og mobilitet Uddannelse Videnskab og teknologi Økonomi
STAR - Styrelsen for Arbejdsmarked og Rekruttering	Regionsnummer Kommunenummer CPR nummer CVR nummer SE nummer Geografisk placering	
Sundhedsdatastyrelsen	Regionsnummer Kommunenummer CPR nummer	Sundhed
Styrelsen for Forskning og Uddannelse	Utallige	Arbejde og beskæftigelse Bolig- og lokalplanlægning Demografi Handel, industri og marked Samfund og kultur Socialpolitik og velfærdssystemer Uddannelse Videnskab og teknologi Økonomi

Navn på organisation, styrelse eller lignende	Koder registreret om populationerne	Hovedemneområder
Trafik-, Bygge- og Boligstyrelsen	Regionsnummer Kommunenummer Matrikelnummer BBR nummer CVR nummer Geografisk placering Landsbyggefondens organisations- og afdelingsnumre BOSSID (kode for det enkelte projekt)	Bolig- og lokalplanlægning Økonomi
Danmarks Statistik	Regionsnummer Kommunenummer Matrikelnummer BBR nummer CPR nummer CVR nummer SE nummer Geografisk placering	Arbejde og beskæftigelse Bolig- og lokalplanlægning Demografi Handel, industri og marked Historie Information og kommunikation Natur og miljø Politik Psykologi Retsvæsen, kriminalitet og retspleje Samfund og kultur Social stratifikation Socialpolitik og velfærdssystemer Sundhed Transport, rejse og mobilitet Uddannelse Videnskab og teknologi Økonomi
Erhvervsstyrelsen	Regionsnummer Kommunenummer CVR nummer	Ingen af de ovenstående
Regionernes Kliniske Kvalitetsprogram	CPR nummer	Sundhed

Delkonklusion 2

Oversigten over populationer i datavarehuse bør kunne give forskere og analytikere et overblik over hvilke oplysninger, organisationerne har registreret i deres datavarehuse.

Det er bemærkelsesværdigt, at der ikke er nogen statslige datavarehuse med oplysninger på det sociale område. KLs datavarehus har kun data på kommuneniveau, ikke på individniveau, hvorfor dette ikke kan supplere det sociale område.

Mange anvender data fra Danmarks Statistiks ministerordning, hvorfor dette kan være et grundlægt for bestræbelser vedr. harmonisering.

For udførlig liste med oplysninger om de forskellige populationer indeholder, se bilag_populationer

3. Muligheder for adgang til data

Kun seks organisationer udleverer data til forskere:

Udleverer data til forskere	Ingen praksis for udlevering af data til forskere
STAR - Styrelsen for Arbejdsmarked og Rekruttering	Styrelsen for Forskning og Uddannelse
Styrelsen for IT og Læring	Trafik-, Bygge- og Boligstyrelsen
Erhvervsstyrelsen	IM, Integrationsafdelingen
Danmarks Statistik	Arbejdstilsynet
Sundhedsdatastyrelsen	NFA
RKKP	KL
	Rigsarkivet

De seks organisationer, der udleverer data til forskning, har alle en relativt formaliseret procedure for, hvordan forskerne kan få adgang til eller udleveret data. Dataudlevering er professionaliseret og let tilgængelig.

De syv organisationer, der ikke har en praksis for at udlevere data til forskere, svarer nej eller ved ikke til alle spørgsmål vedr. udlevering af data, hvorfor det må formodes, at organisationerne ikke har taget stilling til dette, eller det ikke er relevant, fordi forskerne enten ikke henvender sig desangående eller fordi data kan tilgås andre steder - f.eks. ved Danmarks Statistik.

Delkonklusion 3

Det er enkelte større organisationer, der udleverer data til forskere på anmodning. Der er muligvis et uudnyttet potentiale for forskere og analytikere, hvis der træffes aftaler om udlevering af data fra datavarehuse i organisationer, der ikke har nogen praksis for dette. Relevante forskningsmiljøer bør tage stilling til dette, så det her ud fra kan vurderes, om der er behov for at gå videre med dette.

4. Datadokumentation



Spørgeskemaet indeholdt flere spørgsmål om datadokumentation, harmonisering etc. Under hensyn til de få besvarelser og organisationernes beskedne fokus på dette, er de åbne svar samlet i nedenstående tabel:

Datadokumentation Organisation	Retningslinjer for dokumentation	Dokumentationsstandard	Initiativer vedr. samarbejde om harmonisering	Samarbejder vedr. harmonisering?
Sundhedsdatastyrelsen	Dokumentationen kan findes på esundhed.dk	Mest interne standarder	Vi indgår bl.a. i diverse internationale samarbejder	Nej
Styrelsen for IT og læring	[dokumentation kan eftersendes]	[dokumentation kan eftersendes]	Nej	Danmarks Statistik
Styrelsen for Arbejdsmarked og Rekruttering	Der hvor vi udstiller data (fx vores statistikbank) er der fast standard for dokumentation	Vi er igang med at udarbejde ensartede retningslinjer for dette område mhp ensartet implementering i 2018	Nej	Nej
Styrelsen for Forskning og Uddannelse	En formaliseret praksis snarere end egentlige retningslinier. Nøgleord=konsistens!	Egenudviklet	Mangeårigt samarbejde med UVM og DST om uddannelsesstatistikken	I et vist omfang med fx EM og OIM
Trafik-, Bygge- og Boligstyrelsen	Nej	Nej	Nej	Nej
Udlændinge- og Integrationsministeriet, Integrationsafdelingen	Er som udgangspunkt dokumenteret af DST. Ellers egen dokumentation	Nej	Med øvrige ministerier	Med øvrige ministerier og Danmarks Statistik
Danmarks Statistik	se. WWW.dst.dk	flere forskellige	Primært andre statistikbureauer i EU og FN	Utal af arbejdsgrupper, kommissioner og taskforces
Arbejdstilsynet	Dokumentation er struktureret efter interne retningslinier	Nej	med AES	Nej
KL	Kommunerne har adgang til et dokumentationsbibliotek som sikrer transparens ift. data, datakilder og behandling ifb. mapping og nøgletalsproduktion	Nej, ikke så vidt vides.	Der er løbende dialog med relevante ministerier og Danmarks Statistik i opgørelsesmetoder, herunder nøgletalsafgræsning	jf. ovenstående

**KOR**Det Koordinerende Organ
for Registerforligning

Organisation	Retningslinjer for dokumentation	Dokumentationsstandard	Initiativer vedr. samarbejde om harmonisering	Samarbejder vedr. harmonisering?
Kulturministeriet	Interne retningslinjer for registrering	Egenudviklet	Der har været initiativer til samarbejde med andre offentlige og private arkiver i Danmark, men de ligger vist nok stille,	Nej
Erhvervsstyrelsen	Nej	Nej	Nej	Nej
RKKP	Registreringsvejledninger, datadefinitioner	Delvis. De kliniske kvalitetsdabasers indhold skal dokumenteres i en fælles (online) skabelon. Skabelonen indeholder en populationsafgrænsning, beregningsregler, en generel angivelse af datakilder og liste med variable og deres anvendelse.	Der er planer om igangsættelse af et begrebs- og datamodelleringsarbejde for de kliniske kvalitetsdatabaser og klinisk standardisering af indholdet af de kliniske kvalitetsdabaser. Tiltagene har bl.a. til formål at opnå semantiske interoperabilitet, som i højere grad end nu vil øge mulighederne for at genanvende data fra de regionale sundheds it-systemer, forbedre analysemulighederne på tværs af de kliniske kvalitetsdatabaser	RKKP kommer til at indgå i et tæt samarbejde med Sundhedsdatastyrelsen og den regionale "forretning".

Delkonklusion 4

Der er overordnet set ikke det store fokus på dokumentationsstandarder eller harmonisering med øvrige organisationer bortset fra på sundhedsområdet. Eventuel harmonisering er sket i forbindelse med samarbejde med Danmarks Statistik. Såfremt organisationerne kan se et behov for at udveksle data og/eller erfaringer på tværs af deres ressortområder eller ønsker at gøre det lettere for forskere, analytikere og andre at tilgå data, kan der etableres et samarbejdsforum om dette.