

**Nya insamlingsmetoder  
för transportstatistiken  
– erfarenheter från  
Trafikanalys  
utvecklingsprojekt**

**PM  
2018:7**



**Nya insamlingsmetoder  
för transportstatistiken  
– erfarenheter från  
Trafikanalys  
utvecklingsprojekt**

**PM  
2018:7**

**Trafikanalys**

Adress: Torsgatan 30

113 21 Stockholm

Telefon: 010 414 42 00

Fax: 010 414 42 10

E-post: [trafikanalys@trafa.se](mailto:trafikanalys@trafa.se)

Webbadress: [www.trafa.se](http://www.trafa.se)

Ansvarig utgivare: Brita Saxton

Publiceringsdatum: 2018-10-02

# Förord

Trafikanalys ansvarar för produktion av Sveriges officiella statistik om transporter och kommunikationer. I en värld där tiden för att besvara enkäter och intervjuer är begränsad och där mängden digital data som hanteras växer lavinartat, är det viktigt att fråga sig om det finns möjlighet att samla in data med nya metoder. I denna promemoria har erfarenheterna från några av Trafikanalys utvecklingsprojekt på datainsamlingsområdet sammanställts.

Sammanställningen är gjord av Sara Berntsson i samarbete med flera medarbetare vid Trafikanalys.

Stockholm i oktober 2018

Per-Åke Vikman  
Avdelningschef

# Innehåll

<b>Förord</b> .....	<b>3</b>
<b>Bakgrund</b> .....	<b>5</b>
<b>1 Lätta lastbilar</b> .....	<b>7</b>
1.1 Lätta lastbilar i Sverige.....	7
1.2 Lätta lastbilar i Nederländerna .....	8
<b>2 Lastbilsundersökningen</b> .....	<b>11</b>
2.1 Lastbilsundersökningen i Sverige .....	11
2.2 Utvecklingsprojekt för lastbilsundersökningen i Norge .....	11
2.3 Lastbilsundersökningen i Nederländerna.....	12
2.4 Mobilapp i Danmarks lastbilsundersökning.....	15
<b>3 Resvaneundersökningen</b> .....	<b>17</b>
3.1 Utmaningar kopplat till traditionella resvaneundersökningar.....	17
3.2 Resvaneundersökningen i Sverige .....	17
3.3 Resvaneundersökningen i Nederländerna.....	20
<b>4 Big data</b> .....	<b>23</b>
4.1 Big data internationellt.....	23
4.2 Big data inom officiell statistik .....	24
4.3 AIS-data .....	24
4.4 Tidtabellsdata.....	27
<b>5 Slutsatser</b> .....	<b>29</b>
5.1 Specifika observationer.....	29

# Bakgrund

Trafikanalys har de senaste åren sett över insamlingsmetoder i flera statistikprodukter och flera nya utvecklingsprojekt har genomförts. Utöver detta har användning av Big data bland annat för officiell statistik bevakats i olika sammanhang och konferenser. Syftet med detta PM är att sammanfatta några av de projekt som genomförts för att utveckla nya insamlingsmetoder för våra statistikprodukter. Vi har även studerat metoder hos andra statistikansvariga myndigheter utomlands, samt dokumenterat exempel på hur Big data använts för statistikproduktion. Fokus har varit på att sammanställa information som berör och kanske skulle kunna hjälpa till att utveckla Resvaneundersökningen (RVU), lastbilsundersökningen, användningen av AIS-data<sup>1</sup> för sjöfartsstatistiken, samt ge inspiration till ny insamling av statistik om lätta lastbilar.

Trafikanalys har i Resvane- och lastbilsundersökningen på olika sätt påbörjat en översyn av befintliga insamlingsmetoder. Dels för att minska uppgiftslämnarbördan, dels för att få högre kvalitet i insamlade uppgifter. Inom sjöfartstatistiken har metoder utvecklats för att på sikt kunna använda Big data i form av AIS-data. Detta skulle kunna bidra till kvalitetshöjningar av befintlig sjötrafikstatistik, sänkta produktionskostnader och framställning av helt ny och relevant statistik och samtidigt minskad uppgiftslämnarbörda för hamnarna.

Flera av EU:s medlemsländer har kommit långt inom elektronisk insamling och nyttjande av befintliga data för statistikändamål. Bland annat Nederländerna har kommit långt både vad gäller nya insamlingsmetoder för lastbilsundersökningen och nyttjande av AIS-data. De nordiska länderna står inför likartade utmaningar som Sverige när det gäller datainsamling och utveckling av nya insamlingsmetoder. Vi har löpande kontakt med framför allt Norge och Danmark kring statistikfrågor och utbyte av erfarenheter. Vi har också kontakt genom det Nordiska nätverket för transportstatistik, samt vid internationella möten och konferenser. Några av de projekt som är relevanta för vår verksamhet beskrivs i denna PM.

För att dra lärdom av projekt som genomförts hos andra länder gjordes ett studiebesök hos CBS, Nederländernas Statistiska Centralbyrå, i staden Heerlen under april 2017. Trafikanalys fick vid detta tillfälle att ta del av presentationer och diskutera ett urval av deras utvecklingsprojekt. Främst diskuterades lastbilsundersökningen, lätta lastbilsundersökningen, resvaneundersökningen, fordonsregistret, deras Big Data-projekt för AIS, samt ett projekt med vägtrafikkameror.

Under 2017 besöktes även två internationella konferenser, NTTS i Bryssel och TRB i Washington. Smått och gott som hörts på dessa, främst gällande Big data, finns även sammanställda i detta PM tillsammans med användbara länkar som stötts på längs vägen.

Det viktiga i detta projekt har varit aktiviteterna, kunskapsbildningen och dokumentationen av detta.

---

<sup>1</sup> AIS står för Automatic Identification System.





# 1 Lätta lastbilar

Antalet lätta lastbilar har ökat kraftigt de senaste decennierna i många länder, däribland Sverige, men få länder har genomfört undersökningar om deras verksamheter och godstransporter. Det finns ett stort intresse både nationellt och internationellt kring statistik om lätta lastbilars godstransporter, men förordning eller rekommendation om insamling saknas.

Internationellt har DG Move vid bland annat mötet WG Road freight statistics framfört behov av mer kunskap om de lätta lastbilarnas godstransporter. Under CGST 2017 visade flera länder intresse för att delta i en Taskforce om "Light Utility Vehicles". Ämnet diskuterades vidare under WG Road i juni 2018. Det beslutades där att några frivilliga länder (SE, NO, DK, DE, IT och IE) gemensamt formulerar mål och syfte med en Task force kring utvecklingen av statistiken om lätta lastbilar vid ett första möte hösten 2018.

I Sverige utmynnade intresset för mer kunskap om lätta lastbilar i ett regeringsuppdrag till Trafikanalys. Enligt uppdraget (som avrapporterades i januari 2018) skulle vi inventera behov av statistik om lätta lastbilar och distributionsfordon i urbana miljöer samt möjliga källor och modellverktyg för att producera statistik på området. I uppdraget ingick också en intressentanalys. Erfarenheter från olika länders statistikinsamlingar om lätta lastbilar finns presenterade i de två slutrapporterna till uppdraget.<sup>2</sup>

Problemen med att undersöka de lätta lastbilarna består bland annat av att lastbilarna är många till antalet och att de ofta gör många korta transporter dagligen. Att skicka ut en enkätundersökning till dem motsvarande lastbilsundersökningens veckolånga undersökning skulle belasta uppgiftslämnarna kraftigt. Istället bör man sikta på att utveckla möjligheterna att kunna använda registerdata eller andra befintliga datakällor som t.ex. trängselskattepassager.

## 1.1 Lätta lastbilar i Sverige

I de två regeringsuppdrag om lätta lastbilar som nämns ovan, redovisas ett antal förslag till upplägg för statistikproduktion, på kort och lång sikt, samt förslag till kunskapsunderlag baserat på registerdata. Dessutom kartläggs vilka övriga länder som genomfört lätta lastbilsundersökningar, kostnader och metoder för dessa m.m. I en underlagsrapport presenteras dessutom ett undersökningsupplägg för urvalsundersökning av de lätta lastbilarna som Statisticon AB genomfört på uppdrag av Trafikanalys.<sup>3</sup>

Sammanfattningsvis konstateras att tidigare enkätundersökning av lätta lastbilar i Sverige har varit förknippad med en omfattande uppgiftslämnarbörda. Den pilotundersökning som

---

<sup>2</sup> Trafikanalys 2018, Rapport 2017:21, Inventering av datakällor om lätta lastbilars transporter i urbana miljöer. [https://www.trafa.se/globalassets/rapporter/2017/rapport-2017\\_21-inventering-av-datakallor-om-latta-lastbilars-transporter-i-urbana-miljoer.pdf](https://www.trafa.se/globalassets/rapporter/2017/rapport-2017_21-inventering-av-datakallor-om-latta-lastbilars-transporter-i-urbana-miljoer.pdf)

Trafikanalys 2018, Rapport 2018:3, Hur förbättrar vi kunskapen om godstransporter med lätta lastbilar? [https://www.trafa.se/globalassets/rapporter/2018/rapport-2018\\_3-hur-forbattar-vi-kunskapen-om-godstransporter-med-latta-lastbilar.pdf](https://www.trafa.se/globalassets/rapporter/2018/rapport-2018_3-hur-forbattar-vi-kunskapen-om-godstransporter-med-latta-lastbilar.pdf)

<sup>3</sup> Statisticon 2018.

[https://www.trafa.se/globalassets/rapporter/underlagsrapporter/2018/rapport\\_01v02\\_undersokningsupplagg\\_llb\\_u.pdf](https://www.trafa.se/globalassets/rapporter/underlagsrapporter/2018/rapport_01v02_undersokningsupplagg_llb_u.pdf)

genomfördes 2012 hade låg tillförlitlighet. Samtidigt har de sammanställningar som enbart baseras på register (fordonsregistret) om lätta och tunga lastbilar sina begränsningar, eftersom de inte ger någon information om vilken verksamhet fordonen används i eller transporterade godsmängder. Under 2016 genomfördes ett pilotprojekt inom ramen för Varuflödesundersökningen som visade att det är möjligt att hämta vissa uppgifter i företagens administrativa system. För detta behöver dock nya insamlingsystem utvecklas, vilket tar både tid och pengar i anspråk.

Flödesmätningar görs i de flesta kommuner idag med skiftande metoder och kvalitet. Genom metodutveckling för hur insamling och bearbetning av data om flöden kan genomföras bör det finnas möjlighet att producera ny kunskap på sikt.

Trafikanalys föreslår i regeringsuppdraget att en ny statistikprodukt om lätta lastbilar publiceras samt att kunskapsunderlag/nulägesanalyser kontinuerligt tas fram om lätta lastbilar samt övriga distributionsfordons (inkl. tunga fordons) transporter i urbana miljöer. Uppgifterna bör i ett första skede baseras på lättåtkomlig information från fordonsregistret. Uppgifter från en urvalsundersökning riktad till lätta lastbilar (exkl. privatregistrerade lastbilar) kan komplettera redovisningen i ett urbant perspektiv. På längre sikt föreslås statistiken byggas ut både vad gäller detaljeringsgrad och omfång med hjälp av insamling av uppgifter med enkät och/eller från företagens administrativa system tillsammans med flödesuppgifter från kommunala mätningar. På längre sikt, för att öka kunskapen om urbana godstransporter och att inkludera nya datakällor, krävs insatser från flera håll. Arbetet kan innebära stora kostnader både i form av datainsamling och uppgiftslämnarbörda som är svårt att uppskatta. En lämplig avvägning måste göras mellan dagens och framtida tekniker i förhållande till identifierade krav på redovisning.

Ett första steg mot en ny statistikprodukt är att vi under 2018 påbörjar arbetet med en registerbaserad statistikprodukt om lätta lastbilar. Samtidigt pågår i dagsläget en upphandling av en provundersökning och ett undersökningsupplägg för statistik om lätta lastbilar. I uppdraget kommer det att ingå en bedömning av uppgiftslämnarbördan och om efterfrågade uppgifter finns lättillgängliga eller inte.

## 1.2 Lätta lastbilar i Nederländerna

Nederländernas statistikbyrå, CBS, har genomfört två undersökningar av lätta lastbilar de senaste åren. Dels en pilotundersökning 2012 som aldrig publicerades på grund av bristande kvalitet, dels en pilotundersökning med förenklad metod vars resultat publicerades 2016.

Resultaten från pilotundersökningen 2012 var bristande bland annat på grund av små urval, låg svarsfrekvens, och att många svarade att lastbilen var såld eller inte användes under mätperioden. Den totala kostnaden för pilotundersökningen var ungefär 175 000 Euro, varav huvuddelen av kostnaden avsåg utskick av enkäter. Kostnaden kunde hållas nere tack vare att produktionssystemen från lastbilsundersökningen kunde nyttjas.

Enkätundersökningen som genomfördes 2016 skickades till ett representativt urval av 38 000 lätta lastbilar med totalt 32 000 ägare. Endast lätta lastbilar ägda av företag undersöktes och specialfordon såsom ambulanser uteslöts. Urvalet stratifierades efter bransch (SNI-kod för huvudgrupp) och ålder på fordonet (ingen stratifiering skedde efter geografi). Enkäten innehöll frågor om de lätta lastbilarnas användning, när de används, var de körs, godsmängd och typ av gods.

Enkätundersökningen 2016 innebar en betydande förenkling för respondenterna jämfört med den föregående undersökningen. Enkäten gjordes så enkel och lättförståelig som möjligt och målet var att den inte skulle ta mer än 15–20 minuter att besvara. Stor vikt lades vid frågornas formuleringar, att tala "respondentens språk" och att inte ha onödigt många frågor. Uppgiftslämnandet var frivilligt. Mätperioden var 3 dagar och även fredag–söndag täcktes in (att jämföra med mätperioden för de tunga lastbilarna som är 7 dagar) och mätperioderna slumpades ut. Små företag (<30 fordon) fick enkäten endast en gång under året. Mycket energi lades också på att förklara nyttan med statistiken.

Till lätta lastbilar räknas i Nederländerna de lastbilar som har mindre än två tons maxlastkapacitet. (i Sverige utgör denna typ av fordon nästan 560 800, dvs. 88 % av alla lastbilar). De senaste åren har antalet lätta lastbilar ökat kraftigt i antal i Nederländerna, precis som i många andra länder. Antalet transporter för mindre leveranser såsom postpaket, blommor, frukt, paket är högre än någonsin. E-handeln är också högre än någonsin. Antalet webbutiker bedöms som högre än någonsin (källa här oklar). En hypotes de har i Nederländerna är att det pågår en trafikomställning (modal-shift) mot större andel lätta lastbilar i relation till tunga lastbilar. (I Sverige har andel lätta lastbilar de senaste 10 åren utgjort mellan 84–87 % av alla lastbilar i trafik och mellan 84–88 % av de nyregistrerade lastbilarna).

Vid pilotundersökningen av de lätta lastbilarna 2012 gjordes ett urval på 10 315 bilar men enbart 1 680 (16 procent) svarade med transportdata. Många svarade att lastbilen inte användes ("not in use") eller att lastbilen var såld. I 2016 års undersökning var populationen av lätta lastbilar under 2 tons lastkapacitet 793 886 stycken. Urvalet bestod av 38 458 fordon. Antal svar med resor var 8 500 (alltså fem gånger högre än 2012 års undersökning). Populationen delades in i 32 strata efter fordonens bransch och ålder.

Enbart webbenkät användes, respondenten var tvungen att ringa in och be om pappersenkät om de ville svara på papper, vilket väldigt få gjorde. Telefonpåminnelser gjordes inte eftersom det var för dyrt, enbart tre påminnelser via brev skickades ut. Svarefrekvensen var 45 procent. För projektet och undersökningen som genomfördes 2016 fick de pengar och finansiering från Ministry of transport samt från EVO och TLN som är transportorganisationer.

Det är inte uppgiftslämnarplikt på undersökningen om lätta lastbilar eftersom det blir för hög uppgiftslämnarbörda på respondenterna. De flesta som kör lätta lastbilar är egenföretagare med t.ex. serviceyrken såsom elektriker som kör tvättmaskiner, snickare som kör kök/träprodukter och dylikt. De har en egen lastbil de kör med och ofta inte tid för administration. Ska de fylla i en enkät görs det troligtvis på kvällar och helger.

Respondenten fick förifyllda uppgifter om dess "base location" dvs. den adress där lastbilen var registrerad. Denna adress kan enkelt ändras av respondenten om den inte stämmer. Därefter skulle fem lastningsställen anges, även antal stopp (vid distributions- och uppsamlingsrundor) och trafik i utlandet räknas med. Algoritmer hanterar och kalibrerar informationen om körda kilometer, start och stopp efter hur de fem resorna gjordes från "base location" för att minimera bördan. Även antal stopp, antal resor, transporterad godsvikt (i godsvikten ska inte verktyg eller utrustning som hör till chauffören eller bilen ingå), antal km totalt, antal km tomkörningar totalt, antal km utomlands. Dessutom efterfrågades mätarställningen vid start och slut. I enkäten efterfrågas vad som är gods i ekonomisk mening, man ska alltså inte räkna egna verktyg som transporterat gods.

De lätta lastbilarna har många användningsområden. Det skapades fem bransch kategorier som var förtryckta val i enkäten vid insamlingen; bygg, service, transport av varor, post och paket samt privata transporter/persontransporter. Kategorin "övrigt" som ingick i 2012 års

undersökning utgick 2016 med motivet att respondenterna själva kan kontakta statistikbyrån för vägledning om de behöver.

De viktigaste variablerna i undersökningen var:

- "Base location"
- Lastningsplats
- Antal stopp
- Antal resor
- Transporterad godsvikt (exklusive verktyg, utrustning etcetera mm)
- Antal km totalt
- Antal km tomt totalt
- Antal km utomlands

CBS säger sig gärna ge Trafikanalys tillgång till mer metoddokumentation<sup>4</sup>, enkäter m.m. från undersökningen, de anser att ju fler länder som använder samma metod, desto mer jämförbar blir statistiken och det blir "stronger position for funding".

---

<sup>4</sup> En av metoddokumenterna för undersökningen är denna:  
<https://repository.tudelft.nl/view/tno/uuid%3A09dc598b-286f-487f-82bc-f79c8e2fd67b>

## 2 Lastbilsundersökningen

Lastbilsundersökningen styrs av Europaparlamentets och rådets förordning (EU) nr 70/2012 om statistikrapportering om varutransporter på väg. Mängden uppgifter som ska samlas in i undersökningen är omfattande och flera länder har genomfört olika projekt för att minska uppgiftslämnarbördan för respondenterna, dvs. lastbilsägarna. Nedan beskrivs kort vad som pågår i Sverige, Norge, Danmark och Nederländerna kring detta.

### 2.1 Lastbilsundersökningen i Sverige

Ett flertal utvecklingsprojekt har gjorts inom ramen för lastbilsundersökningen de senaste åren för att undersöka möjligheterna att minska uppgiftslämnarbördan.

I lastbilsundersökningen är referensperioden en vecka, det vill säga en uppgiftslämnare ombeds redovisa samtliga sändningar som genomförts under en specifik mätvecka. Under 2015 fick Statisticon AB i uppdrag att genomföra en studie för att utreda om det är möjligt att uppnå precisionskraven med en kortare referensperiod; tre eller fyra dagar. Dessutom att undersöka om en kortare referensperiod är lämplig för hela urvalet eller enbart för delar av urvalet. Resultatet av studien sammanfattas i att precisionen inte tycks försämrats med kortare referensperiod. Däremot har en kortare referensperiod inte någon påvisbar effekt på svarsandelen eller andelen rapporterat stillestånd. Från djupintervjuerna framgår att småföretag skulle föredra en kortare referensperiod. För storföretagen spelar det inte så stor roll. En rekommendation i projektets slutsatser är att se över stratifieringen och att även stratifiera efter företagsstorlek.

Från 2016 har dessutom Statisticon AB genomfört två uppdrag med syfte att kartlägga större systemleverantörer av kördatasystem, besöka några av dessa leverantörer samt att besöka åkerier för att kartlägga vilka uppgifter som finns tillgängliga i deras administrativa system för lastbilsundersökningens ändamål. Fyra leverantörer har intervjuats och exempeldata har erhållits från systemet K2 som ägs av Fleet101. Det mesta av den information som efterfrågas på sändningsnivå i lastbilsundersökningen återfinns i erhållna data. Data är dock inte komplett och vissa saker är oklara. Under våren 2018 har dessutom besök gjorts hos Scania's Transportlaboratorium med förhoppning om gemensamt framtida utvecklingsprojekt av system och överföring av relevanta uppgifter. Hos Scania finns systemutvecklare centralt och besked om eventuella framtida projekt ska ges till Trafikanalys hösten 2018.

### 2.2 Utvecklingsprojekt för lastbilsundersökningen i Norge

Som en del i strävan efter att minska uppgiftslämnarbördan ser Norges Statistiska centralbyrå, SSB, över möjligheten till elektronisk insamling av de uppgifter som krävs för lastbilsundersökningen. Gällande såväl uppgifter som finns tillgängliga hos lastbilsägare som uppgifter som finns i olika transportsystem. Under 2018 kommer de att delta i ett forsknings-

projekt, LIMCO, tillsammans med Transportøkonomisk institutt och en del andra partner såsom större näringslivsaktörer exempelvis SAP, Post Nord, Posten och Volmax (som levererar utrustning till Volvo och Renault). Det är ett större forskningsprojekt där SSB deltar i en del som avser ny datainsamling. Syftet med projektet är att använda data från inbyggda system i lastbilar kombinerat med data från ERP-system<sup>5</sup> och logistikprogramvara för att få en bild av hur effektiviteten kan öka och transportkostnaderna och utsläppen från godstransporter minska. Sekundära mål med projektet är att:

1. Förbättra metoderna för insamling, bearbetning och sammanslagning av data från flera digitala källor.
2. Undersöka hur nya typer av data, såsom data från inbyggda system i lastbilarna, kan bidra till effektivare rapportering till SSB.
3. Använda ny data för att ge kvantitativa uppskattningar av hur förändringar i logistiksystem och körbeteende påverkar transportkostnader och utsläpp.
4. Undersöka hur organisationen av logistikverksamhet kan ändras för att skapa ett effektivare och mer hållbart transportsystem i urbana områden och för långdistanstransporter.
5. Samla bättre kunskaper om godstransporters negativa yttre påverkan av transporter både lokalt och globalt.

I projektet ska SSB bidra med personer från datainsamlingsavdelningen, metodsektionen och de som jobbar med lastbilsundersökningen.

Dessutom hoppas SSB att snart få till en ny varuflödesundersökning, och att få igång aktiviteter med aktörer som är av intresse för lastbilsundersökningen. De konstaterar att det inte är lätt att med nya källor samla in de databehov som krävs för att uppfylla förordningens krav för lastbilsundersökningen.<sup>6</sup>

En tidigare rapport från Norges utvecklingsprojekt kring lastbilsdata finns publicerad.<sup>7</sup>

## 2.3 Lastbilsundersökningen i Nederländerna

Redan på 1950-talet började data till lastbilsundersökningen i Nederländerna samlas in. Förr var det två enkäter, en för firmabilstrafik ("own account") och en för yrkesmässig trafik (hire/reward). En första designförändring av undersökningen gjordes 1992, en andra 2010 då även en webbenkät implementerades. Undersökningen som helhet vidareutvecklades och förbättrades under 2013. Sedan dess har även möjligheten till överföring av lastbilsdata i XML-format (XML är strukturerade filer) utvecklats samt att vissa företag får möjlighet till skräddarsydda lösningar att lämna uppgifter via Excel eller på papper.

Drivkraften inför utvecklingsarbetet 2013–2015 var bland annat budgetnedskärningar, att minska uppgiftslämnarbördan, förbättra kvaliteten, maximera den elektroniska insamlingen samt att minska den partiella non-responser. Branschorganisationer involverades i arbetet

<sup>5</sup> ERP står för Enterprise Resource Planning och avser de system och program som företag använder för dagliga uppgifter som redovisning, inköp, projekthantering och tillverkning.

<sup>6</sup> Mejlkonversation mellan Sara Berntsson och Asbjörn Willy Wethal på SSB.

<sup>7</sup> SINTEF (2016) Rammeverk for nye og bedre godsdata - Elektronisk innrapportering og alternative datakilder <https://brage.bibsys.no/xmlui/handle/11250/2434606>

vilket bland annat resulterade i att de kommunicerade mer positivt om betydelsen av lastbilsundersökningen med sina medlemmar.

I Nederländerna finns omkring 132 000 lastbilar registrerade som ingår i målpopulationen. Urvalsstorleken 2016 var 36 000 lastbilar (600–800 fordon per vecka). Stratifiering görs genom bland annat näringsgrensindelning, NACE, maxmilastvikt samt ålder (+/- 4 år) i 37x2 strata.

Datakvaliteten i lastbilsundersökningen har förbättrats betydligt sedan undersökningen gjordes om inför 2015 års undersökning. Förenklingsarbetet innebar bland annat att frågorna sågs över och omformulerades, dessutom förbättrades enkätens utseende. De arbetar nu mer aktivt med att i enkäter och missiv tala respondenternas "språk", att ha samma "jargong" som respondenterna samt att i största möjliga mån hålla enkäten enkel att förstå och att inte fråga fler frågor än nödvändigt. Påminnelsestrategin är att den baseras på dialog. De jobbar även med effektiviseringar, att minska pappersbördan, bland annat genom att maximera den elektroniska datainsamlingen samt att minska bortfallet. Tack vare förenklingsarbetet har svarsfrekvensen i enkäten ökat från 69 % till 89 % år 2015.

Exempel på genomförda förenklingsåtgärder är att det i webbenkäten är möjligt för respondenterna att skriva i fritext på varuslaget och lastnings- och lossningsort. En algoritm tar hand om detta och klassar till rätt varuslag/NUTS3 kod. Om det finns flera orter med samma namn dvs. om ortnamnen finns på olika ställen i landet, blir det problem. Det är därför bättre att använda postnummer. (Om elektroniskt kan även länet komma upp som komplement till orten). Algoritmer används för att geokoda t ex NUTS 3 där Latitude och longitude används i kombination. Ortnamn eller postkod tas fram genom ett ruttplaneringsprogram (kommersiell mjukvara) som ger ortsnamn och landskod, latitude/longitude, border crossings, distans inrikes och total distans. Optimal rutt beräknas. Tomkörningar baseras på logiska sekvenser av lossnings- och lastningsplatser. Varuslagen kodas genom en automatisk process med hjälp av en referensdatabas (NST2007 och NSTR). 90 % av varuslagen kodas automatiskt, nya beskrivningar kollas upp manuellt och läggs till databasen manuellt. Farligt gods kodas automatiskt från varubeskrivningen (t ex. bensin) till ADR, UN-nummer kan användas och konverteras till ADR.

Projektet med XML-överföringar började 2010. Nu har CBS löpande dataöverföringar från 93 företag. Syftet var att förenkla för uppgiftslämnarna och därigenom få bättre kvalitet på inlämnat data.

För att möjliggöra XML-överföringarna har fem mjukvaruföretag på egen bekostnad gjort skraddasydda lösningar för lastbilsundersökningen, där drivkraften har varit att tillhandahålla och sälja en tjänst till lastbilsägarna. Företagen fick fria händer att utforma tjänsten, statistikbyrån berättade bara vad som skulle ingå. Företagen var tveksamma först eftersom "the business case" var statistisk. Men CBS framhöll att även tid för att samla in och skicka data på traditionellt vis för att besvara enkäterna också kostar pengar för lastbilsägarna. Företagen uppmanades att ta fram tjänsten som en service till kunderna. Att de antingen kunde göra den gratis, eller låta kunderna betala för den, men att det skulle vara företagets egna investering. Även transportbranschen uppmanade medlemmar att använda denna XML-lösning. Med de företag/respondenter som vill ha XML-lösning görs en överenskommelse och det öppnas upp en digital port för uppladdningen. CBS uppskattar att det finns ca 100 mjukvaruföretag inriktade på lastbilsdata totalt, ingen är direkt dominerande.

I Nederländerna har de fältarbetare för de statistiska undersökningarna, för lastbilsundersökningen har de i dagsläget endast en person anställd, förr var de uppemot 10 personer. Denna fältarbetare besöker företag som inte besvarar undersökningen. Inför XML-införandet besöktes omkring 200 företag av fältarbetare som då informerade om systemet. Branschorganisationerna skrev artiklar i deras tidningar om undersökningen, och skapade ett intresse för XML. De uppmanade företagen/åkerierna att fråga sina mjukvaruleverantörer om XML möjligheten.

Vissa uppgifter är knepiga att samla in, den typ av uppgifter som kanske inte direkt är intressant för företagen att spara på och därför ofta saknas, såsom distanser, men CSB är flexibla, de ber då respondenter sätta 0 så fyller CBS i och räknar ut avstånd själva. Även godstyper är ofta okänt, då antar CBS vad det bör vara efter vilken typ av företag det är som kör godset. Detta är inte helt optimalt, men det ger en godtagbar lösning.

Företagen som levererar data via XML väljer själva om de vill lämna data som "sample", "random sample", eller all data året runt. Data sänds manuellt, som upload eller via ftp. Överföringen är helt automatiserad, data kommer in löpande från många företag. Det kommer alltså in ett stort överflöd av data, men bara det specifika urvalet av lastbilar för undersökningen kommer till användning för lastbilsundersökningen, resterande data används av CBS för forskningsändamål eller som referensjämförelse.

Utvecklingen av XML-systemet tog tid, de första 1–3 åren ägnades åt detta. År 2008 implementerades systemet och under 2010 kom den första användaren. De 93 företag som idag nyttjar tjänsten motsvarar 2 % av alla företag, 11 % av de större företagen, 8 % av hela undersökningens population, 10 % av alla transporter 20 % av alla sändningar i undersökningen.

På frågan till CBS om de skulle gjort något annorlunda i processen svarar de att de möjligtvis skulle ha jobbat mer med medfinansiering av utvecklandet av systemen. Samt att ge någon form av värde till transportföretagen som lämnar uppgifter, typ för dem användbar statistik.

De har testat GPS med lösa anordningar, men det var dyrt och tekniskt komplicerat. Smartphone-appar innebär hög belastning på respondenterna, det ska laddas ner en app och appen ska försöka tolka resorna. När det då blir fel måste respondenten själv redigera och kontrollera.

En reflektion vi hade efter studiebesöket hos CBS var att det även hos oss i Sverige kan vara intressant att testa den metod som de använde i dess lastbilsundersökning när det gäller att skriva i fritext på varuslaget, där en algoritm tar hand om detta och klassar till rätt ort. Om elektroniskt kan även länet komma upp som komplement till orten. Detta kanske kan vara något även för den svenska resvaneundersökningen.



## 2.4 Mobilapp i Danmarks lastbilsundersökning

I Danmarks lastbilsundersökning väljs 180 lastbilar ut varje vecka, av totalt 36 000 lastbilar. Insamling sker via webbenkät och sedan 2017 erbjuds även en app. Anledningen till att appen utvecklades är att all rapportering till myndigheter i Danmark ska ske digitalt. 2016 rapporterade fortfarande hälften av respondenterna på pappersenkäter, trots att webbenkäten funnits i 10 år. Företagen förklarar att det är billigare att låta chauffören fylla i pappersenkäten i lastbilen snarare än att använda kontorspersonal för detta, ofta är det dessutom enklare för chauffören att fylla i uppgifter om varuslag och exakt destination. Nackdelen är dock att pappersenkäten anges vara svår att fylla i. Jämförelser med färdskrivardata visar att det troligtvis förekommer undertäckning på grund av partiell underrapportering. Trots att pappersinsamlingen inte längre erbjuds är det fortfarande 10–15 respondenter per år som lämnar uppgifter på detta sätt av gammal vana.

Det varierar stort mellan företag på vilket sätt det är enklast att lämna uppgifterna till lastbilsundersökningen. Strategin har varit att ha några få standardiserade lösningar.

1. Mobilappen riktar sig främst till de små/medelstora företag där det ofta finns en ägare som själv kör lastbilen och inte har något administrativt stöd på kontoret.
2. Webbenkäten riktar sig till de som redan lämnar uppgifter på detta sätt, små/medelstora företag med administrativt stöd och god dokumentation.
3. Möjligheten till Uppladdningssystem med data från tex. fleet management systems ses över. Fokus ligger då på medium/stora företag. Implementeringskostnaden för detta blir troligtvis den största utmaningen.

Mobilappen är utvecklad för tre mobila plattformar, Android, iOS och Windows. Den är utvecklad hos Statistics Denmark och kan bara laddas ner från deras webbplats. Ett problem har varit Apple licensen hos Google, att hitta sätt för att försöka minska dataanvändningen och att undvika ständig GPS-reading. Statistikbyrån jobbar vidare med utveckling, förbättringar och användarvänlighet. Vid inloggningen anges företagets identifikationsnummer och fordonets registreringsnummer.

När det gäller hur multistop-resor registreras används ett "journey scheme" med förenklade antaganden och samma metod används för webbenkäten. Databasen byggdes efter webbenkätens struktur och appen följer samma struktur. Google används för avståndsberäkning. Denna distansberäkningsmodul är tydligen en del av ett stort paket och det var inte helt lätt för Danmark statistik att komma överens med Google kring bara denna modul utan att betala för hela paketet.

Trafikanalys har fått tillgång till länkar för att testa appen, även tillgång till källkoden för appen kan fås från Statistics Denmark vid behov.



## 3 Resvaneundersökningen

### 3.1 Utmaningar kopplat till traditionella resvaneundersökningar

För att kunna utforma ett framtida transportsystem som på ett hållbart sätt kan möta befintliga och kommande samhällsutmaningar, behövs utvecklad förståelse av hur och varför människor reser som de gör. I resvaneundersökningar är man intresserad av att förstå hur människor rör sig, det vill säga *transportdata*. En del datakällor ger dock information om hur fordon rör sig, det vill säga *trafikdata*. Hur transport- och trafikdata relaterar till varandra är olika, trafikdata för cykel kan i det närmaste sägas vara detsamma som transportdata, medan det för personbil och kollektivtrafik är stora skillnader.

Dagens metoder för resvaneundersökningar (RVU) har i vissa avseenden problem och brister, såsom,

- misstanke om bristande representativitet,
- sjunkande svarsfrekvens,
- brister i datakvalitet när det gäller reslängder och restider,
- brister i uppgifter om vägval,
- hög uppgiftslämnarbräda,
- långa mellanrum mellan mätningarna,
- och höga kostnader.

Problemen med traditionella metoder för att samla in resvaneundersökningsdata upplevs på många håll runt om i Europa och i resten av världen. Dessutom är det många som ser potentialer i nya sätt att samla in data. Detta gör att det pågår en del hel utveckling inom området både i Sverige och i resten av världen. Sedan 2014 arbetar en tillfällig arbetsgrupp<sup>8</sup> inom Eurostat för att harmonisera resvaneundersökningar i Europa. Bakgrunden är behovet att utveckla indikatorer för att följa upp EU:s vitbok om transporter från 2011.

### 3.2 Resvaneundersökningen i Sverige

I ett av regeringens fem strategiska samverkansprogram, som presenterades i juni 2016 "Nästa generations resor och transporter", identifierades ett behov av att finna och utveckla nya lösningar för resvaneundersökningar. Trafikanalys har lett detta arbete<sup>9</sup> och slutredovisning beräknas ske i september 2018.

---

<sup>8</sup> Task Force on Passenger Mobility.

<sup>9</sup> Övriga deltagare i projektet är Trivector, VTI, Sweco, Ericsson, Göteborgs stad, Lindholmens Science Park/Västra Götalandsregionen och Stockholms läns landsting Trafikförvaltningen. Som referensgruppsmedlemmar ingår Trafikverket, Nobina Technology, Samtrafiken och Linköpings Universitet.

Drivkraften för projektet har varit en kombination av sjunkande svarsfrekvenser, kostnader, behov av detaljeringsgrad för bland annat geografisk nivå samt att teknikutvecklingen gett nya möjligheter att följa resmönster. Syftet har varit att undersöka och testa nya tekniska lösningar för datainsamling, samt att undersöka hur olika datainsamlingsmetoder skulle kunna användas, enskilt eller i kombination, i framtida resvaneundersökningar.

Projektet har bestått av olika delar. Arbetet inleddes med en intressentanalys<sup>10</sup> för att kartlägga de behov av resvanor som finns att identifiera, samt frågeställningar som resvaneundersökningar ska kunna hantera och besvara. Resultaten visar att intressenterna har spridda behov och intressen avseende resvanedata.

Utöver intressentanalysen gjordes en kartläggning<sup>11</sup> av verktyg och applikationer för insamling av resvanor. Syftet var där att undersöka nya tekniska lösningar för datainsamling och hur dessa skulle kunna användas, enskilt eller i kombination, i framtida resvaneundersökningar, som underlag till statistik.

I kartläggningen identifierades ett 50-tal internationella verktyg och applikationer som kan samla in RVU-relevant data. De juridiska förutsättningarna för att samla in och använda/sälja personliga data skiljer sig mellan olika länder. Spelreglerna ändrades också när EU:s nya direktiv kring datasäkerhet (GDPR) trädde i kraft 25 maj 2018. Detta kan påverka vilka verktyg som kan användas i vilka länder, och vilka juridiska anpassningar som behöver göras. Det finns internationellt flera liknande produkter (t.ex. RVU-appar, vägsidesutrustningar, etc) som de svenska. Det är inte möjligt att entydigt peka ut vilka som är bäst, då för- och nackdelar beror på tillämpningar.

Kartläggningen skedde genom dokumentanalys, litteraturstudie samt kontakter med experter/utvecklare och en referensgrupp. Resultatet presenterades i en översikt av olika typer av trafik- och resandemätningar (datakällor/datainsamlingsmetoder) och en kategorisering av verktyg och applikationer för att samla in data för resvaneundersökningar – både Big och small data, passiv och aktiv insamling.

Utifrån resultaten i intressentanalysen och kartläggningen föreslogs och genomfördes tre pilotprojekt.

### **1) Jämförbarhet och representativitet med webbenkät och resvaneapp i mobiltelefon.**

Syftet med projektet var att göra en demonstrationsstudie som visar hur två datainsamlingsmetoder (webbenkät med karta och mobilapp) för resevaneundersökningsdata presterar i fält, samt hur de presterar jämfört med "traditionell" RVU (postal enkät och/eller telefonintervjuer). Projektet syftar också till att få bättre förståelse för hur olika rekryteringsmetoder kan användas för att framtida RVU:er ska kunna bli mer kostnadseffektiva.

Tre rekryteringsmetoder har använts för att testa insamlingsmetoderna för webbenkät med karta och mobilapp; rekrytering via slumpmässigt urval, webbpanel och crowdsourcing. Sammantaget visar studien att mobilappen registrerade betydligt fler delresor jämfört med webbenkäten medan reslängderna är jämförbara.

---

<sup>10</sup> Intressentanalysen genomfördes av VTI under våren 2017 på uppdrag av Trafikanalys och med medfinansiering från Trafikverket.  
[http://www.trafa.se/globalassets/rapporter/underlagsrapporter/2017/vti\\_intressent-och-behovsanalys-for-resvaneundersokningar.pdf](http://www.trafa.se/globalassets/rapporter/underlagsrapporter/2017/vti_intressent-och-behovsanalys-for-resvaneundersokningar.pdf)

<sup>11</sup> Trivector gjorde denna kartläggning på uppdrag av Trafikanalys.  
<http://www.trafa.se/globalassets/rapporter/underlagsrapporter/2017/detaljerad-kartlaggning-av-verktyg-och-applikationer-for-resvaneundersokningar.pdf>

Crowdsourcingen fick in flest antal svar jämfört med både slumpmässigt urval och webbpanel. Den var dessutom billigast att genomföra. Dock är det en problematisk metod att använda om man vill använda statistisk teori eftersom målpopulationen är okänd.

Vad gäller svarsfrekvens presterar de nya insamlingsmetoderna (webbenkät och mobilapp) sämre jämfört med traditionella metoder (telefonintervjuer och pappersenkät). Incitament bedöms bara ha en marginell positiv effekt på svarsfrekvensen. Det grundläggande problemet med låg svarsfrekvens är att risken ökar för att svaren inte är representativa för målpopulationen. Ett sätt att hantera att respondenterna inte motsvarar målpopulationen är att vikta svaren så svaren från de grupper som är överrepresenterade blir mindre betydelsefulla och svaren från de grupper som är underrepresenterade blir mer betydelsefulla.

**2) Plattform för integrering av data.** Sannolikt behövs i framtiden flera datakällor för att förstå människors resvanor. För att kunna dra nytta av olika typer av datakällor behöver de integreras. I en sådan plattform kan både traditionella datainsamlingsmetoder (till exempel "slangmätningar") och data från nya verktyg/metoder inkluderas.

Detta projekt har ännu inte genomförts.

### **3) Mobilnätsdata för OD-matriser och långväga resor.**

Syftet med pilotprojektet var att utföra en första utvärdering av möjligheter och utmaningar med att skatta resestatistik för långväga resor baserat på aggregerad och anonymiserad mobilnätsdata. Jämförelserna visar på potential att skapa statistik för långväga resande med hjälp av mobilnätsdata med hög tidsupplösning, exempelvis per dag.

Utöver dessa tre ovanstående projekt har ett projekt avseende data från register genomförts. I detta projekt fick SCB hösten 2017 i uppdrag av Trafikanalys att se över intervjufrågorna som använts till resvaneundersökningen 2011–2016, för att se vilka frågor som kan ersättas med administrativa data från befintliga register hos SCB. Fördelarna med att hämta data från administrativa register består bland annat i att minska uppgiftslämnarbördan, kortare frågeformulär och eventuellt få ökad svarsfrekvens. SCB:s undersökning har identifierat 36 bakgrundsfrågor i Trafikanalys tidigare frågeformulär som kan besvaras fullt ut med hjälp av administrativa register och sex frågor som delvis kan besvaras, bland annat på grund av att registren t.ex. endast har uppgifter om ägande och inte information om tillgång till fritidsfastighet. Ett annat exempel är att avståndet i km mellan bostad – jobb/ skola är uppmätt som fågelvägen snarare än som faktisk reslängd i transportnätet.

### **Slutsatser från RVU-projekten**

För att frågeställningar kring förståelse av mobilitetsval ska kunna besvaras krävs i många fall att datamaterialet innehåller information om individen och om ärendet. Denna typ av data är svår att samla in med många av de verktyg/metoder som utformats för andra syften än RVU, såsom mobilnätsdata, biljett- och betalssystem, mätning på fordon samt vägsidesutrustning. Därför behöver resvanedata oftast samlas in genom traditionella metoder, webbenkäter och/eller appar designade för RVU-data. Dock är dessa individfokuserade verktyg/metoder svåra att genomföra kontinuerligt och i tillräckligt stor skala för att kunna klara av alla de nedbrytningar som vissa frågeställningar kräver. Om man kan lyckas kombinera de två typerna av datakällor, så att intervjuer/webbenkäter/appar kan stå för förståelse på en individnivå och andra, större data, kan stå för totala resandemängder och aggregerade mönster skulle en mycket användbar RVU-databas kunna skapas.

Generellt kan konstateras att det verkar finnas en trade-off mellan stora datamängder och information om individen och resans ärende. Frågeställningarna är avgörande för var denna gräns går.

Data bör i framtiden kombineras på olika sätt för att spegla en helhetsbild av trafik och transporter, och för att svara på de olika frågeställningar som finns. T.ex. kan individdata ge en djup förståelse av hur olika grupper reser, och detta kan kombineras med kontinuerlig och passiv datainsamling som ger en bild av hur trafik och transporter förändras över tid. Snitt- eller platsmätningar på cykelstråk skulle kunna kombineras med data från låncykelsystem för att skapa en förståelse av resväg för cyklister, etc. Det finns många möjligheter, och de bör utgå ifrån de frågeställningar som finns att besvara.

Den data som samlas in med de nya metoderna finns oftast hos privata aktörer. För att en offentlig myndighet ska kunna använda sig av dessa data, måste det finnas kvalitetssäkring för de olika datakällorna. T.ex. för individdata genom bekräftelse från användaren, och för passiva data genom jämförelse med andra datakällor, eller med stickprov etc.

Juridiska aspekter behöver i de flesta fall också studeras närmre. För data som samlas in med andra syften än för RVU kan det behövas medgivande/informerande av de som lämnar data för att datamaterialet juridiskt sett ska kunna användas. Spelreglerna ändrades också något när EU:s nya direktiv kring datasäkerhet (GDPR) trädde i kraft den 25 maj 2018. Integritetsskyddet för individen stärktes, men det skapade också enhetliga regler inom hela EU, vilket kan underlätta för både import och export av verktyg/metoder inom EU. Det är dock oklart vad som gäller vid import av verktyg från utomeuropeiska länder, då juridiska aspekter behandlas på olika sätt i olika länder.

### 3.3 Resvaneundersökningen i Nederländerna

Sammanfattningsvis jobbar CBS med successiv förbättring av traditionella metoder för resvanestatistiken. De använder mer registerdata jämfört med tidigare, för t.ex. hushålls-sammansättningen. Eftersom de använder registerdata måste de informera om detta i följbrevet, men de tror inte att det påverkar svarsvilligheten. De tester som gjorts med registerdata för hushåll har fallit väl ut.

Skatteuppgifter används i de fall personen har deklarerat för skatteförmån för bil, istället för att fråga om tillgång till leasingbil finns. (I Sverige frågar vi i dagsläget efter om hushållet har leasingbil.)

Följbrevet har setts över och de funderar på riktade följbrev beroende på målgrupp och ålder. De har även omarbetat frågeformuläret. Nu utgår de från platserna där personen har varit istället för att fråga om resan och att respondenten ska berätta hur man reste mellan platserna. De säger sig ha lyckats bra med att genom mindre förändringar få upp svarsfrekvensen från 35 % till 60 % genom att ändra brev, slimma frågor, m.m.

Olika belöningar har testats för att öka svarsfrekvensen i resvaneundersökningen, varav ett av dem gav snedare urval. Människor med högre inkomst reagerade mer positivt på belöningar, låginkomsttagarna påverkades inte lika mycket. Exempelvis skickades fem euro ut tillsammans med enkäten, som presentkort. Presentkortet fick behållas oavsett om enkäten fylldes i eller inte. Ett annat test var ett lotteri om en Ipad, då var respondenten tvungen att svara på något sätt. Lotteriets vinnare var avgjorda redan vid undersökningens start för att förenkla vinnardragningen, redan när intervjuaren ringde upp kunde denne se om respondenten skulle vinna om denne besvarade enkäten. Intervjuaren kunde alltså se redan

från början om det var vinst eller inte om respondenten valde att svara. Flera Ipads lottades ut på detta sätt.

Data samlas in på tre olika sätt. Följebrevet innehåller inloggningsbeskrivning till webbenkäten. Webbenkäten ska fyllas i för en specifik dag. Det skickas ut två påminnelser, då flyttas svarsdatum framåt, har respondenten inte svarat ännu, ringer CBS om de hittar ett telefonnummer. De ringer nio gånger tills de får tag på någon. De utvärderar i dagsläget om verkligen är lönt att ringa så många gånger för att nå vissa grupper, eller om de bara ska skicka påminnelser till dem. Om inget telenummer hittas åker de hem till respondenten.

De planerar att skapa och använda "Google autolist fill" till webbenkäten för att underlätta ifyllande.

Vid telefonintervju och "face to face" intervjuer är det knappt 50 % svarsfrekvens, omkring 20 % svarar på webbenkäten av totalen. Vägrarna är dock inte inräknade i detta. Vägrarna utgör ca 4 %. 76 % går vidare till nästa steg. Den totala svarsfrekvensen är omkring 55 %.

Efter 2018 ska CBS bara använda webbenkät, det är för dyrt med andra metoder. Då räknar de med att svarsfrekvensen förmodligen sjunker till ca 20 %, men större stickprov kommer att dras i grupper där man vet att det är låg svarsfrekvens. Större stickprov bör ge högre kvalitet och lägre statistisk osäkerhet.





## 4 Big data

Big data, dvs. stora mängder digitalt lagrad information, har de senaste åren ökat i takt med möjligheten att samla in och lagra information från bland annat internet, telefoner, betalkort, m.m. Utmaningen med Big data handlar bland annat om att kunna hantera, strukturera och aggregera dessa stora datamängder så att de blir användbara. En annan utmaning med Big data är bristen på erfarenhet, kunskap och kompetens på att använda de extremt stora dataset som det ofta handlar om. En annan frågeställning kring stora dataset, ofta insamlade för helt andra syften än att producera statistik, är dessutom ofta äganderätten.

Det finns ingen formell definition av Big data men det handlar om stora mängder data som genereras snabbt och de kan ha en mängd olika format. UNECE, FN:s ekonomiska kommission för Europa, har identifierat tre olika typer av källor till Big data,

- sociala nätverk som till exempel Facebook, Twitter, bloggar, foton, videofilmer, Youtube och internetsökningar,
- traditionella företagssystem, till exempel transaktioner, kundregister, kreditkortsdata och medicinska register,
- Internet of Things, olika typer av sensordata, till exempel väder, trafik, övervakning, mobiltelefoner och satellitbilder.

Många av de nya datakällor som inkluderas i begreppet Big data kommer från specifika källor och innehåller endast några få variabler. Dessa behöver oftast kombineras med andra datakällor för att de ska ge intressant och ny information.<sup>12</sup>

### 4.1 Big data internationellt

#### Essnet Big data

Inom EU pågår bland annat ESSnet<sup>13</sup> nätverket. Målet är att utnyttja synergierna från samarbete mellan vissa medlemsstater för att dela kompetens och spara kostnader för att lösa gemensamma problem av europeiskt intresse. Överföringen av resultat och kunskap till icke-deltagande partner till förmån för hela ESS är också en viktig egenskap för ett ESSnet-projekt. En del av detta är ESSnet Big data, som är det största konsortiet någonsin, med 22 deltagande stater.

Syftet med ESSnet Big data är att integrera stora data i den regelbundna produktionen av officiell statistik, genom piloter som undersöker potentialen hos utvalda stora datakällor och bygger konkreta tillämpningar. ESSnet Big Data startade i februari 2016 och pågick i 28 månader fram till maj 2018. Den består av 10 arbetspaket: åtta av dessa är innehållsinriktade,

<sup>12</sup> SCB och Trafikanalys, 2017. Metodutveckling för förbättrad sjötrafikstatistik med Big Data. <https://www.scb.se/contentassets/b764060b80a74b2baecc78d6b14d4ab7/slutrappport-metodutveckling-for-forbattrad-sjotrafikstatistik.pdf>

<sup>13</sup> ESS står för European Statistical System. Ett ESSnet projekt kännetecknas av ett nätverk av flera ESS organisationer som syftar till att ge resultat som kommer till nytta för hela ESS. [https://ec.europa.eu/eurostat/cros/page/essnet\\_en](https://ec.europa.eu/eurostat/cros/page/essnet_en)

medan två består av samordning och spridning som dessutom stöder det övergripande projektet.

## 4.2 Big data inom officiell statistik

Under senare år har Big data som möjligt underlag för officiell statistik aktualiserats.

Exempel på hur Big data används i produktion av officiell statistik är datainsamling till Konsumentprisindex, KPI. Många länder inom EU har ersatt en del av datainsamlingen för att beräkna KPI, med uppgifter från butikskedjors kassasystem. Även SCB har sedan ett par år tillbaka ersatt en del av den traditionella datainsamlingen, där intervjuare besöker olika butiker och gör prismätningar, med data direkt från dagligvaruhandeln. Detta har både rationaliserat insamlingen av prisuppgifter och höjt kvaliteten i KPI.

Ett annat exempel på hur Big data används av SCB är för att ta fram statistik om grönområden i tätorter. Då kombineras information från satellitbilder med register om fastigheter och befolkning. Tillsammans gör dessa datakällor att man kan se hur mycket grönområden som finns i en tätort och hur mycket av grönyta som är tillgänglig för allmänheten.

AIS-data är ett exempel på Big data som dessutom har en geografisk komponent. De har i detta avseende stora likheter med andra geografiska Big data som t.ex. mobiltelefondata. AIS-data är närmast att hänföras till UNECE:s tredje kategori, massgenererade data från sensorer.<sup>14</sup>

## 4.3 AIS-data

I sjötrafiksammanhang handlar Big data i stor utsträckning om AIS-data (Automatic Identification System). AIS är ett globalt system som används för att identifiera fartyg och följa deras rörelser genom koordinater. Varje fartyg med en AIS-transponder sänder kontinuerligt och automatiskt ut information via VHF-radio som vem som helst med en AIS-mottagare kan ta emot. Mottagare finns installerade på fartyg och samlas även in av fartygstrafikservice, och används av dessa som ett led i säkerhetsarbetet.

AIS-data används i praktiken på de flesta kommersiella fartyg och många fritidsbåtar och sänder automatiskt ut information som fartygets position och hastighet samt fartygets identitet, storlek och hemvist med täta mellanrum.

### AIS i Sverige

I Sverige samlas AIS data in av Sjöfartsverket, dock inte i först hand för statistikändamål.

Ett planeringsprojekt<sup>15</sup> som genomfördes av SCB och Trafikanalys under 2015 visade att AIS-data har stor potential att förbättra sjötrafikstatistiken, genom bättre kvalitet till lägre kostnad, och med minskad uppgiftslämnarbörd för hamnarna.

<sup>14</sup> SCB och Trafikanalys, 2017. Metodutveckling för förbättrad sjötrafikstatistik med Big Data.

<https://www.scb.se/contentassets/b764060b80a74b2baecc78d6b14d4ab7/slutrapport-metodutveckling-for-forbatttrad-sjotrafikstatistik.pdf>

<sup>15</sup> Förbättrad sjötrafikstatistik med Big Data - en utvärdering, SCB:s diarienummer 2014-05466 (Vinnovas dnr: 2014-05466).

Några av resultaten var.

- Det visade sig vara fullt möjligt att använda AIS-data för att generera en lista över de fartyg som anlöpt en given hamn. Om fartygslistor rutinmässigt kunde genereras på detta sätt skulle hamnarnas uppgiftslämnarbörda minska avsevärt.
- AIS-data kan användas för att skapa en distansmatris mellan svenska hamnar som är bättre än den som tagits fram av Eurostat och i dagsläget används av Trafikanalys.
- Två metoder för att beräkna avstånd mellan hamnar provades, en vektorbaserad och en rasterbaserad metod, där den senare verkade mest lovande.

Projektet gav också tydliga indikationer om att AIS-data skulle kunna vara användbara för statistikframställning även för olika områden gränsande till sjötrafikstatistiken, speciellt miljöområdet. En bedömning gjordes att de preliminära metoder som togs fram för identifiering av fartygsrutter mellan hamnar skulle kunna användas för att bearbeta andra typer av massgenererade positioneringsdata (såsom mobiltelefondata eller GPS-data).

Trafikanalys och SCB har i ett fortsättningsprojekt tagit fram metoder för att använda positionsdata från sjötrafikens AIS-data i framställning av officiell statistik. Projektet har delvis finansierats av Vinnova. I detta fortsättningsprojekt var målet att ta fram validerade metoder för att med hjälp av AIS-data,

- skapa en högupplöst avståndsmatris mellan svenska hamnar samt beräkna avstånd mellan svenska och utländska hamnar,
- ta fram relevanta data om anlöp till varje svensk hamn under en avgränsad tid,
- beräkna antalet fartyg som passerar en given passagelinje under en avgränsad tid, samt beräkna trafikarbete för fartyg med olika egenskaper och för olika geografiska områden.

Projektet levererade validerade metoder samt verktyg i form av dokumenterad programkod, däremot inte färdiga produktionssystem.

Länk till SCB:s webbplats med information om projektet och rapporten:

<https://www.scb.se/om-scb/nyheter-och-pessmeddelanden/big-data-i-sjofartsstatistiken/>

Ett tredje steg i projektet om AIS-data har varit att vidareutveckla verktyget, och göra det fullt fungerande för ett års data. Nu är 2015 års data inlagt i verktyget och WSP har avropats för implementeringen. En slutredovisning till Vinnova gjordes 11 juli 2018. Då kommer verktyget och resultaten för 2015 års data integrering att beskrivas. Eventuellt blir det även ett PM hos Trafikanalys.

Metoderna inklusive verktyg ska i ett senare skede kunna implementeras i Trafikanalys produktion av sjötrafikstatistik. Tanken är att detta på sikt ska bidra till kvalitetshöjningar av befintlig sjötrafikstatistik och sänkta produktionskostnader, ge helt ny och relevant statistik, och samtidigt minska uppgiftslämnarbördan för hamnarna. Trafikanalys kan även använda uppgifterna för granskning av inlämnade uppgifter från hamnarna. Exempel på användningen är en ny matris för beräkning av persontransportarbete per linje och godstransportarbete. Sjöfartsverket erbjuder Trafikanalys att köpa in uppgifter för 30 000 kr per år.

Projektet har dessutom gett en generell kunskaps- och erfarenhetsutveckling kring positions-data som långsiktigt kan leda till innovation inom helt andra områden.

Andra aktörer som också har använt AIS-data är Chalmers. De har gjort egna simuleringsmodeller för exempelvis olycksrisker att fartyg kraschar med en nybyggd bro. Det har dock inte förekommit något samarbete med Trafikanalys kring datahantering.

## **AIS i Nederländerna**

Fördelarna med AIS (Automatic Identification Systems) är enligt CBS i Nederländerna att data om hamnbesöken kan generera mer detaljer än befintlig sjöstatistik, bland annat information om tid och distanser inom hamnen. Genom information om fartygets hastighet skattas om ett fartyg rör sig eller inte. Beroende på var stoppen görs kan man avgöra om det är en plats för av- eller pålastning. Utifrån denna information kan resvägen avgöras för fartygen. Bristerna med AIS-datan består dock av att det saknas information om lastat och lossat gods, bogserbåtar samt information om fartygstyp. CBS har framtida planer på att beskriva trafikdensitet, körrutter och emissioner med AIS-data.

De har utvecklat ett webbverktyg som kan visualisera och analysera AIS-data. (Trafikanalys kan få inloggningsuppgifter vid behov.) Visualiseringen av hamnbesöken görs av designstudenter och ett kommunikationsföretag tillsammans vid Big data-avdelningen på CBS, där många studenter gör praktik.

I Nederländernas hantering av AIS-data utnyttjar de bara mottagarsändarna på land som skickar signaler med en minuts intervall och med begränsade timeslots. AIS-datan sammanställs i läsbara filer från decoders.

Ett problem med AIS-data från transpondrarna på fartygen är att de ibland slås av. Detta gör fartygen eftersom de inte vill avslöja sina oljetankar på fartyg i hamn, eftersom det är marknadskonkurrens om raffinerade oljeprodukter med avseende på prissättningen längs kustlinjen vill man inte avslöja sina "lager" i hamn. Detta gör att det blir "glapp" i statistiken. Det är egentligen inte tillåtet att slå av transpondern inom territorialvattnet för handelsfartyg av viss storlek. Kvalitetsproblem genom bland annat fördröjd tidsangivelse som leder till glapp/stopp hanteras genom en algoritm för "port visits".

Nederländerna är med i ett ESSnet europeiskt projekt. För ESSnet Big data<sup>16</sup> är syftet att integrera stora data i den regelbundna produktionen av officiell statistik, genom piloter som undersöker potentialen hos utvalda stora datakällor och att bygga konkreta tillämpningar.

För inlandsvattenvägar har de inte fullständig täckning på AIS-datan eftersom fartygen inte rapporterar sitt gods med "River Information System". Men CBS:s mål är att de ska kunna använda AIS data för att beräkna bortfallet i statistiken för de större floderna.

---

<sup>16</sup>2017-05-22: [https://webgate.ec.europa.eu/pfis/mwikis/essnetbigdata/index.php/ESSnet\\_Big\\_Data](https://webgate.ec.europa.eu/pfis/mwikis/essnetbigdata/index.php/ESSnet_Big_Data)

## 4.4 Tidtabellsdata

Enligt föreskrift från Transportstyrelsen, ska alla företag som bedriver kollektivtrafik lämna uppgifter för trafikantinformation till Samtrafiken AB. De uppgifter som lämnas omfattar namn och geografiska positioner för hållplatser, tidtabeller, färdmedel, huvudmän och utförare. Uppgifterna sammanställs i en databas. Samtrafiken AB skickar veckovis underlag från databasen till Trafikanalys.

Underlagen från Samtrafiken AB bearbetas och analyseras av Trafikanalys. Med hjälp av kartbaserade väg- och spårnät tillsammans samt hållplatsernas positioner beräknas antal körda kilometer. Uppgifterna om huvudmän och utförare används för att identifiera vem som är ansvarig för trafiken.

Resultaten används för flera ändamål. Ett är att de används som stöd till uppgiftslämnare när övriga uppgifter om kollektivtrafik samlas in. I kollektivtrafiken publiceras numera också ett par tabeller baserade på tidtabelldata. Vidare används de också i Trafikanalys analysarbete, bl.a. i ett par uppdrag som kommit genom Trafikanalys regleringsbrev. Med tidtabellsdata beräknas invånarnas tillgång till kollektivtrafik efter i vilket funktionellt geografiskt område de bor. De geografiska indelningarna baseras på invånarnas socioekonomiska situation eller kommungruppsindelningar, som spänner mellan storstadskommuner och avlägset belägna landsbygdskommuner.



## 5 Slutsatser

Den tekniska utvecklingen har förvisso lett till att mängden digital information i samhället ökat lavinartat. Trafikanalys har på olika sätt prövat möjligheten att nyttiggöra dessa datamängder som underlag för sin statistikproduktion. I något fall har möjligheter att ersätta eller utveckla befintliga insamlingsmetoder identifierats och dessa kommer också att implementeras i myndighetens produktion.

Det finns dock mycket data som förefaller relevanta men som inte befunnits lämpliga för användning i produktion av officiell statistik. Skälen består bland annat av kvalitet, transparens, äganderätt samt resurser för utveckling av informationen till användbar statistik. Detta kanske får betraktas som naturligt då många nya datakällor utvecklats för helt andra syften än produktion av officiell statistik.

Det är Trafikanalys förhoppning att den fortsatta utvecklingen av datahantering i samhället kan göras med dessa erfarenheter i minnet. Med en bra dialog, mellan de som utvecklar system för datahantering och myndigheter med uppdrag att samla statistik för samhällets kunskapsförsörjning, bör det finnas goda möjligheter att både minska uppgiftslämnarbördan och effektivisera den allmännyttiga statistikproduktionen. Trafikanalys kommer att fortsätta söka efter sådana möjligheter.

Några av de lärdomar vi fått genom detta projekt sammanfattas nedan såsom specifika observationer för respektive statistikområde.

### 5.1 Specifika observationer

De viktigaste slutsatserna i denna PM är följande:

#### *Lastbilsundersökningen (tung lastbil)*

- Nederländernas satsning på överföring av systemdata för lastbilsundersökningen (XML-överföring) fungerar väl, men det tog många år och stora insatser för att få systemutvecklare att ta sig an uppgiften och för att få företag att leverera data på detta sätt. Tids- och resursaspekten är bra att ha i åtanke inför eventuell utveckling i Sverige.
- Kontakt och samarbete med branschorganisationer kan vara bra för att få förståelse för uppgiftsinsamlingen och för att använda deras kontaktytor för att sprida kunskap om statistikinsamlingen.
- Även hos oss i Sverige kan det vara intressant att testa metoden med att skriva i fritext på varuslaget, där en algoritm tar hand om detta och klassar till rätt ort. Då kan även länet komma upp elektroniskt som komplement till orten. Detta kanske även kan vara något för den svenska resvaneundersökningen. Även autofill i variabler vid ifyllande av webbenkät bör kunna testas i Sverige.
- Fortsätt hålla kontakten med Norges statistikbyrå för att följa deras utveckling av hur nya typer av data, såsom data från inbyggda system i lastbilarna, kan bidra till effektivare rapportering.

- Inga pappersenkäter används i Norge och Danmark. Enbart webbinsamling (samt en app i Danmark). Det fungerar bra.
- Den danska app-insamlingen av lastbilsdata kommer vara intressant att följa, om användandet ökar över tid. Att Trafikanalys dessutom kan ta del av systemdokumentation och test-inlogg från denna kan vara värdefullt.

#### *Lätta lastbilar*

- Nederländernas slimmade enkät till ägare av lätta lastbilar, som enbart ska ta 15–20 minuter att besvara kan vara en förebild för oss.
- Tänk på att som i Nederländerna hålla instruktioner, missiv och enkäter m.m. lättbegripliga och tydliga, att tala respondenternas språk. Det underlättar för respondenterna.
- Enbart webbenkät användes i Nederländerna. Respondenten var tvungen att ringa in och be om pappersenkät vid behov, vilket väldigt få gjorde. Av kostnadsskäl gjordes enbart telefonpåminnelser, tre påminnelser via brev skickades ut. Detta gav ändå en svarsfrekvens på 45 procent, utan svarsplikt. Deras erfarenheter av att det går bra att använda webbenkät och att helt skippa pappersenkät verkar alltså lovande.
- Trafikanalys har tillgång till metoddokumentation från Nederländernas undersökning, kanske kan vi få fler idéer därifrån när vi kommit längre med vår svenska undersökning.

#### *Resvaneundersökningen*

- En lärdom från Holland var att hämta in registeruppgifter istället för att belasta respondenterna med extra frågor. I nya RVU Sverige kommer många uppgifter hämtas från register för att minska uppgiftslämnarbördan.
- Pilotundersökningar i Sverige visar att när enbart webbinsamling används tycks det ge en låg svarsfrekvens (omkring 14–16 %). Däremot gav kombinationen webb och pappersenkät dubbelt så hög svarsfrekvens (30 %) pappersenkäten skickades i dessa fall ut tillsammans med påminnelsemissivet. Detta kan tyda på att det kan löna sig att behålla möjligheten att skicka ut pappersenkäter.
- Utskick av presentkort till de respondenter som svarade på webben gav liten effekt. Svarsfrekvensen med presentkort på 100 kr om urvalspersonen svarade var 16 procent, att jämföra med 14 procents svarsfrekvens utan incitament.
- Vid jämförelse av genomförda resvaneundersökningar verkar pappersenkäter vara billigare än telefonintervjuer. Svarsfrekvensen för pappersenkäter verkar heller inte ha minskat lika kraftigt över tid som för telefonintervjuer.
- Vid byte av insamlingsmetoder är det viktigt att analysera om det finns förklaringar till eventuellt ändrade svarmönster. Att jämföra resultaten och definitioner av det man mätt för att se eventuella skillnader.
- När det gäller mobilnätsdata kan det vara en fördel om Trafikanalys aktivt deltar i det framtida utvecklingsarbetet av att sätta standarder för hur data kan levereras. Då kan



vi lättare göra kontroller och förstå vägen från mobilnätdata till statistik. Annars sätter någon annan standarden.

- Den kommande RVU:n kommer genomföras som en kombinerat webb- och pappersenkät. De nya insamlingsmetoderna, såsom resvaneapp och mobilnätdata, är inte fullt ut redo att ersätta de traditionella metoderna.

#### *Big data*

- AIS-projektet fortsätter i Sverige och verkar lovande med ett fungerande verktyg. Eventuellt kommer tidsserier av transportarbetet att revideras bakåt i tiden, med den nya metoden blir beräkningarna mer tillförlitliga.
- Nederländernas webbverktyg som kan visualisera och analysera AIS-data var snyggt. Ett sådant verktyg skulle säkert öka användbarheten och intresset av AIS-data även i Sverige.
- Tidtabellsdata används tillsammans med uppgifter om geografiska positioner för hållplatser, färdmedel, huvudmän och utförare för att beräkna antal körda kilometer. Denna data används även som stöd till uppgiftslämnare och för analysarbete.
- Svårt att hitta direkt matnyttiga exempel på projekt baserat på Big data för vår verksamhet via konferenser.



Trafikanalys är en kunskapsmyndighet för transportpolitiken. Vi analyserar och utvärderar föreslagna och genomförda åtgärder inom transportpolitiken. Vi ansvarar även för officiell statistik inom områdena transporter och kommunikationer. Trafikanalys bildades den 1 april 2010 och har huvudkontor i Stockholm samt kontor i Östersund.