



Statistiken vid Trafikanalys och dess potential för geografiska analyser PM
2010:3

Statistiken vid Trafikanalys och dess potential för geografiska analyser

**PM
2010:3**

Trafikanalys

Adress: Sveavägen 90

113 59 Stockholm

Telefon: 010 414 42 00

Fax: 010 414 42 10

E-post: trafikanalys@trafa.se

Webbadress: www.trafa.se

Ansvarig utgivare: Brita Saxton

Publiceringsdatum: 2010-12-30

Förord

Trafikanalys bildades den 1 april 2010 och övertog då ansvaret för den officiella statistiken om transporter och kommunikationer. Som ett led i utvecklingen av den nya myndighetens verksamhet påbörjades en genomlysning av statistikproduktionen både med avseende på relevans och på kvalitet. I den svenska transportplaneringen ökar efterfrågan på nya typer av analyser, bland annat med trafikslagsövergripande perspektiv och med en högre geografisk upplösning än den nationella nivån. I syfte att beskriva förutsättningarna för denna typ av analyser har Trafikanalys genomfört en inventering av sina statistikprodukter.

Föreliggande PM sammanfattar resultaten av den genomförda inventeringen. Resultatet utgör samtidigt en översikt över samtliga av myndigheten publicerade statistiktabeller och därmed ett underlag för den som önskar en helhetsbild över dessa. Arbetet med inventering och rapportskrivning har utförts av utredaren Johan Holmer i samarbete med personal som arbetar med olika statistikprodukter vid Trafikanalys.

Östersund december 2010

Per-Åke Vikman
Avdelningschef

Innehåll

Förord	3
1 Inledning	7
1.1 Bakgrund	7
1.2 Syfte och innehåll	7
2 Metodbeskrivning	9
2.1 Registrets framställningsprocess och innehåll	9
2.2 Grunder för bedömning av GIS-potential	11
2.3 Användning.....	11
3 Resultat	13
3.1 Allmän statistik.....	14
3.2 GIS-statistik	17
4 Sammanfattning och diskussion	21
4.1 Sammanfattning.....	21
4.2 Utvecklingsmöjligheter	21

1 Inledning

1.1 Bakgrund

Analys och beskrivningar av nuläget inom transportsektorn utgör en viktig del av Trafikanalys uppdrag, liksom att tillgängliggöra denna kunskap för intresserade parter i samhället. Dessa uppgifter faller främst på Avdelningen för utvärdering, nulägesanalys och statistik (UNS). För att kunna utföra dem effektivt behövs en god överblick över den statistik som myndigheten innehar, bearbetar och presenterar. På nästa sida finns en tabell över vilka publikationer som är aktuella för utgivning av Trafikanalys under 2011. Ett behov av att analysera och presentera denna statistik på en finare detaljnivå rent geografiskt finns också. Anledningen är en ökande förekomst och betydelse av regional delaktighet i infrastrukturprojekt; både vad gäller finansiering och planering. Med detta följer förmodligen ett växande behov av bland annat flödesberäkningar och transportekonomiska analyser på den regionala nivån.

Att gå från att tillhandahålla statistik på riksnivå till att göra detsamma på regional nivå, eller lägre, är dock förenat med vissa begränsande förutsättningar. En sådan är sekretessen, då aktörerna på regional nivå kan vara för få eller dominanta lokalt för att transportarbetet som utförs regionalt ska kunna offentliggöras utan att informationen kan kopplas till ett visst företag. Detsamma gäller för individer, främst i samband med olycksstatistik och rese mönster. Men ett ibland för litet antal observationer riskerar inte bara att leda till risk för röjande av företag och individer när den geografiska aggregationsnivån blir lägre, risken är att även den statistiska säkerheten blir lidande.

1.2 Syfte och innehåll

Tabellregistret, som denna PM handlar om, är avsett att ge den tidigare nämnda överblick över Trafikanalys statistikprodukter och i vilken utsträckning de kan presenteras och jämföras på en regional aggregeringsnivå. Vid sådana jämförelser är olika GIS-programvaror, som ArcGIS och MapInfo, användbara och därför finns i registret en bedömning av lämpligheten hos varje enskild statistiktabel för gestaltning och analys med sådana verktyg. Tillvägagångssättet vid bedömningarna och framtagandet av registrets övriga delar kommer ett metodkapitel på de följande sidorna att beskriva. Kapitlet därefter presenterar registrets resultat i form av en statistisk sammanställning av dess datainnehåll. Det fjärde och avslutande kapitlet för en diskussion kring de viktigaste slutsatserna registret leder till och hur Trafikanalys skulle kunna använda det i sin verksamhet.

Tabell 1.1: Trafikanalys publikationer som är aktuella för utgivning år 2011.

Publikationsnamn	Undersökning	Utgivningsfrekvens	Sveriges Officiella Statistik
Fordon 2010	Fordon	Årlig	SOS
Fordon i län och kommuner	Fordon	Årlig	SOS
Fordon nyregistreringar	Fordon	Månatlig	SOS
Vägtrafikskador – antal dödade 2010	Vägtrafikskador	Årlig	SOS
Vägtrafikskador 2010	Vägtrafikskador	Årlig	SOS
Vägtrafikskadade i sjukvården	Vägtrafikskador	Årlig	
Lastbilstrafik – Inrikes och utrikes trafik med svenska lastbilar	Lastbilstrafik	Kvartalsvis	SOS
Lastbilstrafik – Inrikes och utrikes trafik med svenska lastbilar	Lastbilstrafik	Årlig	SOS
Utländska lastbilstransporter	Lastbilstrafik	Intermittent	
Varuflödesundersökningen	Varuflödesundersökningen	Intermittent	SOS
Bantrafik	Bantrafik	Årlig	SOS
Person- och godstransporter på järnväg	Bantrafik	Kvartalsvis	
Bantrafikskador	Bantrafik	Årlig	SOS
Sjötrafik – Utrikes och inrikes trafik med fartyg	Sjötrafik	Kvartalsvis	SOS
Sjötrafik – Utrikes och inrikes trafik med fartyg	Sjötrafik	Årlig	SOS
Fartyg – Svenska och utländska fartyg i svensk regi	Fartyg	Årlig	SOS
Sjöfartsföretag	Sjöfartsföretag	Årlig	SOS
Luftfart	Luftfart	Årlig	SOS
Postverksamhet	Postverksamhet	Årlig	SOS
Televerksamhet	Televerksamhet	Årlig	SOS
Färdtjänst och riksfärdtjänst	Färdtjänst och riksfärdtjänst	Årlig	SOS
Lokal och regional kollektivtrafik	Kollektivtrafik och samhällsbetalda resor	Årlig	
Kollektivtrafik och samhällsbetalda resor	Kollektivtrafik och samhällsbetalda resor	Årlig	
Långväga buss	Kollektivtrafik och samhällsbetalda resor	Årlig	
Transportbranschen – hur står det till?	Transportbranschen	Intermittent	
Trafik- och transportarbetet i Sverige	Trafik- & transportarbete	Årlig	
Uppföljning av de transportpolitiska målen	Måluppföljning	Årlig	
Marginalkostnadsberäkningar i transportsektorn	Marginalkostnader	Årlig	

2 Metodbeskrivning

Det här kapitlet kommer till att börja med att beskriva tabellregistrets innehåll och hur det togs fram. Bakgrunden till varför det ser ut som det gör och vilka definitionerna har varit kan vara bra att känna till för den som ska använda det. Ett särskilt stycke kommer också att ta upp andra saker som är bra att känna till för en eventuell användare.

2.1 Registrets framställningsprocess och innehåll

På Trafikanalys hemsida finns tydligt angivna ämnesområden dit myndighetens statistikprodukter hör. Vid tabellregistrets framställning var utgångspunkten hemsidans statistikindelning där samtliga statistikprodukter finns publicerade med undantag för "Långväga buss" som istället återfinns under rubriken "publikationer". Alla publikationers senast utgivna version inventerades. Diskussioner fördes med de ansvariga för respektive statistikområde för att klargöra de fakta som inte alltid framgår i produkterna, exempelvis vilken myndighet som rent fysiskt innehar mikrodataunderlaget.

De flesta publikationer innehåller ett flertal tabeller och det är dessa som utgör grunden i registret. De är alla listade med namn, nummer, vilken statistikprodukt de kommer från samt i förekommande fall även produktens nummer. Statistik som tidigare år utgjort självständiga produkter men som numera istället gått upp som delar i andra produkter är inte angivna för sig.

För varje tabell i publikationerna angavs vem den ansvarige kontaktpersonen för statistikområdet är och vilken myndighet som publicerat den då Somlig statistik som producerades av SIKA ännu inte har publicerats i ny version av Trafikanalys. I två kolumner angavs länkar till pdf- respektive Excel-versionerna av statistikprodukten om de fanns på hemsidan, så att en användare av registret snabbt ska kunna få tillgång till de tabeller som ser intressanta ut. En nackdel med detta är förstås att länkar kan bli inaktuella, så visst underhåll kan krävas för att denna del av registret ska behålla sin funktion. I flera av statistikprodukterna finns diskussioner om datas tillförlitlighet och jämförbarhet över tid och med andra data. I förekommande fall har sådana diskussioner sammanfattats i två kolumner dedikerade till ämnena.

Om ett register av det här slaget ska kunna fylla sitt syfte och utgöra ett stöd i arbetet för de som använder det är det viktigt att det är lätt att söka i dess innehåll. I skrivande stund innehåller registret beskrivningar av 731 tabeller och att manuellt leta fram tabeller av ett visst slag är då inte det lättaste. Därför har varje tabell tilldelats nyckelord i upp till sex olika kolumner, beroende på innehållet. Tanken är att det med hjälp av nyckelorden och Excels "filter"-funktioner ska gå lätt att sortera fram tabeller med ett visst innehåll. Till att börja

med innehåller den första av dessa kolumner; "Trafikslag/Kategori", information om vilken slags trafik tabellerna handlar om. Det kan vara exempelvis luftfart eller vägtrafik. Men det är inte all statistik som handlar om fysisk trafik, det kan även röra sig om kommunikationer via internet eller telefon och sådana tabeller har istället kategoriserats som just "Kommunikationer". Relativt snabbt gick det också att avgöra om en tabell berörde godstransporter, persontransporter eller både och så detta anges i ämneskolumn nummer två. Återigen fanns tabeller som inte riktigt kunde kategoriseras enligt denna modell; de tilldelades nyckelorden "IT" respektive "infrastruktur" beroende på ämne. Dessa två sorteringsmöjligheter räcker dock inte särskilt långt för den som vill göra detaljerade sökningar, varför ytterligare fyra kolumner skapades med olika nyckelord beroende på vad tabellerna innehåller. Till exempel är det möjligt att sortera fram alla tabeller som på något sätt kan vara aktuella i miljösammanhang, eller alla tabeller som innehåller statistik om körsträckor. Anledningen till att dessa mer detaljerade sorteringsmöjligheter fordrar ytterligare fyra kolumner finner vi i Excels begränsningar och behovet av överskådlighet; ämnena är långt ifrån ömsesidigt uteslutande och det hade varit mycket opraktiskt att kombinera dem i alla potentiella sorteringsfilter i samma kolumn.

För den som söker statistik med registrets hjälp kan det även vara användbart att snabbt se vilka tidsperioder tabellerna omfattar. Detta finns angivet för varje tabell och bedömdes dels med hjälp av tabellens titel, som ofta anger denna information, men även genom att granska innehållet då titel och faktiskt innehåll inte alltid överensstämmer. I de fall de inte gör det är tidsperioden för innehållet och inte den som nämns i titeln angivet i registret. Även periodiciteten med vilken statistiken utkommer finns nedtecknad.

Flera kolumner behandlar möjligheterna att redovisa statistiken på en mer detaljerad nivå än riksnivå och hur lämplig den är för GIS-analyser. Den första av dessa kolumner redovisar aggregationsnivån i produkternas tabeller. I de flesta fall är data aggregerad på nationell nivå vilket innebär att den inte kan användas till geografiska analyser i sin behandlade form. Men det finns också ett flertal tabeller där aggregeringsnivån är regional, kommunal, eller på till exempel flygplats- eller hamnnivå. Då det kan vara intressant att se vilken statistik som finns på olika NUTS-nivåer finns en kolumn även för detta. Som beskrevs i inledningen kan sekretess och statistisk osäkerhet vara anledningar till att data aggregerats till riksnivå. I vissa statistikprodukter finns sådana begränsningar angivna och i sådana fall finns de med, i förkortad form, även i tabellregistret. Ytterligare en kolumn beskriver vilken "geografisk omfattning" data i en tabell har. I de flesta fall gäller statistiken Sverige, men ibland kan det även finnas statistik för grannländer, EU-länder eller hela världen. För GIS-ändamål kan det vara bra att se om statistiken har någon typ av geografisk kodning som underlättar överföringen till kartprogram. Endast i undantagsfall fanns sådan kodning och då i form av kommunkoder, men presenteras data med någon typ av geografisk indelning är det relativt lätt att ordna en egen kodning. Huruvida detta är möjligt eller ej är angivet för varje tabell i en kolumn. I ytterligare en kolumn gjordes en bedömning av om data i tabellen lämpar sig för GIS eller om hinder föreligger. En motsvarande kolumn upprättades för mikrodatas lämplighet, efter samtal med den ansvarige för det aktuella statistikområdet.

Utöver information om mikrodatas lämplighet för GIS-analyser samlades ett antal andra variabler in vid samtalen med de statistikområdesansvariga. Den första är var ursprungsdata finns och vilken källan är. En stor del av datamaterialet finns rent fysisk inte hos Trafikanalys utan hos andra myndigheter som sammanställt och skickat tabellerna. I vissa fall finns även sekundära datakällor och då är det angivet i en kolumn. När data faktiskt finns hos Trafikanalys är det vanligtvis lagrat i någon typ av databas. En kolumn innehåller i dessa fall information om vilken typ av databas det är. Slutligen tillfrågades de statistikansvariga om möjligheterna att jämföra data med motsvarigheter från andra länder. Dessa möjligheter varierade förstås, men för alla tabeller finns kommentarer om dem i en kolumn. En "övrigt"-kolumn finns också där all information som inte passade in i något av de andra sammanhangen matades in.

2.2 Grunder för bedömning av GIS-potential

Med all information om alla statistikprodukter insamlad var nästa steg att bedöma lämpligheten hos de olika tabellernas innehåll för GIS-användning. I bedömningen gjordes en sammanvägning av alla relevanta variabler; aggregationsnivå, sekretessfrågor, geografisk omfattning och lämplighet i på hemsidan presenterat data och mikrodata. Det som vägde allra tyngst var de statistikansvarigas kommentarer om materialet.

Graden av lämplighet hos ett datamaterial indikeras med en skala om tre steg. När ingen tvekan råder om att materialet i en tabell kan användas i GIS-sammanhang har lämpligheten klassats som god. I dessa fall föreligger inga problem med sekretess eller mikrodata och ofta har även tabellens data en gynnsam aggregationsnivå. Data som saknar dessa kvaliteter har klassats som "olämpligt", ofta på grund av sekretessproblem eller för höga aggregationsnivåer men i vissa fall också för att statistikansvarig bedömt att en orealistiskt stor arbetsinsats skulle behövas för att få rätt format på data. För de data som faller i gränslandet mellan dessa kategorier finns klassningen "tveksam" lämplighet. Här rör det sig i de flesta fall om data som på många sätt är lämplig men har någon faktor som talar mot sig, exempelvis oklarheter i sekretessfrågan. Det kan också röra sig om data som helt klart kan användas i GIS-sammanhang, men som statistikansvarig bedömt som ointressant. Bedömningen av datamaterialet har gjorts i en fast ordning. Först har relevansen för GIS-analyser värderats, följt av om mikrodata har det som krävs och slutligen huruvida sekretess utgör ett problem. Tabeller vars mikrodata inte duger för GIS-användning skulle alltså mycket möjligt kunna fastna på sekretessen även om grundmaterialet på något sätt kunde förbättras. Notera att data som bara utgör ett aggregerat värde för hela Sverige och inget annat land automatiskt klassas som olämpligt för GIS-användning, det finns ingen mening med att illustrera ett enda värde på en karta.

2.3 Användning

Tabellregistret har två huvudsakliga syften. Det ena, att ge en överblick över Trafikanalys statistik innehav, har det redan uppfyllt när denna PM är färdigställd.

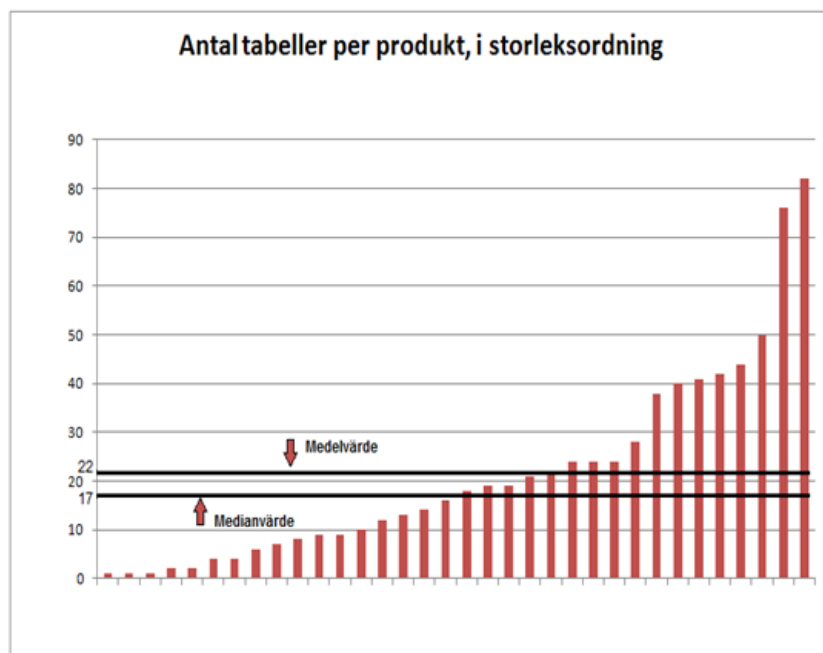
Det andra syftet är att tjäna som ett hjälpmedel för den som letar statistik med vissa egenskaper; exempelvis sådan med kommunal aggregeringsnivå. Det är här som nyckelorden kommer in i bilden, med deras hjälp kan användaren lätt få fram den statistik som har ett visst innehåll. Problemet är att Excels begränsningar tvingat fram en fördelning av nyckelorden över flera kolumner. Detta försvårar sökprocessen för ovana användare och den kompliceras ytterligare av att vissa nyckelord finns i två kolumner; exempelvis återfinns "tillstånd och innehav" som nyckelord både i kolumnen "Ämne 1" och i kolumnen "Ämne 2". För att få fram samtliga tabeller med sådant innehåll måste användaren alltså göra två olika sökningar, en på varje kolumn. Nyckelorden är fördelade över kolumnerna så att detta undviks i största möjliga mån, men tabellernas innehåll gör att problemet inte helt har kunnat elimineras. Givetvis går det att söka i tabellregistret med hjälp av de andra kolumnerna också, om användaren exempelvis bara vill se tabeller innehållande data från åren 1970-1990 eller vad en viss produkt innehåller.

Tabellregistret kan alltså vara ett användbart hjälpmedel, men för att få ut mesta möjliga av det krävs en viss förtrogenhet med hur det är upplagt, som lyckligtvis bör vara relativt snabbt uppnådd. Det krävs också att användaren kan bedöma hur relevant tabellregistrets data är i det enskilda fallet. När tabellerna kommer som de är från en yttre leverantör kan Trafikanalys knappast göra något annat av mikrodata än vad som går att förstå av tabellerna. Men i de fall mikrodata finns på myndigheten behöver inte begränsningen gå vid de enskilda tabellernas upplägg och då är de senare inget mer än en upplysning om vad mikrodata garanterat innehåller.

3 Resultat

Det färdigställda registret kan avslöja mycket om Trafikanalys statistikinnehav. Tack vare de många sorterbara kolumnerna finns ett stort antal kombinationer att undersöka. Att presentera alla potentiellt intressanta kombinationer här skulle vara opraktiskt; istället kommer bara grundläggande statistik att presenteras. Det står dock läsaren fritt att själv undersöka mer specifika kombinationer och den ansvarige för tabellregistret kan givetvis också tillhandahålla sådana på förfrågan.

Statistiken som nu kommer att presenteras är baserad på den version av tabellregistret som fanns mellan den 5 och 25 oktober 2010. Sedan dess har det uppdaterats och ändrats, så alla värden här stämmer inte helt med de aktuella. Helhetsbilden torde dock vara oförändrad. I den beskrivna versionen av tabellregistret finns 731 tabeller fördelade på 34 statistikprodukter så varje produkt innehåller i medeltal 22 tabeller. Men spridningen är relativt stor, vilket syns i Figur 3.1. Gini-koefficienten¹ för fördelningen är cirka 0,49. Detta kan vara värt att ha i åtanke vid tolkning av den övriga statistik som presenteras i detta kapitel. Exempelvis kommer den totala andelen av tabellerna med sekretessrelaterade begränsningar förstås att bli mycket större om en produkt med många tabeller har den egenskapen, än om en produkt innehållandes färre tabeller har den.



Figur 3.1: Fördelning av tabeller i statistikprodukterna.

¹ Ett mått på hur jämnt något, till exempel inkomster, är fördelat i en population. Ju lägre koefficient desto jämligare fördelning. När koefficientens värde är 1 innebär det att en individ har allt av det som mäts och övriga inget.

3.1 Allmän statistik

I nedanstående stycke kommer mer övergripande statistik om datamaterialet att presenteras. Därefter följer i det nästkommande stycket den GIS-relaterade statistiken.

Tabell 3.1: Överblick.

Antal statistiska produkter	34
Publicerat i Excelformat	25
Publicerat i pdf-format	22
Med tabellförklaringar på både engelska och svenska	16
Utgives årligen	26
Utgives kvartalsvis	3
Utgives intermittent	5
Senaste upplaga producerad av SIKA	20
Senaste upplaga producerad av Trafikanalys	14
Totalt antal tabeller	731

I Tabell 3.1 kan vi se att det är de årligen utgivna statistikprodukterna som dominerar vad gäller innehåll av antal tabeller. De flesta är tillgängliga på hemsidan både i Excel- och pdf-format, men långt ifrån alla. De flesta produkter utgavs också i senaste version av SIKA men detta kommer givetvis att förändras de kommande månaderna.

Tabell 3.2: Nyckelord i "trafikslag" och "typ av transport".

Trafikslag	Antal	Andel, %	Typ av transport	Antal	Andel, %
Alla trafikslag	251	34,3	Enbart gods	242	33,1
Bantrafik	39	5,3	Enbart person	246	33,7
Lufttrafik	29	4,0	Gods & person	129	17,7
Sjötrafik	76	10,4	Infrastruktur (ej transport)	21	2,9
Vägtrafik	236	32,3	Gods & Infrastruktur	2	0,3
Övrigt/Kommunikationer	100	13,7	IT (ej transport)	91	12,5
			Totalt gods	373	51,0
			Totalt person	375	51,3
			Totalt infrastruktur	23	3,1

Tabell 3.2 visar antalet och andelen av tabellerna som är kopplade till nyckelorden i de två första sorteringskategorierna: "Trafikslag" och "Typ av transport". I den första avdelningen ser vi att kategorin "alla trafikslag" är den största, det vill säga alltså tabeller vars data berör alla de övriga kategorierna

utom "kommunikationer". "Vägtrafik" är det överlägset vanligaste enskilda trafikslaget, cirka 32 procent av tabellerna innehåller sådan data. De övriga trafikslagen är avsevärt mindre till antalet, men det innebär inte nödvändigtvis att informationsunderlaget för dessa är sämre. "Kommunikationer" är också en relativt stor kategori då de flesta tabeller från Resvane- och Kommunikationsvaneundersökningarna sorterar under den. Dessa behandlar till stor del IT-användning, men också viss resestatistik och postverksamhet. I tabellens andra avdelning ser vi att ungefär en tredjedel av tabellerna handlar enbart om persontransporter och en tredjedel enbart om godstransporter. Cirka 18 procent innehåller data om båda ämnena medan cirka 12 procent är IT-relaterade. Den lilla resterande andelen handlar om infrastruktur. Totalt innehåller cirka 3 procent av tabellerna infrastrukturdata, medan motsvarande siffra är över 50 procent för både gods- och persontransporter, tack vare att en tabell kan innehålla data om båda ämnena.

Tabell 3.3: Nyckelord i ämneskolumnerna 1 och 2.

Nyckelord i ämneskolumn 1	Antal	Andel, %	Nyckelord i ämneskolumn 2	Antal	Andel, %
Ekonomi	124	17,1	Aktörer	34	4,7
Fordonsstatistik	168	23,0	Arbete	11	1,5
Internet	25	3,4	Fordonsstatistik	10*	1,4 *
Telefoni	23	3,1	Förbindelsestatistik	8	1,1
Internet och Telefoni	5	0,7	Förbindelsestatistik och fordonsstatistik	1*	0,1*
Olyckor	46	6,3	Färdsträckor	39	5,3
Resestatistik	64	8,8	Lasttyp	26	3,6
Svars- & Urvalsstatistik	46	6,3	Färdsträckor och lasttyp	4*	0,5*
Tillgänglighet	3	0,4	Resestatistik	8*	1,1*
Tillstånd & innehav	29	4,0	Tillstånd & innehav	9*	1,2*
Transportvolym	187	25,6	Trafikarbete	4	0,6
[Tomma]	54	7,4	Trafiknoder	12	1,6
			Transportvolym	5*	0,7*
			Varugrupper	43	5,9
			[Tomma]	526	72,0

I Tabell 3.3 visas hur datatabellerna fördelar sig över nyckelorden i registrets första och andra ämneskolumner. I ämneskolumn 1 är det "transportvolym" som dominerar, då både gods- och persontransporter räknas till denna kategori är det inte så konstigt. Att kategorin "fordonsstatistik" är näst störst bör inte heller komma som någon överraskning. Däremot är det förvånansvärt många tabeller som innehåller någon typ av ekonomisk statistik. Svars- & Urvalsstatistik, som inte är till någon nytta för de ändamål för vilka registret skapats, utgör mer än 6 procent av datamaterialet. Notera att vissa av nyckelorden från ämneskolumn 1 återkommer i kolumn 2. Anledningen är att flera nyckelord från samma ämneskolumn passar för dessa tabeller; det har inte alltid varit möjligt att gruppera kolliderande nyckelord i olika ämneskolumner och när de kolliderat har ett av dem flyttats till den andra ämneskolumnen. "Fordonsstatistik" är exempelvis ett nyckelord som huvudsakligen huserar i ämneskolumn 1, men

även återfinns som nyckelord för totalt elva tabeller i ämneskolumn 2. Det totala antalet och andelen tabeller med ett nyckelord anges första gången det presenteras. Notera också att "och" skrivs på två olika sätt i tabellen; "&" och "och". När &-tecknet används utgör det del av ämnets rubricering, men när och istället skrivs "och" betyder det att två ämnen som annars står var för sig har kombinerats i en kategori för att tabellen innehåller båda typerna av data. I den andra ämneskolumnen finns inga ämneskategorier med särskilt många tabeller. "Varugrupper" är störst och "aktörer" och "färdsträckor" är relativt stora. Men de flesta, nästan 72 procent, av tabellerna saknar nyckelord i denna kolumn.

Tabell 3.4: Nyckelord i ämneskolumnerna 3 och 4.

<i>Nyckelord i ämneskolumn 3</i>	<i>Antal</i>	<i>Andel, %</i>	<i>Nyckelord i ämneskolumn 4</i>	<i>Antal</i>	<i>Andel, %</i>
Kollektivtrafik	250	34,2	Cabotage & transit	4	0,6
Kön	102	14,0	Miljö	43	5,9
Kollektivtrafik och kön	30*	4,1*	Personal	32	4,4
Utrikeshandel	68	9,3	Miljö och personal	1*	0,1*
Kollektivtrafik och utrikeshandel	8*	1,1*	Transportmedelsstatistik	39	5,3
[Tomma]	349	47,7	[Tomma]	614	84,0

Tabell 3.4 visar de två sista ämneskolumnerna och deras nyckelord. Ett stort antal tabeller, totalt 250 stycken, har med "kollektivtrafik" att göra. Med kollektivtrafik menas här inte bara bussar och färdtjänst utan alla typer av kollektivt resande, inklusive taxi. Många tabeller innehåller statistik uppdelad på män och kvinnor, totalt nästan 14 procent. Flertalet av dessa tabeller ingår i Resvaneundersökningen och Fordon 2009 så de 102 tabellerna är inte direkt jämnt fördelade över statistikområdena. Uppdelningen på kön utgör dessutom bara en mindre del av statistiken i en tabell i de flesta fall. Likväl är det ett material som kan användas till många intressanta jämförelser. I den fjärde ämneskolumnen är det "miljö" och "transportmedelsstatistik" som flest tabeller sorterar under. Nyckelordet miljö har tilldelats alla tabeller som innehåller något om utsläpp, energiförbrukning eller drivmedelsanvändning så det spänner över relativt mycket. "Transportmedelsstatistik" är ett nyckelord som tilldelats de tabeller som har information om med vilka transportmedel persontransporter sker och de tabellerna är näst flest. Gemensamt för dessa båda kolumner är dock att flertalet av tabellerna saknar nyckelord här.

3.2 GIS-statistik

Tabell 3.5: Bedömningar av lämplighet för GIS-användning.

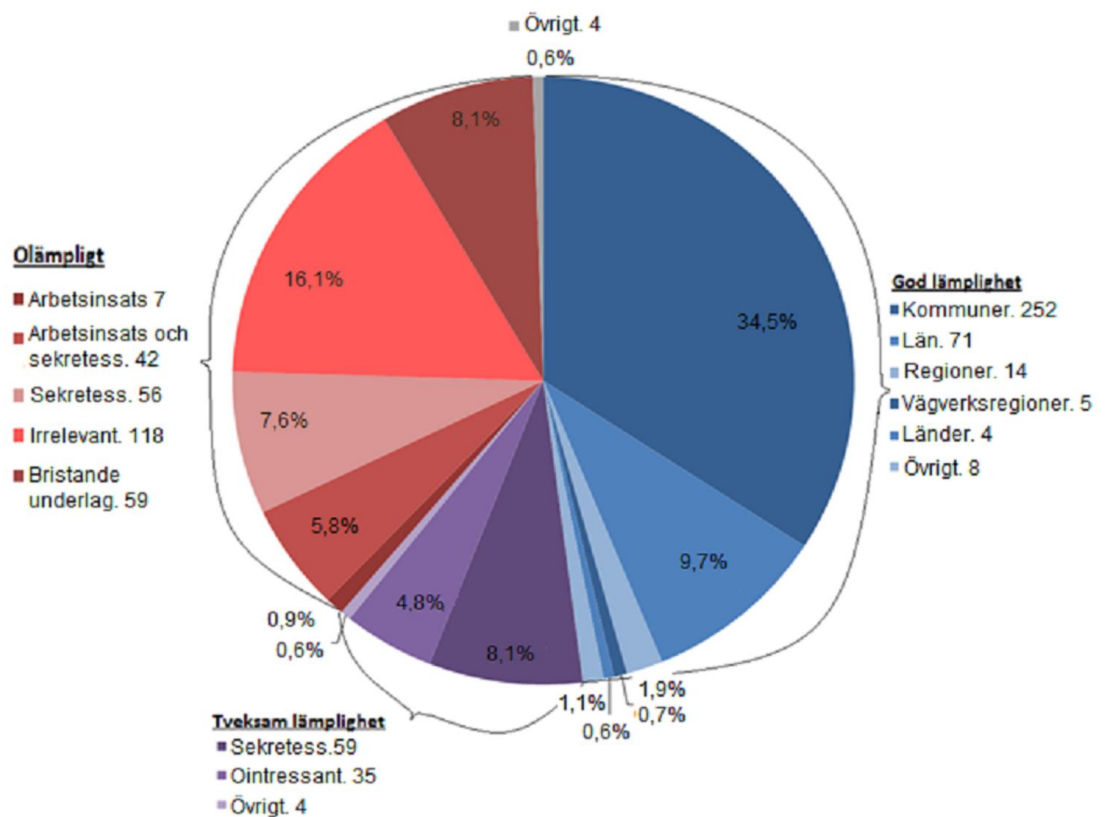
Bedömning av publicerade tabeller	Antal	Andel, %	Bedömning av mikrodata	Antal	Andel, %
God	158	21,6	God	335	45,8
Begränsad	14	1,9	God, men ointressant	13	1,8
Låg	513	70,2	Begränsad	102	14,0
Irrelevant	42	5,8	Låg	216	30,0
			Irrelevant	42	5,8
			Ej bedömt	19	2,6
Samlad bedömning	Antal	Andel, %			
God	354	48,4			
Tveksamt	88	12,0			
Olämpligt	285	39,0			
Övrigt	4	0,6			

I den första avdelning av Tabell 3.5 visas hur de publicerade tabellerna har bedömts utifrån lämplighet för GIS-användning. 22 procent av tabellerna bedömdes ha god lämplighet för någon typ av användning med GIS-verktyg. Användningsområdet behöver inte vara särskilt brett, men så länge det står utom tvivel att data kan användas till någonting har lämpligheten klassats som "god". 14 tabeller, 1,9 procent, bedömdes av olika anledningar ha ytterst begränsade möjligheter till GIS-användning. Den stora majoriteten av tabellerna, strax över 70 procent, går inte att använda till GIS-ändamål i sin nuvarande publicerade form. Ytterligare 42 tabeller är inte relevanta för GIS-användning.

I den andra avdelningen av Tabell 3.5 presenteras bedömningen av de mikrodata som de publicerade tabellerna är baserade på. Mikrodatas lämplighet för användning med GIS-program bedömdes utifrån de statistikansvarigas förklaringar och kommentarer kring datamaterialet. Nästan 46 procent tycks ha god lämplighet för någon typ av GIS-användning. Ytterligare 1,8 procent har bedömts vara lämpligt men förmodas av de statistikansvariga vara ointressanta för GIS-analyser. Cirka 14 procent, mikrodata till 102 tabeller, bedöms ha begränsad lämplighet. Anledningen är då ofta att sekretess potentiellt kan hindra användningen av vissa delar av materialet men att problemets omfattning är svår att bedöma. 30 procent av materialet har fått kategoriseringen "låg" lämplighet, vilket är mycket mindre än i bedömningen av tabellerna. Samma 42 tabeller som tidigare är bedömda som irrelevanta och 19 tabeller har ej kunnat bedömas. Sammantaget ser bilden ljusare ut för GIS-användning av materialet när bedömningen gjorts utifrån mikrodata och inte bara baserat på de publicerade tabellernas egenskaper.

Den nedre och sista avdelningen av Tabell 3.5 innehåller statistik över de samlade bedömningarna av Trafikanalys datamaterial. Här har hänsyn tagits till tabellernas innehåll, mikrodata, aggregeringsnivåer, eventuella sekretessproblem med mera. Här användes en annan bedömningskala som gick från "god lämplighet" via "tveksam" till "olämpligt" för GIS-användning. Strax under hälften, 48,4 procent, bedömdes ha god lämplighet för GIS-användning och detta är något mer än i bedömningen av mikrodata. Förklaringen är att det finns fall där mikrodata och tabelldata kanske inte är användbart, men enligt de statistikansvarigas utsago kan bli det med relativt liten arbetsinsats. Det finns också material som kan användas i GIS-analyser men först efter mycket stora arbetsinsatser. 6,7 procent av materialet har bedömts som olämpligt av denna anledning, alltså 49 av de totalt 285 tabeller med tillhörande mikrodata som bedömts som olämpliga.

Ungefär en tredjedel av det olämpliga materialet, eller 13,4 procent av allt material, är bedömt som olämpligt på grund av sekretessproblem. Ett annat vanligt skäl är helt enkelt bristande underlag. Den fullständiga fördelningen av anledningar till olika bedömningar återfinns i Figur 3.2: Aggregeringsnivåer i GIS-lämplig data och skäl till bedömningen som "olämpligt" för övrig data. Värt att notera är att en långt större andel av materialet skulle fastna på sekretessproblemen om det inte vore för bedömningsordningen; nu är det mycket data som faller bort på grund av irrelevans och bristande mikrodata. Kvarstår gör den andel av materialet vars lämplighet bedömts som "tveksam". Här är anledningen oftast svårbedömda sekretessproblem eller att det förmodligen inte skulle vara intressant att använda materialet i GIS-analyser. I tabellregistret har data-materialet bedömts mer detaljerat än hur det beskrivs här; här har bedömningskategorierna aggregerats något för enklare presentation.



Figur 3.2: Aggregeringsnivåer i GIS-lämplig data och skäl till bedömningen som "olämpligt" för övrig data.

De två avslutande tabellerna beskriver de geografiska egenskaperna hos data i de publicerade tabellerna. I Tabell 3.6 framgår de olika aggregeringsnivåerna och deras frekvens. Som synes är data aggregerat på nationell nivå i mer än 80 procent av tabellerna och kan därför inte användas i GIS-analyser i obearbetad form annat än i de fall det rör sig om fler nationer än Sverige. Strax under 9 procent av tabellernas datainnehåll är aggregerat och presenterat på länsnivå. På kommunal nivå är det bara strax över 1,5 procent, 12 tabeller.

Tabell 3.6: Aggregeringsnivåer för data i statistiktabeller.

Aggregeringsnivå	Antal	Andel, %	Aggregeringsnivå	Antal	Andel, %
Företag	1	0,1	Nationell	589	80,6
Flygplatser	6	0,8	NUTS 2	4	0,6
Hamnar	1	0,1	Regional	13	1,8
Kommuner	12	1,6	Världen	1	0,1
Län	64	8,8	Urvalspopulation	35	4,8

Tabell 3.7: Geografisk omfattning hos data i statistiktabeller

<i>Geografisk omfattning</i>	<i>Antal</i>	<i>Andel, %</i>	<i>Geografisk omfattning</i>	<i>Antal</i>	<i>Andel, %</i>
Sverige	673	92,1	EU27+Norge	12	1,6
Sverige och några till	20	2,7	EU27+USA	1	0,1
Europa	2	0,3	Världen	14	1,9
EU25+Norge	2	0,3	Utvalda kuststräckor	2	0,3

Tabell 3.7 avslöjar att mer än 92 procent av tabellerna omfattar Sverige och varken mer eller mindre. Men i vissa produkter återfinns även tabeller med data, oftast från Eurostat, för andra länder. Dock varierar det vilka övriga länder som omnämns. Via Eurostat går det i många fall att få fram data för andra länder som motsvarar det som statistikprodukterna berättar om Sverige, men det rör sig främst om enklare typer av transportstatistik och trafikslagsdata som exempelvis beståndsuppgifter.

4 Sammanfattning och diskussion

4.1 Sammanfattning

Det är relativt få av de publicerade statistiktablerna som har data på länsnivå och ännu färre på kommunal nivå. Ser man istället till mikrodata är möjligheterna att presentera och analysera data på dessa lägre aggregeringsnivåer bättre, men det gäller främst för data på länsnivå. Ska data presenteras på kommunal nivå kan det bli problem med den statistiska tillförlitligheten men främst sekretessen. Det som säkert kan jämföras mellan kommuner är statistik relaterad till färdtjänstnyttjande och fordonsbestånd. Förmodligen möjliggör mikrodata jämförelser på fler områden, men för att säga vilka behövs en djupare analys. På länsnivå går förstås samma statistik som nämnts ovan att jämföra och illustrera med GIS, men dessutom ytterligare beståndsuppgifter och trafikskadedata. Det finns också med säkerhet begränsade möjligheter att jämföra data från varuflödes- och resvaneundersökningarna, men sekretessen är ett partiellt hinder. Dock kan ju dessa data generera ny, icke sekretessbehäftad, data när den används i modellsystem som till exempel Sampers och Samgods. I Varuflödesundersökningen, liksom inom sjötrafiktabellerna, finns också data aggregerat på olika regionala nivåer som möjliggör vissa jämförelser.

Det står också klart att data från bantrafik liksom televerksamhet samt merparten av luftfartsdata inte kan användas i GIS sammanhang. För Postverksamhet 2009 och statistiken över vägtrafikdöda har vissa kartor gjorts, men de är de enda möjliga; allt som kan göras med befintliga mikrodata har redan gjorts. De övriga, icke nämnda statistikområdena ligger i något av en gråzon; vissa tabeller här och där kan passa för GIS användning men för att säga någonting säkert behöver mikrodata undersökas noggrannare än vad som gjorts vid sammansättningen av tabellregistret.

4.2 Utvecklingsmöjligheter

Trafikanalys statistik utgör ett omfattande dataunderlag för analyser och beskrivningar av transportsystemet. Av historiska skäl består statistiken dock av en samling produkter utan mer gemensamt än att de berör trafik-, transport- och kommunikationsfrågor; vilket begränsar dess användbarhet för djupare och eventuellt geografiska analyser. På vissa statistikområden är det lätt att tillämpa GIS baserade analys- och illustrationsmetoder medan det på andra skulle fordras kraftiga omarbetningar. Med nya, trafikslagsövergripande statistikprodukter kan detta komma att förändras och en gemensam standard möjliggör nya sorters jämförelser och användningsområden. Kanske kunde det vara intressant att kombinera alla flöden av gods och personer till och från geografiska områden, som exempelvis städer. Städernas del i den nationella

tillväxten verkar vara något som Tillväxtanalys just nu lägger visst fokus på. Om faktiskt svenska städer kommer att tas med i någon typ av beräkning eller analys kan det säkerligen vara av intresse hur flödena av gods och människor till och från dessa städer ser ut. Något liknande har ju redan gjorts i den kartografiska analysen av interregionala resestråk som SIKAs publicerade 2009². Ett annat möjligt utvecklingsområde är jämförelser med andra länders statistik. I nuläget görs sådana jämförelser endast i begränsad omfattning och enbart inom områdena fordonsbestånd och trafikskadestatistik. Kanske kunde ytterligare jämförelser med andra länder utgöra del i en utökad omvärldsbevakning. Fler lärdomar kan säkert dras från andra länders trender, framgångar och misstag. Vissa möjligheter finns hur som helst via Eurostat, även om framställningstiden i många fall är avsevärd.

Själva tabellregistret uppdateras i nuläget löpande allt eftersom nya produkter eller produktutgåvor publiceras. Det finns många delar av det som kan raffinerats för en förbättrad funktion, men om det istället ska utvecklas med nya funktioner krävs förmodligen att registret tar steget från Excels begränsade miljö till ett mer avancerat program.

² SIKAs PM 2009:2.



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.