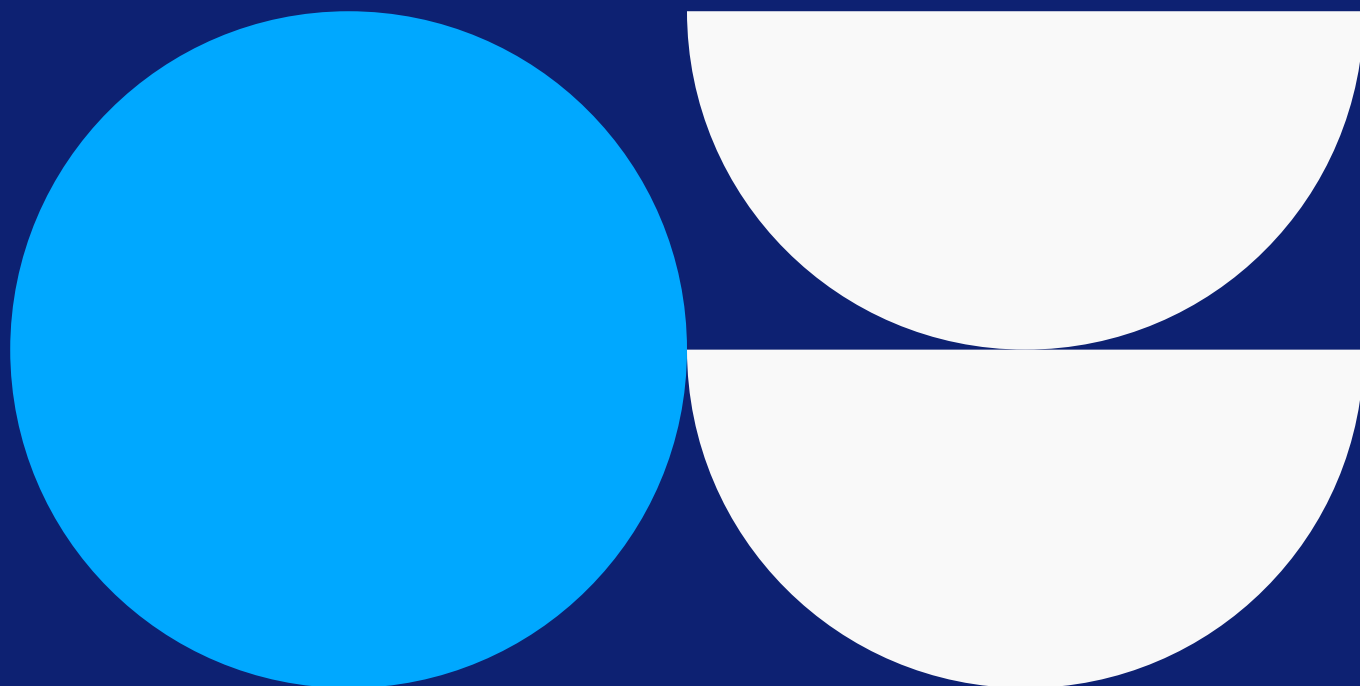


# Slutrapport från behovsinventering avseende data och infrastruktur för utbildningsvetenskaplig forskning



# **Slutrapport från behovsinventering avseende data och infrastruktur för utbildningsvetenskaplig forskning**

VR2401  
Dnr 3.1-2023-00065  
ISBN 978-91-89845-12-1

Swedish Research Council  
Vetenskapsrådet  
Box 1035  
SE-101 38 Stockholm, Sweden

## Innehållsförteckning

<b>Förord</b> .....	<b>6</b>
<b>1 Bakgrund</b> .....	<b>7</b>
1.1 Frågeställningar .....	7
1.2 Omvärldsorientering .....	8
<b>2 Tillvägagångssätt</b> .....	<b>10</b>
2.1 Dataförberedelse .....	10
2.1.1 Tillgång till data, åtkomst och sekretess .....	11
2.1.2 FAIR .....	13
<b>3 Beskrivning av datamängder inom skolan</b> .....	<b>14</b>
3.1 Tillgängliga datamängder inom skolområdet .....	14
3.1.1 Förskola, grundskola, gymnasium, kommunal vuxenutbildning och motsvarande verksamhetsformer .....	14
3.1.2 Skolinformationsutredningen .....	17
Skolverket ska ansvara för informationsförsörjningen .....	18
3.1.3 Skolinspektionens statistik .....	18
3.1.4 Statistikens tillgänglighet och trovärdighet .....	18
FoU-programmet Datadriven skolförbättring .....	19
Skolverkets och SCB:s uppfattning om Datadriven skolförbättring .....	20
3.1.5 FAIR-principerna inom skolsektorn .....	21
3.2 Brister i den nationella skolstatistiken .....	21
3.2.1 Datamängder på lokal nivå .....	23
3.2.2 Inläsning av data i verksamhetssystem .....	23
Många system försvårar interoperabilitet .....	24
Vem äger data? .....	25
Samarbete mellan huvudmän, leverantörer och forskare .....	25
3.2.3 Tillgänglighet, trovärdighet och FAIR-principerna för datamängder på lokal nivå .....	26
<b>4 Beskrivning av datamängder inom folkhögskolan</b> .....	<b>27</b>
4.1 Statistikens tillgänglighet och trovärdighet .....	28
4.1.1 FAIR-principerna inom folkhögskolesektorn .....	28
<b>5 Beskrivning av datamängder inom yrkeshögskolan</b> .....	<b>29</b>
5.1 Data om deltagare .....	29
5.2 Data om utbildningar .....	30
Yrkeshögskoleutbildningar .....	30
Konst- och kulturutbildningar .....	30
Tolkutbildningar .....	31
Statligt stöd i form av tillsyn .....	31
5.3 Statistikens tillgänglighet och trovärdighet .....	31
5.3.1 FAIR-principerna inom yrkeshögskolesektorn .....	32
<b>6 Beskrivning av datamängder inom högskolan</b> .....	<b>33</b>
6.1 Universitetskanslersämbetet .....	33
6.2 Universitets- och högskolerådet .....	34
Beda – betygsdatabas .....	34
Antagningssystemet NyA .....	35

Valda – ansöknings- och administrationssystem för validering .....	35
6.3 Högskolestatistikens tillgänglighet och trovärdighet.....	36
6.3.1 FAIR .....	36
6.4 Brister i högskolestatistiken.....	36
<b>7 Övriga relevanta datamängder .....</b>	<b>38</b>
7.1 Utbildningsregistret – UREG.....	38
7.2 Utvärdering Genom Uppföljning – UGU .....	38
7.3 Gothenburg Educational Longitudinal Database – GOLD .....	39
7.4 COMPEAT .....	39
7.5 ILSA Gateway till internationella studier .....	39
7.6 Skolvalet .....	40
7.7 Ungidag.....	40
7.8 Barnen och internet.....	40
7.9 Högskoleprovet.....	41
7.10 Tvillingregistret .....	41
7.11 Folkhälsomyndigheten.....	41
7.12 Ungdomsbarometern.....	41
7.13 Undersökning av levnadsförhållanden – ULF .....	42
7.14 IFAU .....	42
7.15 Citeringsdatabaser.....	42
<b>8 Nordisk utblick .....</b>	<b>43</b>
8.1 Danmark .....	43
8.1.1 LER.....	45
8.1.2 STIL.....	46
8.1.3 Avslutande kommentar .....	46
8.2 Finland .....	47
8.2.1 Vipunen .....	47
8.2.2 Statistikcentralen .....	48
8.2.3 Utbildningsstyrelsen .....	48
8.2.4 Undervisnings- och kulturministeriet .....	49
Övriga .....	49
8.2.5 Finlands samhällsvetenskapliga dataarkiv .....	49
8.2.6 Avslutande kommentar .....	50
8.3 Norge .....	51
8.3.1 Nationella utbildningsdatabasen .....	51
8.3.2 Database for statistikk om høyere utdanning.....	51
8.3.3 Grunnskola och gymnasium .....	52
8.3.4 Gymnasieutbildning.....	53
8.3.5 Datainsamling på högre nivå (universitet och högskolor, yrkeshögskolor) .....	53
Databehandling/kontroller .....	53
Nationell utbildningsdatabas .....	54
Datas tillgänglighet.....	54
8.3.6 Avslutande kommentar .....	54
8.4 Sammanfattning av nordisk utblick .....	54
<b>9 Behov av data och infrastruktur .....</b>	<b>55</b>
9.1 Behov av ytterligare datamängder .....	55
Öka tillgängligheten genom bättre överblick.....	55
Möjlighet att koppla undervisning till lärare .....	55
Data om elevers frånvaro.....	56

Data från nationella prov .....	56
Möjlighet att koppla register på mikronivå till internationella undersökningar .....	56
Ökad transparens från Skolverket.....	57
Möjlighet att jämföra lärosäten.....	57
9.2 Behov av infrastrukturella insatser .....	58
Register och databaser .....	58
GDPR-relaterade frågor.....	58
Behov av kompetensutveckling.....	59
Portal till utbildningsdata.....	59
Utveckla kvalitetsregister .....	59
<b>Bilaga 1 - Översikt över utbildningsstatistiska register .....</b>	<b>61</b>
Skolan .....	61
Förskolan .....	61
Grundskolan.....	61
Specialskolan .....	61
Fritidshem.....	62
Gymnasieskolan.....	62
Kommunal vuxenutbildning inkl. svenska för invandrare.....	62
Anpassad utbildning för vuxna .....	63
Lärarregistret.....	63
Folkhögskolan.....	63
Studieförbundsstatistik.....	63
Eftergymnasial yrkesutbildning.....	63
Högskolan.....	64
Befolkningens utbildning och studiedeltagande .....	64
Övergångar mellan utbildningsformer och till arbetsmarknaden .....	65
Studiestöd .....	65
Övriga undersökningar med mikrodata som SCB ansvarar för .....	65
<b>Bilaga 2 - Översikt över tillgänglig statistik .....</b>	<b>66</b>

## Förord

Vetenskapsrådet har en avgörande roll i det svenska forskningssystemet för att garantera forskningens vetenskapliga kvalitet, i termer av såväl finansiering och kontinuitet som kvalitetsvärdering. Den utbildningsvetenskapliga kommittén (UVK) inrättades 2001 och utbildningsvetenskap har under de senaste dryga 20 åren vuxit fram som ett väl etablerat forskningsfält. I sin forskningsöversikt 2023 betonade den UVK behovet av att utveckla infrastrukturer som bidrar till och stärker utbildningsvetenskaplig forskning. Det finns generellt en varierande grad av kunskap om hur till exempel registerdata kan användas inom utbildningsvetenskaplig forskning för att följa skolor, elever, syskon, föräldrar och lärare under en längre period. Dessutom finns ett växande krav på att data ska kunna göras öppet tillgänglig på ett etiskt hållbart sätt. Under 2023 genomfördes på uppdrag av UVK en kartläggning och behovsinventering avseende data samt behov av infrastrukturer för utbildningsvetenskap. Uppdraget omfattade en nationell översyn och behovsinventering av hur utbildningsdata på elev- och studentnivå, skolnivå, huvudmannanivå och nationell nivå kan göras tillgänglig, lagras och användas av forskarsamhället. Arbetet innefattade även en nordisk utblick för att samla inspiration om hur utbildningsvetenskapligt relevant data hanteras i de nordiska länderna.

Behovsinventeringen har utförts av Jan Hylén på Education Analytics. Den fångar in behovet av stöd för forskare för att använda och lagra olika former av data samt vilken data finns tillgänglig på elevnivå, skolnivå, huvudmannanivå och på nationell nivå för att användas av forskarsamhället.

Utbildningsvetenskapliga kommittén vill med detta rikta ett stort tack till Jan Hylén för det viktiga arbete som genomförts med föreliggande rapport.

Stockholm, februari 2024

Vibeke Grøver  
Ordförande UVK  
Vetenskapsrådet

Pernilla Nilsson  
Huvudsekreterare  
Utbildningsvetenskap,  
Vetenskapsrådet

# 1 Bakgrund

Vetenskapsrådets utbildningsvetenskapliga kommitté (UVK) har beslutat att genomföra en kartläggning och behovsinventering avseende data samt behov av infrastrukturer för utbildningsvetenskap. Uppdraget omfattar en nationell översyn och behovsinventering av hur utbildningsdata på elev- och studentnivå, skolnivå, huvudmannanivå och nationell nivå kan göras tillgänglig, lagras och användas av forskarsamhället på ett etiskt hållbart sätt.

## 1.1 Frågeställningar

De frågeställningar som enligt uppdragsbeskrivningen ska utredas är följande:

1. Vilken data finns tillgänglig på elevnivå, skolnivå, huvudmannanivå och på nationell nivå för att användas av forskarsamhället? Vilka ytterligare behov ser forskarsamhället av tillgängliga data för forskningsändamål?
2. Hur kan data från kommunala, regionala, nationella och privata digitala plattformar för skola och utbildning insamlas, samordnas och tillgängliggöras för att skapa underlag för forskning? Vilka eventuella inlåsningar finns inom huvudmäns avtal med Edtech-branschen, hur utbredda är de och vad finns att göra åt det?
3. Hur kan forskningsdata insamlas, lagras och tillgängliggöras på ett etiskt hållbart sätt (samordning mellan olika aktörer och huvudmän)?
4. Vilket behov av stöd behövs för forskare för att använda och lagra olika former av data?
5. Finns det andra behov av data/infrastruktur för utbildningsvetenskaplig forskning utöver registerdata? I så fall vad?

Vidare ska arbetet innefatta en nordisk utblick för att samla inspiration om hur utbildningsvetenskapligt relevant data hanteras i de nordiska länderna.

Vad är utbildningsdata? Någon entydig definition finns för närvarande inte. I följande utredning avses data på elev- och studentnivå, skolnivå respektive lärosätetsnivå, huvudmannanivå och nationell nivå. Vidare avses både individdata och verksamhetsdata. Ett preliminärt förslag till definition är: Data om förutsättningar för, genomförande och resultat av utbildningsverksamhet inom det formella och icke-formella utbildningsväsendet. Uppgifter om det informella lärandet, som sker när individen tar initiativ med syftet att förbättra sina kunskaper eller färdigheter på ett område, kan per definition inte samlas in.

## 1.2 Omvärldsorientering

Det pågår flera initiativ och utredningar inom myndighetsvärlden som är av relevans för föreliggande kartläggning. Det är utrednings- och myndighetsuppdrag som inte är specifikt inriktade mot att skapa underlag för forskning, men vars resultat på sikt kan komma att påverka tillgången till forskningsdata.

Ett viktigt initiativ är uppdraget till åtta myndigheter att utveckla en gemensam datainfrastruktur för kompetensförsörjning och livslångt lärande.<sup>1</sup> Regeringsuppdraget har getts till Arbetsförmedlingen, Skolverket, Myndigheten för yrkeshögskolan, Vetenskapsrådet, Universitets- och högskolerådet, Statistiska centralbyrån, Myndigheten för digital förvaltning (Digg) och Vinnova. Uppdraget har sex olika deluppdrag: utveckling av gemensamma begrepp, utveckling av säkra metoder för hantering av individdata, utveckling av en nationell databas för offentligt finansierade utbildningar, tillgängliggöra data om kvalifikationer, datadelning för stärkt innovationskraft samt utveckling och förvaltning av myndigheternas sammanhållna datainfrastruktur. Myndigheterna har lämnat två delrapporter från arbetet och begärt förlängning av uppdraget. Sammantaget visar delrapporterna på stor potential om uppdraget kan genomföras men att det samtidigt präglas av svårigheter och utmaningar främst ifråga om samverkan och samsyn mellan myndigheterna. Uppdraget slutredovisades i januari 2024<sup>2</sup>.

Ett annat relevant uppdrag avser en utredning om utvecklad reglering och styrning av interoperabilitet vid datadelning inom den offentliga förvaltningen och från den offentliga förvaltningen till externa aktörer.<sup>3</sup> En särskild utredare ska analysera befintlig styrning och reglering av interoperabilitet vid datadelning inom den offentliga förvaltningen och från den offentliga förvaltningen till externa aktörer. Utredaren ska utifrån analysen föreslå om och hur styrning och reglering av interoperabilitet bör och kan utvecklas. Syftet är att öka den offentliga förvaltningens förmåga att samverka digitalt genom att dela data på ett effektivt och säkert sätt samtidigt som den personliga integriteten värnas och säkerheten i systemen och i samhället förblir robust. Uppdraget redovisades i december 2023.<sup>4</sup>

Vidare kan arbetet inom Forum för informationsstandardisering i skolväsendet (FFIS) nämnas.<sup>5</sup> Initiativet är en följd av att Skolverket sedan 2020 blivit sektorsansvarig myndighet för skolans digitalisering. I FFIS möts branschföreträdare, företrädare för Sveriges Kommuner och Regioner (SKR) och Skolverket samt Research Institutes of Sweden (RISE) och företrädare för standardiseringsarbetet inom skolsektorn. Forumets arbete syftar till att främja

---

<sup>1</sup> [Uppdrag att utveckla en sammanhållen datainfrastruktur för kompetensförsörjning och livslångt lärande](#)

<sup>2</sup> [Slutredovisning av uppdraget \(pdf\)](#)

<sup>3</sup> [Utvecklad reglering och styrning av interoperabilitet vid datadelning inom den offentliga förvaltningen och från den offentliga förvaltningen till externa aktörer](#)

<sup>4</sup> [En reform för datadelning \(SOU 2023:96\)](#)

<sup>5</sup> [Forum för informationsstandardisering i skolväsendet](#)



interoperabilitet och därmed möjliggöra ett säkert och effektivt informationsutbyte inom och mellan huvudmän samt mellan huvudmän och andra aktörer som myndigheter och leverantörer av digitala tjänster. Syftet är också att undvika framtida inlåsningar i specifika teknologier eller leverantörer. Det faktum att Skolverket är drivande i standardiseringsfrågorna har bland annat inneburit att Svenska institutet för standarder (SIS) för några år sedan kunnat fastställa en standard kallad SS12000.<sup>6</sup> Den ger stöd för processer som skoladministration, tjänstefördelning, schemaläggning, närvarohantering och betygshantering. Skolverket rekommenderar huvudmän och leverantörer att implementera denna standard.

I vilken utsträckning SS12000 är implementerad kommer sannolikt att uppdagas i ett nytt uppdrag som SCB fått. SCB skulle genomföra en undersökning om möjligheterna att samla in registerdata om skolverksamhet direkt från systemleverantörer.<sup>7</sup> I undersökningen ingick att inhämta synpunkter från forskare och andra användare om vilka behov av data om skolverksamhet de har. Resultatet avrapporterades i januari 2024.<sup>8</sup>

---

<sup>6</sup> [Samlad information om standarden SS 12000](#)

<sup>7</sup> [Regleringsbrev för budgetåret 2023 avseende anslag 4:4 inom utgiftsområde 16 Utbildning och universitetsforskning](#)

<sup>8</sup> [SCN: Förutsättningar för att samla in nya skoldata från systemleverantörer](#)

## 2 Tillvägagångssätt

Arbetet har genomförts i form av intervjuer med myndigheter, forskare, organisationer och skolhuvudmän. Detta har kompletterats med dokumentstudier.

Berörda myndigheter är Skolverket, Skolinspektionen, Specialpedagogiska myndigheten (SPSM), Statistikmyndigheten (SCB), Centrala studiestödsnämnden (CSN), Myndigheten för yrkeshögskolan (MYH), Universitetskanslersämbetet (UKÄ), Universitets- och högskolerådet (UHR) och Vetenskapsrådet (VR). Folkbildningsrådet är egentligen inte en myndighet men har vissa myndighetsuppgifter som att fördela statsbidrag till studieförbund och folkhögskolor samt att ta in statistik om folkhögskolornas verksamhet. Vidare har SKR en central roll som företrädare för samtliga kommuner som tillika är centrala skolhuvudmän. Branschorganisationen Swedish Edtech Industry företräder de företag som levererar it-system och läromedel till utbildningssektorn.

Samtliga uppräknade myndigheter har intervjuats. I flera fall har flera intervjuer gjorts med flera företrädare för samma myndighet. Likaså för SKR och Swedish Edtech Industry. Ytterligare några myndigheter har tillfrågats per mail i olika ärenden. Fem skolhuvudmän och tio forskare har intervjuats. Forskarna har medvetet hämtats från olika discipliner som pedagogik, nationalekonomi, data- och systemvetenskap, informatik, utbildningsteknologi med mera. Förutom att intervjuas forskare från olika discipliner så finns även olika forskningsansatser representerade bland intervjupersonerna. Både kvantitativt inriktade forskare och forskare som arbetar med kvalitativa data har intervjuats. Sammantaget har mellan 35–40 personer intervjuats muntligt eller per mail.

### 2.1 Dataförberedelse

Möjligheterna att använda data ökar ständigt. Nya beräkningsmetoder utvecklas, datorernas processorkapacitet växer och visualiseringsmöjligheterna ökar. Allt mer data eftersöks för strategiskt beslutsfattande, forskning och andra ändamål. Men det vilar en förrädisk enkelhet över att arbeta med data. Innan data matas in i en formel, en algoritm eller liknande måste man ha kunskap om i vilket skick datamängden befinner sig. Emellanåt har de som står bakom insamlingen av data mycket grund förståelse för komplikationerna med att samla in, bearbeta och bevara data. Utmaningarna omfattar bland annat: dåliga metoder för datainsamling, saknade värden, obekväma lagringsmekanismer, immateriella rättigheter, uppfyllandet av säkerhetskrav och hantering av sekretess. Alla dessa aspekter hindrar delning och samkörning av data.

För att underlätta diskussionen om det som ibland med ett övergripande begrepp benämns som datakvalitet så har den brittiska forskaren inom maskinlärande N. D. Lawrence (2017) utvecklat ett resonemang om tre aspekter, eller möjligen nivåer, av dataförberedelse.<sup>9</sup>

På den mest grundläggande nivån, nivå C, behandlas datas tillgänglighet. Här ställs frågor som: finns data, var finns de och går det att få tillgång till datamängden? Vilket format har datamängden (Excel, Word, PDF)? Tillåter formatet vidare bearbetning utan omfattande extraarbete? Finns det etiska aspekter samt legalitets- och integritetsaspekter att ta hänsyn till? Föreligger det tekniska svårigheter för att få tillgång till data? Är de till exempel spridda över flera system som kräver arbetstid för att sammanföra?

När data uppfyller kraven i nivå C är mängden redo att laddas in i analysprogram, eller göras tillgängliga för andra. Datasetet är då maskinläsbart och etiska förfaranden för datahantering har tagits upp.

Nästa nivå, nivå B, handlar om datas trovärdighet och representation. Nu när datamängden har laddats in i programvaran, motsvarar det som registreras det som påstås vara registrerat? Hur hanteras saknade värden, hur har data kodats? Hur är bruset karakteriserat (för sensorer) eller finns det fel i inmatningen vid eventuell manuell hantering?

När kraven i nivå B är uppfylld vet analytikern hur trogna uppgifterna är mot det datamängden beskriver. Detta är vanligen den stora utmaningen för en forskare – skillnaden mellan det människor tror att de har i sina datamängder jämfört med vad som är faktiskt finns där. Det är först i slutet av nivå B som analytikern börjar få en blick för trovärdigheten och vad som verkligen är möjligt att göra med datamängden.

Den sista nivån, nivå A, handlar om sammanhang och relevans. Hur väl ämnad är datamängden för att besvara en given frågeställning med en vald analysmetod? Denna nivå är inte generell utan sammanhangsspecifik och kan enbart besvaras av den forskare som överväger att använda en viss datamängd. Den kommer därför inte vidare att beröras i följande text.

Vid en genomgång av vilken data som finns tillgänglig för forskning i utbildningsvetenskapliga frågor kan ovanstående resonemang ge en god utgångspunkt. De datakällor som identifieras i föreliggande rapport kommer därför att i görligaste mån karakteriseras utifrån tillgänglighet (nivå C) och trovärdighet (nivå B).

### **2.1.1 Tillgång till data, åtkomst och sekretess**

Möjligheten att ta del av data från offentliga myndigheter utgår ifrån den grundlagsskyddade offentlighetsprincipen. Denna garanterar en rätt till insyn för att möjliggöra undersökning och bevakning av myndigheternas arbete. En del av offentlighetsprincipen är den så kallade handlingsoffentligheten. Den innebär en

---

<sup>9</sup> [Lawrence, N. D. \(2017\). Data Readiness Levels](#)

rätt att ta del av allmänna handlingar som är offentliga. Det finns dock begränsningar i offentlighetsprincipen, dessa begränsningar regleras i offentlighet- och sekretesslagen, OSL, och har utformats som undantag till handlingsoffentligheten. Det innebär alltså att vissa typer av handlingar betraktas som att de omfattas av sekretess och därför inte kan lämnas ut. Bedömningen av detta är vad som vanligen kallas för sekretessprövning.

Inom ramen för utbildningsvetenskap torde sekretessregler av intresse röra den så kallade statistiksekretessen och den s.k. dataskydd (eller GDPR) sekretessen.

Statistiksekretessen innebär att uppgifterna i Statistiska Centralbyråns (SCB:s) register är belagda med absolut sekretess men data kan lämnas ut för ändamålet forskning efter en särskild prövning vilket framgår av offentlighets- och sekretesslagen (2009:400) (OSL)<sup>10</sup>. För registerforskare innebär detta alltså att absolut sekretess är huvudregeln för ett flertal av de vanligaste källorna till registerdata. Ett undantag i OSL anger dock att registerdata som är belagda med absolut sekretess trots detta får lämnas ut för just forskningsändamål om det efter en så kallad men prövning står klart att de kan röjas utan att den enskilde eller någon närstående till denne lider skada eller men. Det här gäller alltså individdata, sådan data som direkt eller indirekt kan knytas till en person, aggregat data, eller med andra ord statistik omfattas inte av sekretessen.

Den så kallade dataskyddssekretessen innebär att utlämnande myndighet inte får lämna ut begärd data om den innehåller personuppgifter och de finns anledning att anta att den begärande kommer att använda personuppgifter på ett sätt som står i strid med dataskyddsförordningen (GDPR).

Vad gäller dataskydd, eller GDPR, så måste forskare generellt vara noga med att beakta dataskyddslagar och etiska riktlinjer för att skydda individers integritet. Det innebär bland annat att personuppgifter som kan identifiera individer inte får användas utan rättslig grund, t.ex. samtycke, och om sådana uppgifter används, bör det alltid övervägas om de lika gärna kan anonymiseras eller pseudonymiseras. Dessutom bör forskningsprojekt genomgå etisk granskning för att säkerställa att de följer gällande normer och lagar, beroende på vilket typ av data som forskningen innefattar kan det i vissa fall vara ett krav att söka etiskt prövningstillstånd enligt Etikprövningslagen. Det är också viktigt att transparent kommunicera syftet med forskningen, hur data kommer att användas, och vilka säkerhetsåtgärder som vidtas för att skydda informationen.

---

<sup>10</sup> OSL 24 kap. 8 §: ”Sekretess gäller i sådan särskild verksamhet hos en myndighet som avser framställning av statistik för uppgift som avser en enskilds personliga eller ekonomiska förhållanden och som kan hänföras till den enskilde. [...] Uppgift som behövs för forsknings- eller statistikändamål och uppgift som inte genom namn, annan identitetsbeteckning eller liknande förhållande är direkt hänförlig till den enskilde får dock lämnas ut, om det står klart att uppgiften kan röjas utan att den enskilde eller någon närstående till denne lider skada eller men. Detsamma gäller en uppgift som avser en avliden och som rör dödsorsak eller dödsdatum, om uppgiften behövs i ett nationellt eller regionalt kvalitetsregister enligt patientdatalagen (2008:355).”

Att skydda personers integritet när det gäller användning av data innefattar flera viktiga principer och åtgärder som inhämtning av samtycke, minimering av datainsamling, anonymisering och pseudonymisering av data, säker lagring och hantering av data, transparens om hur data samlas in, lagras och används, rätt till radering och i vissa fall återkallande av samtycke samt etisk granskning.

### 2.1.2 FAIR

Ett annat sätt att karakterisera data används ofta av förespråkare för öppna data. FAIR är en akronym som står för Findable, Accessible, Interoperable och Reusable.<sup>11</sup> FAIR-principerna innebär att forskningsdata ska gå att hitta, det ska finnas information om hur man får tillgång till dem, de ska vara kompatibla med andra data, och de ska vara möjliga att återanvända. FAIR-principerna spelar en viktig roll i arbetet för öppen vetenskap och beskriver några av de mest centrala riktlinjerna för god datahantering och öppen tillgång till forskningsdata.

Svensk Nationell Datatjänst, som har i uppdrag att skapa möjligheter för forskare att beskriva, dela och återanvända forskningsdata, arbetar med att utveckla tjänster och verktyg som följer FAIR-principerna.

De två modellerna har delvis olika fokus. FAIR-principerna fokuserar på öppenhet och tillgänglighet, men säger mindre om trovärdigheten i data. Ett sannolikt skäl till detta är att FAIR-principerna i första hand tycks användas på dataset som härstammar från tidigare forskningsprojekt. Därmed räknar man med att trovärdighetsfrågan har lösts av det tidigare forskningsprojektet.

Eftersom föreliggande genomgång av datamängder av relevans för utbildningsvetenskaplig forskning kommer att omfatta en rad datakällor som inte härstammar från tidigare forskningsprojekt, så förefaller FAIR-principerna mindre relevanta i detta fall. Men i gärligaste mån kommer olika datamängders uppfyllande av FAIR-principerna också att diskuteras.

---

<sup>11</sup> [FAIR-principerna](#)

## 3 Beskrivning av datamängder inom skolan

Data från utbildningssektorn är i de flesta fall indelad efter utbildningsform. Därför följer denna redovisning samma form. Även om listan över datamängder är relativt omfattande går det inte att påstå att den är uttömmande. Det gäller framför allt olika datamängder hos skolhuvudmän. En mängd variabler, framför allt i den officiella statistiken, beskrivs också men inte heller här är listan fullständig.

Den statistik som myndigheter inom utbildningssektorn samlar in kan sägas vara av två slag: antingen sker insamlingen för att Utbildningsdepartementet ålagt myndigheten att vara statistikansvarig (och mer eller mindre i detalj bestämt vad som ska samlas in), eller så sker annan insamling för att myndigheten ska kunna fullgöra sitt uppdrag. Någon allmän datainsamling för att underlätta forskning i största allmänhet förekommer inte. För att något sådant ska kunna ske skulle myndigheternas instruktion sannolikt behöva ändras så att främjande av forskning blir en del av deras kärnuppgifter. I den meningen föreligger det en viss skillnad i uppdrag och behov mellan myndigheter och forskare.

### 3.1 Tillgängliga datamängder inom skolområdet

#### 3.1.1 Förskola, grundskola, gymnasium, kommunal vuxenutbildning och motsvarande verksamhetsformer

Skolverket är utsett av regeringen att ansvara för Sveriges officiella statistik om skolväsende och barnomsorg.<sup>12</sup> Den officiella statistiken redovisas i de flesta fall en gång per år. Officiell statistik ska finnas för allmän information, utredningsverksamhet och forskning. Som statistikansvarig myndighet har Skolverket bland annat ansvar för att statistiken är objektiv, att den dokumenteras samt att den är kvalitetsdeklarerad. Den officiella statistiken ska utan avgift finnas tillgänglig i elektronisk form. Skolhuvudmännen har uppgiftslämnarplikt.

Skolverket publicerar årligen en kvalitetsdeklaration för varje underområde där objekt, population och variabler beskrivs. Statistikens tillförlitlighet och eventuella osäkerhetskällor diskuteras också.<sup>13</sup> Här behandlas även frågor om sekretess och personuppgiftsbehandling.

Utöver officiell statistik publicerar Skolverket även annan statistik. Den härrör vanligen från specifika insamlingstillfällen och har sällan längre tidsserier.

---

<sup>12</sup> [Om Skolverkets statistik](#)

<sup>13</sup> Se exempelvis [SCB: KVALITETSDEKLARATION Grundskolan: elever per 15 oktober \(pdf\)](#)

Dessa datamängder publiceras vanligen i rapportform som PDF-filer, men kan ibland även finnas tillgängliga i Excel-format.

För att ge inblick i hur omfattande insamlingen av statistik är kommer här tre exempel att ges. I övrigt hänvisas till Bilaga 1 och 2. Exempel ett och två avser uppgifter om elever och det tredje uppgifter om skolans personal. De uppgifter som årligen rapporteras in om elever är elevers resultat i form av terminsbetyg i åk 6, slutbetyg i åk 9, resultat från nationella prov och delprov samt meritvärde. Vidare rapporteras bakgrundsinformation som kön, födelseland, om man är nyinvandrad (folkbokförda i Sverige under de senaste fyra åren), föräldrars födelseland och föräldrars utbildningsnivå. Slutligen rapporteras också en del data om elevens skolsituation som enskild undervisning, särskild undervisningsgrupp, anpassad studiegång, studiehandledning på modersmålet samt åtgärdsprogram.

Det tredje exemplet hämtas från kommunal vuxenutbildning. Avseende kommunal vuxenutbildning så samlas en hel del information in utöver rena resultatdata. Kommunal vuxenutbildningsstatistiken avseende elever har följande variabler<sup>14</sup>:

- Personnummer
- Kurskod
- Utbildningsnivå
- Poäng för kursen
- Kursdeltagarens genomförda verksamhetspoäng för kursen
- Kursstart
- Kurslut
- Kursens förläggning
- Utbildningsanordnare
- Betyg
- Orsak till att betygsvärde ej finns
- Avbrott
- Avbrottsdatum
- Prövning/validering

Listan över datavariabler kopplade till personal av olika kategorier är än mer omfattande än uppgifterna om elever. Den omfattar:

- Antal lärare
- Kön
- Ålder
- Högsta utbildningsnivå
- Född utomlands/född i Sverige
- Lärare med pedagogisk högskoleexamen
- Lärare med specialpedagogisk högskoleexamen
- Antal år personen funnits som lärare i registret över pedagogisk personal
- Tillsvidareanställd

---

<sup>14</sup> [SCB: Data från Komvuxregistret](#)

- Typ av tjänst (lärare, rektor, övrig skolledare, speciallärare, specialpedagog, yrkes- och studievägledare, studiehandledare på modersmål, lärarassistent, fritidspedagog, fritidsledare eller annan pedagogisk personal)
- Skolform där personalen är verksam
- Skolenhet där personalen är verksam
- Kommun och län där skolenheten är belägen
- Tjänstgöringsgrad omräknad till heltidstjänstgöring
- Anställningsform (tillsvidareanställning eller tidsbegränsad anställning)
- Undervisningsämnen och årskurser där undervisningen bedrivs (för samtliga skolformer förutom utbildning i svenska för invandrare, förskoleklass och fritidshem)
- Tjänstledighet (om personen är helt tjänstledig längre än 30 dagar)
- Utbildningsanordnare (för kommunal vuxenutbildning och kommunal vuxenutbildning i svenska för invandrare)

Data om lärare är baserad på personnummer, vilket innebär att det finns mikrodata att begära ut från SCB.

Den mest samlade information om vilka uppgifter som samlas in finns i Skolverkets ”Föreskrifter om uppgiftsinsamling från huvudmännen inom skolväsendet”.<sup>15</sup> Där anges tydligt vilka uppgifter skolhuvudmännen har skyldighet att rapportera in till Skolverket för olika verksamhetsformer och hur ofta det ska ske. De flesta uppgifter samlas in årligen men vissa enbart var tredje verksamhetsår. Det sistnämnda gäller till exempel uppgifter om elevhälsopersonal. I en bilaga finns också definitioner för samtliga begrepp. Däremot listas inte samtliga variabler.

SCB har en del mikrodatabaser på utbildningsområdet. Tillgängliga register beskrivs på webben.<sup>16</sup> Listan är inte fullständig men ger en överblick över vad som finns.<sup>17</sup> Går man in på underrubriken Skolväsende och barnomsorg så återfinns all skolstatistik. Under rubriken Dokumentation hittar man information om respektive register.<sup>18</sup> Se vidare Bilaga 1.

Skolverkets redovisning av statistiken återfinns i en databas kallad ”Jämförelsetal” fördelat efter verksamhetsform samt kommun eller län.<sup>19</sup> Databasen går tillbaka till 1992 för de verksamhetsformer som fanns redan då och från de år de startade för övriga. Det innebär till exempel att data om förskoleklass finns först från och med 1998. Utöver utbildningsdata finns även ett antal strukturvariabler kopplade till kommunen som andel barn i bidragshushåll, åldersstruktur på innevanarna, skattekraft med mera.

Skolans kostnader redovisas uppdelade efter ett antal kostnadslag som undervisning, lokaler, skolmåltider, lärverktyg, elevhälsa, elevavgifter och

---

<sup>15</sup> [Senaste lydelse av SKOLFS 2011:142](#)

<sup>16</sup> [SCB: Data från individregister](#)

<sup>17</sup> [SCB: Utbildning och forskning](#)

<sup>18</sup> Se till exempel [SCB: Grundskolan: elever per 15 oktober](#)

<sup>19</sup> [Skolverket: Sök statistik om förskola, skola och vuxenutbildning](#)



övrigt.<sup>20</sup> Det är endast de kommunala och statliga huvudmännens kostnad som redovisas.

Den officiella statistiken redovisas i aggregerad form och den tillgängliggörs i Excel-format samt som CSV eller XML.<sup>21</sup> Mikrodata, det vill säga uppgifter om enskilda objekt som personer, är inte tillgängligt via Skolverkets webb utan måste beställas av SCB och kräver en sekretessprövning innan utlämnande.<sup>22</sup>

Ett undantag från detta är Skolenhetsregistret som har mikrodata i form av uppgifter om lägeskommun, huvudmannatyp, inriktning för skolenheter med enskild huvudman, huvudmannens organisationsnummer och juridiska form.<sup>23</sup> Registret innehåller också uppgifter om förekommande årskurser för skolenheter med skolformerna grundskola, sameskola, grundsärskola och specialskola samt förekommande program för skolenheter med skolformerna gymnasieskola och gymnasiesärskola. Registret innehåller även uppgifter om skolenheter med skolformen grundskola som begränsar sitt mottagande till elever i behov av särskilt stöd (så kallade resursskolor) eller är sjukhusskolor. Registret har nyligen överförs från SCB till Skolverket.

### 3.1.2 Skolinformationsutredningen

I föreliggande uppdrag ingår att följa Skolinformationsutredningens arbete (SOU 2023:21).<sup>24</sup> Utredningen avlämnade ett delbetänkande i juni 2023. Där föreslås att:

- Skolverket ska ansvara för informationsförsörjningen på skolområdet,
- Skolverket ska bedriva uppdragsverksamhet,
- andra myndigheter och enskilda utbildningsanordnare inom universitets- och högskoleutbildning ska lämna uppgifter till Skolverket samt att Skolverket ska lämna uppgifter till Skolinspektionen och SCB,
- nya bestämmelser om behandling av personuppgifter i Skolverkets verksamhet med informationsförsörjning,
- en ny sekretessbestämmelse till skydd för uppgifter om enskilds personliga förhållanden i Skolverkets verksamhet med informationsförsörjning, och
- en ny förordning om Skolverkets informationsförsörjning.

Förslagen syftar bland annat till att myndigheter och kommuner ska kunna utföra sina författningsreglerade uppgifter genom att de får tillgång till information även från fristående skolhuvudmän och att föräldrar och elever ska ha god tillgång till information inför skolval.

---

<sup>20</sup> [SCB: KVALITETSDEKLARATION - Kostnader för förskola, fritidshem, skola och vuxenutbildning](#)

<sup>21</sup> [Skolverket: Sök statistik om förskola, skola och vuxenutbildning](#)

<sup>22</sup> [SCB: Vad är MONA](#)

<sup>23</sup> [Skolverket: Skolenhetsregistret](#)

<sup>24</sup> [Informationsförsörjning på skolområdet - Skolverkets ansvar \(SOU 2023:21\)](#)

### ***Skolverket ska ansvara för informationsförsörjningen***

I delbetänkandet föreslås att Skolverket ska ta över insamlingarna av skolinformation från SCB. Genom att Skolverket blir den myndighet som sköter insamlingen så omfattas inte de insamlade uppgifterna av offentlighets- och sekretesslagens bestämmelser om sekretess för enskilda ekonomiska förhållanden. Det innebär att uppgifter om enskilda huvudmäns verksamhet och fristående förskolor och skolor bör kunna publiceras. Enligt utredningens förslag bedöms Skolverket kunna ta över insamlingen från SCB från och med den 1 januari 2026. Det framgår inte av delbetänkandet om retroaktiva data från den period då Skolverket inte fick publicera informationen kommer att kunna tillgängliggöras.<sup>25</sup>

#### **3.1.3 Skolinspektionens statistik**

Skolinspektionen har inget statistikansvar men samlar in viss statistik av relevans för myndighetens inspektionsroll. Så genomför man årligen en stor enkätundersökning riktad till elever, vårdnadshavare och pedagogisk personal i grundskolan. Enkäten behandlar frågor om trygghet och studiero.<sup>26</sup> Den besvaras årligen av cirka 360 000 personer. Enkäten har genomförts sedan år 2000 med vissa förändringar över tid. All data från undersökningen finns tillgänglig i Excel-format på webben från 2015 och framåt men myndigheten skriver också årligen en sammanfattande rapport.

Vidare publiceras statistik över inspektionens tillsynsverksamhet. Det gäller så kallad regelbunden tillsyn, där statistiken visar hur många skolor som granskats och vilka olika brister som upptäckts.<sup>27</sup> Statistiken redovisas i PDF-format och går tillbaka till 2008. Se vidare Bilaga 2.

#### **3.1.4 Statistikens tillgänglighet och trovärdighet**

Den officiella statistiken om skolväsendet måste sägas ha hög tillgänglighet, så länge statistik från de kommunala skolhuvudmännen avses. För fristående skolhuvudmän gäller att de har samma skyldighet att lämna uppgifter till Skolverket som kommunala huvudmän, men Skolverket och SCB lämnar för närvarande inte ut denna data efter en nytolkning av sekretesslagstiftningen i september 2020. Som framgick av tidigare avsnitt föreslår en utredning nu en nyordning som innebär att uppgifter från fristående huvudmän åter ska kunna publiceras sannolikt från och med 2026.

Förutom att redovisas på Skolverkets webb så kan den officiella statistiken även återfinnas hos SCB. Vidare finns ett verktyg kallat Kolada för att underlätta för kommuner och regioner att jämföra sig över tid eller med andra huvudmän i olika avseenden.<sup>28</sup> Bakom Kolada står Rådet för främjande av kommunala

---

<sup>25</sup> [Remissvar från bland andra Skolverket och SCB](#)

<sup>26</sup> [Skolinspektionen: Resultat från Skolenkäten](#)

<sup>27</sup> [Skolinspektionen: Statistik om regelbunden tillsyn](#)

<sup>28</sup> [Rådet för främjande av kommunala analyser \(RKA\): Kolada - en tjänst från RKA](#)

analyser, RKA, som är en ideell förening som bildats i samarbete mellan staten och Sveriges Kommuner och Regioner, SKR.

Kolada består av statistik från Skolverket, Skolinspektionen och SCB. Här finns ytterst begränsat med unika eller nya data jämfört med vad de tre myndigheterna själva publicerar. Undantaget är en enkät till samtliga förskollärare. Detta eftersom Skolinspektionens stora skolenkät inte går ut till förskolan.

Vad gäller trovärdigheten så publicerar Skolverket, som tidigare nämnts, årligen kvalitetsdeklarationer för de olika statistikområdena.<sup>29</sup> Där resoneras bland annat om uppgifternas tillförlitlighet, vilka osäkerhetskällor som kan finnas och hur stort bortfallet kan tänkas vara. Den övergripande slutsatsen i samtliga kvalitetsrapporter är att tillförlitligheten anses hög. Men en viktig brasklapp i sammanhanget är att Skolverkets resonemang berör verkets egen hantering av inrapporterade uppgifter – inte hur insamlingen och inrapporteringen går till hos skolhuvudmännen.

Här visar både intervjuer med skolhuvudmän och tidigare erfarenhet att med många uppgiftslämnare och med uppgiftslämnare som kan ha olika bild av vad de olika begreppen står för, samt hur inrapporterade data ska användas, så ökar felkällorna. Skolledare har ofta ansvaret för att samla in skolenhetens uppgifter för vidare distribution till förvaltningsnivån, men det är inte ovanligt att rektor delegerar denna uppgift till skoladministratörer med lägre förståelse för och dålig insyn i hur uppgifterna ska användas. Vissa uppgifter, som deltagande i särskilda undervisningsgrupper, anpassad studiegång och liknande, svarar ofta speciallärare för att rapportera in. Även om ”Skolverkets föreskrifter om uppgiftsinsamling från huvudmännen inom skolväsendet” (SFS 2011:142) innehåller definitioner av de flesta begrepp så når dessa beskrivningar inte alltid längst ut i verksamhetens kapillärer. Inte minst är skolans kostnader ett område där olika huvudmän har olika praxis runt hur man beräknar och redovisar kostnader. Det tydligaste exemplet är att olika kommuner har olika sätt att beräkna sina lokalkostnader.

### ***FoU-programmet Datadriven skolförbättring***

I ett FoU-program drivet av Ifous har sex skolhuvudmän tillsammans med forskare från Stockholms universitet genomfört ett treårigt program för datadriven skolutveckling.<sup>30</sup> Programmets mål var att undersöka och förstå hur datadrivna metoder och arbetssätt kan användas för att förbättra undervisning och lärande i grund- och gymnasieskolan.

Några av de resultat som kan nämnas är att deltagarnas bristande förmåga att formulera relevanta frågeställningar, att samla och analysera data, att bedöma datakvalitet samt att använda digitala verktyg och system utgjorde ett betydande

---

<sup>29</sup> Se till exempel [SCB: KVALITETSDEKLARATION - Grundskolan: elever per 15 oktober \(pdf\)](#)

<sup>30</sup> [Ifous: Datadriven skolförbättring](#) De skolhuvudmän som deltog i projektet var Eskilstuna kommun, Fridaskolorna, Nyköping Strand Utbildningscentrum, Tyresö kommun, Salems kommun och Värmdö kommun

hinder för att uppnå programmålen. Vidare konstateras att huvudmännens data var utspridda på flera digitala system och verktyg, vilket gjorde det svårt att aggregera och sammanställa informationen. Det krävdes ofta teknisk kompetens, som deltagarna inte alltid hade tillgång till, och i vissa fall var det helt enkelt inte genomförbart att samla ihop alla data på ett effektivt sätt. Dessutom fanns det problem med att extrahera data ur vissa system, antingen för att deltagarna själva inte ägde och hade kontroll över data, eller för att systemen inte var uppbyggda för att möjliggöra dataexport. Det faktum att leverantörer inte hade implementerat datastandarderna SS 12000 försvårade även nödvändiga dataflöden för aggregering av data.

Deltagarna upplevde även att datakvaliteten ofta var låg, vilket yttrade sig i form av saknade data, felaktiga data eller alltför få data. I många fall fanns data tillgänglig för en önskad variabel, men i fel detaljnivå (för låg eller för hög abstraktionsnivå) eller i otillräcklig mängd, vilket påverkade analys- och aktionsmöjligheterna på ett negativt sätt.

Vidare konstaterar man att brist på samarbete och kommunikation mellan skolans olika aktörer, som lärare, rektorer, administratörer, IT-stöd och förvaltningar, hämmade en effektiv implementering av datadrivna metoder och skapade hinder för en fungerande datadriven skolorganisation.

En av projektets mest centrala lärdomar är att skolor och huvudmän i hög utsträckning arbetar utifrån en kultur och en metodik som primärt premierar rapportering uppåt i systemet snarare än datadriven skolförbättring – rapportering från lärare till rektorer, rektorer till förvaltningar och förvaltningar till politiska nämnder och myndigheter. Vidare innebär mätinstrumentens begränsningar och konstruktion tydliga hinder. Mätningarna är ofta lågfrekventa (en gång per termin eller år, eller i vissa fall enbart vart tredje år) och genomförs ofta bristfälligt.

Sammantaget konstaterar följeforskarna i programmet att det finns betydande svagheter i skolhuvudmännens kunskap om och arbete med informationsinsamling och -hantering. FoU-programmets slutsatser ligger med andra ord helt i linje med de brister som påtalats ovan.

### ***Skolverkets och SCB:s uppfattning om Datadriven skolförbättring***

SCB menar att den stora merparten av det statistiska materialet håller bra kvalitet. Man menar att skolorna rimligtvis har korrekta uppgifter om vilka elever som finns på skolan och vilka betyg de får. Här är utrymmet att misstolka ganska litet. Givetvis finns det statistik där tolkningsutrymmet är större, säger man. Skolverket svarar skriftligt att ”Kvaliteten i de lämnade uppgifterna bedöms som hög på kommunnivå och riksnivå. Det är alltid uppgiftslämnaren som ansvarar för att de lämnade uppgifterna är korrekta.”

En påtaglig brist i sammanhanget är att det för närvarande saknas en definition i skollagen av giltig och ogiltig frånvaro.<sup>31</sup> Skolan bedömer om en elevs frånvaro

---

<sup>31</sup> [Skolverket: Frånvaro i skolan](#)

är giltig eller inte. Detta gäller inte bara grundskolan där skolplikt råder, utan även gymnasieskolan där skolenheter är skyldiga att rapportera in ogiltig frånvaro till SCB, eftersom för mycket frånvaro kan leda till indragen studiehjälp från CSN. I intervjun med CSN framkom också att myndigheten antar att olika skolhuvudmän har olika syn på vad som ska rapporteras som frånvaro.<sup>32</sup>

För att ge en nationell bild av frånvarosituationen genomförde Skolverket 2021 en enkätundersökning till skolhuvudmännen.<sup>33</sup> Där konstateras att skolhuvudmännen av olika skäl inte alltid rapporterar all frånvaro. Av intervjun med CSN framgick att myndigheten tror sig veta, utan att kunna belägga påståendet, att olika huvudmän har olika policy för när de rapporterar att en elev är frånvarande. Vissa huvudmän tycks systematiskt underrapportera. Ur forskningssynpunkt är det olyckligt att en så central variabel i relation till skolresultat och studieavhopp inte är bättre säkerställd.

En annan variabel där många huvudmän vittnar om att det är svårt att jämföra sig med andra huvudmän är barngruppernas storlek i förskolan. Det saknas tydliga riktlinjer om vilken tidpunkt på dagen antalet barn ska räknas och eftersom antalet barn i grupperna varierar över dagen blir statistiken mindre tillförlitlig.

### 3.1.5 FAIR-principerna inom skolsektorn

Den skolstatistik som finns på nationell nivå tycks i hög grad följa FAIR-principerna. Data går att hitta (Findable), den är till största delen tillgänglig (Accessible), både data och metadata följer vedertagna standarder och data publiceras i filformat som stöds av olika operativsystem och kan öppnas i flera program. Slutligen beskrivs data med tillräckliga metadata som kan läsas av både människor och datorer, och att det finns tydliga upplysningar om i vilket sammanhang data samlades in och villkor för hur data får användas måste anges tydligt. Därmed uppfylls även kravet på återanvändbarhet (Reusable).

Forskarintervjuerna visar dock att inte alla forskare delar bilden att det är enkelt att hitta i den officiella statistiken inom skolområdet. Det upplevs som svårt att veta vilka variabler som finns och var man hittar mikrodata. Som ett svar på denna kritik har SCB utarbetat en handledning för forskare inom utbildningsvetenskap om hur man orienterar sig på SCB:s webb. Denna återfinns som Bilaga 1.

## 3.2 Brister i den nationella skolstatistiken

De brister som kan skönjas i den nationella statistiken från skolväsendet är av tre slag. Det gäller bristande trovärdighet inom vissa statistikområden, bristande

---

<sup>32</sup> Intervju 2023-03-15

<sup>33</sup> [Skolverket: Nationell kartläggning av elevfrånvaro](#)

tillgänglighet till data från fristående skolhuvudmän och slutligen områden där för lite statistik samlas in.

Det första området berördes ovan. Åtgärder som sätts in för att komma tillrätta med kvalitetsbrister relaterade till kompetensbrist runt insamling och inrapportering av data hos skolhuvudmän, är framför allt kompetensutvecklingsinsatser som RKA (det vill säga den ideella förening som SKR och staten bildat för att utveckla Kolada) står för. Vissa skolhuvudmän har också regelbundna samlingar för skolledare för att öka förståelsen för vikten av att skolenheternas data rapporteras in på ett korrekt sätt. Men hur vanligt detta är, går inte att fastställa.

Vad gäller den bristande tillgängligheten till data från fristående huvudmän kommer, som tidigare nämnts, detta sannolikt att åtgärdas till 2026.

Den tredje bristen avser områden där det möjligen samlas in för lite statistik. Samtidigt som uppgiftslämnarbristen vad avser grundskola och gymnasieskola är omfattande, så finns det andra skolformer där den sannolikt skulle behöva utökas. Det gäller förskolan samt den anpassade skolan (tidigare särskolan) och specialskolan. Mest påtagligt är hur jämförelsevis lite data som samlas in om förskolans verksamhet. Den information som ligger under uppgiftslämnarskyldigheten är:

- Antal barn
- Ålder
- Andel barn i kommunal regi
- Andel barn i enskild regi
- Andel anställda årsarbetare med pedagogisk högskoleexamen och utbildning för arbete med barn
- Totalt antal anställda
- Antal verksamheter i kommunal resp. enskild regi

Resultatdata är ju inte aktuellt för förskolan, men det är svårt att jämföra förskoleverksamheter hos en och samma huvudman eller hos olika huvudmän enbart på basis av ovanstående uppgifter. För att kunna utveckla analys- och jämförelsemöjligheterna har, som tidigare nämnts, Kolada gått ut med en enkät till förskolepersonal. Men hur svarsstatistiken ser ut för enkäten är oklart.

Det ska framhållas att datainsamlingen från förskolan är personnummerbaserad sedan 2014, vilket innebär att det finns mikrodata att begära ut från SCB. Detsamma gäller för anpassad skola sedan 2019.

Frågan om statistik och informationsinsamling inom anpassad skola och specialskola är betydligt mer känslig. Det rör sig ofta om mycket små målgrupper och informationen är ofta integritetskänslig. Specialpedagogiska skolmyndigheten (SPSM) är inte statistikansvarig myndighet men gör vissa begränsade datainsamlingar. Man har också tillgång till vissa verksamhetsdata från specialskolorna, där SPSM är skolhuvudman. Myndigheten samarbetar även med intressegrupper som finns runt olika funktionsnedsättningar. Sannantaget

finns det inte skäl att anta att mängden data inom SPSM:s verksamhetsfält kommer att öka inom den närmaste tiden.

### 3.2.1 Datamängder på lokal nivå

Utöver de uppgifter som skolhuvudmän samlar in som en följd av sitt uppgiftslämnaransvar gentemot staten, så samlar många skolhuvudmän in egna data. Det kan till exempel gälla diagnostiska prov; frånvaro; enkäter till elever och personal om arbetsklimat, trygghet och hälsa; söktryck till olika gymnasieprogram; kursutvärderingar inom gymnasieskolan med mera. Vidare samlas vanligen en hel del ekonomiska uppgifter på skolenhetsnivå in. Uppgifterna omvandlas till tidsserier och nyckeltal som används av skolledningar, förvaltningsledningar och nämnder för strategiskt beslutsfattande.

Intervjuerna omfattar bland annat forskare med erfarenhet att samarbeta med skolhuvudmän runt deras datatillgång, samt både stora och små skolhuvudmän. Intervjuerna ger vid handen att det föreligger stora skillnader i kompetens och kapacitet mellan olika huvudmän, vanligen beroende på storlek. De mest avancerade gör ambitiösa insatser för att arbeta datadrivet med omfattande egen datainsamling och bearbetning. Andra har varken tid eller kompetens till detta. De mest avancerade huvudmännen har byggt integrationer mellan sina olika verksamhetssystem som gör det möjligt att få ut de rapporter de vill. Andra har medvetet köpt flera system från samma leverantör för att i någon mån få motsvarande nytta. Åter andra har en flora av system där data inte kan flöda mellan systemen vilket ger huvudmannen få möjligheter att exportera data på annat sätt och vid andra tidpunkter än de som systemleverantören bestämt.

Det är ovanligt med samarbete mellan huvudmän ifråga om insamling och analys av data på förskole- och skolnivå. Sannolikt är det något vanligare på gymnasienivå, eftersom flera regioner har gemensam gymnasieantagning. Ett av de få exemplen på grundskolenivån är de tretton västsvenska kommuner som bildar GR (Göteborgsregionen) Utbildning. På sin webbplats skriver GR att man ”samordnar regiongemensamma enkäter till elever och vårdnadshavare i förskolan, grundskolan och gymnasiet för att följa upp och jämföra olika delar av verksamheten, från trygghet till lärande. Undersökningarna ger ett underlag till det systematiska kvalitetsarbetet i kommunerna och regionen.”<sup>34</sup>

### 3.2.2 Inläsning av data i verksamhetssystem

I uppdraget från UVK att göra en behovsinventering avseende utbildningsdata ingår också att studera i vilken utsträckning data från kommunala, regionala, nationella och privata digitala plattformar för skola och utbildning samlas in, samt om dessa datamängder skulle kunna samordnas och tillgängliggöras för att skapa underlag för forskning. Vidare ställs frågan vilka eventuella inläsningar som finns inom huvudmäns avtal för att få tillgång till sina egna datamängder.

---

<sup>34</sup> [Göteborgsregionen: Förskola och grundskola](#)

### ***Många system försvårar interoperabilitet***

Det första man kan konstatera är att det finns en stor mängd verksamhetssystem för skolor att använda. För att underlätta upphandlingsprocesser för skolhuvudmän har branschföreningen Swedish Edtech Industry tagit fram en så kallad edtech-karta.<sup>35</sup> Tanken är att utifrån ett antal verksamhetsprocesser inom tre huvudområden (huvudman, skolledning, undervisning) visa vilka leverantörer som har system att sälja. Under området ”Processer på huvudmannanivå” finns en del kallad ”Analysera verksamhetsdata”. Här listas 25 olika företag, vart och ett med ett eget system som erbjuds skolan. Vissa begränsningar har införts vad gäller vem som får synas på kartan. Till exempel måste företaget vara medlem i branschorganisationen för att få vara med. Trots det är alltså utbudet svåröverskådligt.

Sannolikt kompletterar snarare än konkurrerar flera av systemen med varandra men exemplet belyser ändå den stora splittring som råder. Det skulle inte behöva vara ett svårt problem om leverantörerna byggde sina system utifrån överenskomna standarder.<sup>36</sup> I så fall skulle datamängderna vara strukturerade och uppmärkta på samma sätt. Oavsett vilket system data samlats in och bearbetats i så skulle informationen kunna exporteras ur systemen i samma format. Tyvärr är så sällan fallet av flera skäl. Det första är att det saknas standarder inom många av skolans områden. Först på senare år har detta arbete skjutit fart främst genom att Skolverket 2020 fått sektorsansvar för skolans digitalisering och därmed kan engagera sig i arbetet att utveckla nya standarder.<sup>37</sup> Det andra skälet är att även om det finns utarbetade standarder så är det inte säkert att alla leverantörer implementerar dem i sina system.

Det tredje, och sannolikt viktigaste skälet till att det brister i interoperabiliteten mellan de datamängder olika skolhuvudmän samlar in, är att skolhuvudmännen ofta brister i upphandlingskompetens och inte ställer tillräckliga krav vid upphandlingstillfället eller inte har kraft eller kompetens att i efterhand följa upp att kraven i upphandlingen verkligen efterlevs. En rapport från Konkurrensverket visar skrämmande tydligt att bristande kompetens om standarder och om upphandling leder till en lång rad inläsningar för offentliga aktörer.<sup>38</sup> Rapportens exempel är i huvudsak hämtade från skolsektorn. Rapporten är visserligen från 2016 men det finns skäl att tro att grundproblemet med bristande upphandlingskompetens hos skolhuvudmännen inte genomgått någon större förändring.

Slutsatsen är att datamängder i skolans verksamhetssystem ofta håller en låg grad av interoperabilitet. Det gäller främst mellan olika skolhuvudmän, men det kan också gälla inom en skolhuvudmans olika verksamhetsformer och -system. I en rapport från 2019 visas svårigheter som kan finnas inom en och samma

---

<sup>35</sup> [Swedish Edtech Industry: Edtechkartan](#)

<sup>36</sup> [Svenska institutet för standarder \(SiS\): Vad är en standard?](#)

<sup>37</sup> För mer information om pågående standardiseringsarbete, se [Svenska institutet för standarder \(SiS\): Informationshantering inom utbildningssektorn](#)

<sup>38</sup> [Konkurrensverket: IT-standarder, inläsning och konkurrens. En analys av policy och praktik inom svensk förvaltning](#)



skolhuvudman.<sup>39</sup> Exemplet handlar om elev som byter skola, till exempel vid 12 års ålder när eleven flyttar från en F-5 skola till en 6–9 skola, eller från grundskolan till gymnasiet. Att säkerställa utbyte av information i samband med övergångar är en komplex utmaning med flera delar som till exempel harmonisering av dokumentation och begrepp för att underlätta utbyte och mottagande av information, beslut om vilken information som omfattas och hantering av samtycke med mera. Om eleven dessutom byter skolhuvudman, till exempel från en fristående skola till en kommunal, är processen ännu mer komplex.

SCB:s nya uppdrag, att undersöka möjligheterna till datainhämtning direkt från leverantörernas system, är i högsta grad relevant i detta sammanhang. Frågan är hur stora möjligheter myndigheten har att hämta in data från systemleverantörer. Det handlar om i vilken utsträckning leverantörerna är villiga att släppa ifrån sig denna data och om de finns lagliga möjligheter att tvinga dem att göra det. Vidare finns det, som tidigare nämdes, drygt 20 leverantörer av verksamhetssystem som sinsemellan har olika teknisk uppbyggnad, så de tekniska svårigheterna ska nog inte underskattas. En annan fråga är i vilken utsträckning sådan data kan användas, på grund av brister i trovärdigheten (Nivå B).

### ***Vem äger data?***

Trots en ökad medvetenhet bland skolhuvudmän att de i upphandlingssituationer måste ställa kravet att man själv har fri tillgång till den data som lärare och elever genererar, så tycks det inte ovanligt att den typen av klausul saknas i avtalen. Därmed fortsätter leverantörer att hävda att det är de själva som äger datamängden som finns i deras system. Likaså kan det bereda svårigheter för huvudmannen att snabbt, enkelt och kostnadsfritt kan få ut informationen ur systemet. Det kan kräva många, dyra konsulttimmar att få ut rapporter i det format man vill. Det kan också finnas begränsningar i avtal för hur ofta, hur många gånger per termin eller år, som dataexporter kan göras utan extra kostnader. Det är kommun/skolhuvudman som är personuppgiftsansvarig för alla personuppgifter inom ramen för sin verksamhet och i egenskap av det har ett långtgående ansvar genom GDPR att säkerställa att de rättigheter och friheter som de registrerade har också tas om hand.

Här spelar återigen huvudmannens upphandlingskompetens roll, men givetvis också storleken på huvudmannen. Stora huvudmän har enklare att förhandla fram förmånliga avtal än små.

### ***Samarbete mellan huvudmän, leverantörer och forskare***

Givet de begränsningar som beskrivits ovan, som innebär att en forskare kan behöva göra betydande insatser för att tvätta data och omformatera datamängder för att kunna sammanställa och bearbeta dem, finns ändå stort intresse hos huvudmännen att samarbeta med akademien. De huvudmän som intervjuats inom ramen för denna studie är alla öppna för samarbete med utbildningsforskare. Det

---

<sup>39</sup> [Swedish Edtech Industry: Standarder för datadrivna processer i skolan \(pdf\)](#)

som sätter begränsningar är att kunna frigöra tid för personal att delta. Men man tror sig ha mycket att lära av sådant samarbete. De intervjuade huvudmännen är också öppna för att ställa verksamhetsdata till förfogande, givet att elevers och personals integritet kan garanteras.

Även leverantörer kan ha intresse av samarbete om de ser att det på sikt kan leda till en utveckling av deras produkt. Några av de intervjuade forskarna har frågat om möjlighet att använda data ur leverantörernas system. Det har visat sig fullt möjligt, under förutsättning att de huvudmän som matat in uppgifterna också ger sitt godkännande. En begränsning som nämnts är dock att vissa leverantörer kan vilja ta betalt för att genomföra den dataexport forskaren önskar, och då blir det antingen huvudmannen eller forskaren som får betala.

De datamängder som vanligen diskuteras är data om elevers resultat, frånvaro och liknande. Emellanåt också enkätresultat från elev- och medarbetarundersökningar. Mer sällan uppmärksammas den data som samlas in när elever arbetar i digitala läromedel. Där tycks det ovanligt att vare sig huvudmän eller forskare får tillgång till annat än översiktliga uppgifter. I övrigt betraktar leverantörerna vanligen denna data som affärshemligheter. Även inom detta område pågår dock arbete inom standardiseringsorganet SIS. En svensk applikationsprofil för metadata för digitala lärresurser från 2014 är under revidering.<sup>40</sup>

### **3.2.3 Tillgänglighet, trovärdighet och FAIR-principerna för datamängder på lokal nivå**

Som framgår uppfyller få lokala datamängder dataförberedelsenivå C. Även om en forskare har kännedom om att en datamängd existerar, så innebär det inte att den är tillgänglig för bearbetning utan överenskommelse med huvudmannen. Det finns en rad etiska aspekter samt legalitets- och integritetsaspekter att ta hänsyn till. Även om det är svårt att uttala sig generellt om förhållandena hos Sveriges över 1 100 skolhuvudmän så är det sannolikt att det ofta föreligger tekniska svårigheter för att få tillgång till data. Vad gäller nivå B, trovärdighet, kan resonemanget avseende brister i den officiella skolstatistiken återopas.

FAIR-principerna är inte heller uppfyllda i de lokala datamängderna.

---

<sup>40</sup> [Svenska Institutet för Standarder \(SiS\): Informationshantering inom utbildningssektorn](#)

## 4 Beskrivning av datamängder inom folkhögskolan

Det finns omkring 150 folkhögskolor i landet. De flesta är knutna till och drivna av olika folkrörelser men ett antal drivs av regioner. SCB är statistikansvarig myndighet för folkhögskolornas verksamhet och svarar för insamling och sammanställning i samråd med Folkbildningsrådet. I SCB:s statistikdatabas finns historisk folkhögskolestatistik under område Utbildning och forskning – Folkhögskola.<sup>41</sup> Där presenteras årlig statistik över folkhögskolans deltagare. De uppgifter som redovisas är:

- kön
- åldersgrupper
- nationell bakgrund
- kurstyp
- kursinriktning
- studieomfattning
- typ av erhållen behörighet
- studienivå
- deltagarens tidigare utbildningsnivå
- deltagarens föräldrars utbildningsnivå
- region

Här redovisas även årlig statistik över folkhögskolans pedagogiska personal och kostnader. Vad gäller den pedagogiska personalen så redovisas personalens kön, åldersgrupp, kommungrupp samt huvudsakligt undervisningsämne allmän kurs respektive särskild kurs. Dessa uppgifter fördelas vidare på personaltyper som skolläring och administrativt arbete, studie- och yrkesvägledning samt olika typer av pedagogiskt arbete. Kostnaderna beskrivs utifrån undervisning, administration och marknadsföring, kök och hushåll samt fastigheter och lokaler. För varje område görs en uppdelning mellan personal och övrigt. Se vidare Bilaga 1 och 2.

På Folkbildningsrådets webbplats finns även statistiken indelad regionvis och per skola.<sup>42</sup> Den baseras på aggregerade uppgifter och icke unika deltagare, medan den som SCB publicerar avser unika deltagare.

Folkhögskoleverksamheten presenteras årsvis grupperad på län, folkhögskola och kurstyp med antalsuppgifter för deltagarveckor, deltagare, kön, ålder och utbildningsbakgrund bedömd av folkhögskolan utifrån studiernas nivå och syfte. Den till SCB totalt inrapporterade verksamheten sammanfattas och datafilen ger underlag till en övergripande verksamhetsanalys på regions- och/eller skolnivå. På hemsidan finns uppgifter för 2020 och 2021 tillgängliga. Det är möjligt att få

---

<sup>41</sup> [SCB: Folkhögskolestatistik](#)

<sup>42</sup> [Folkbildningsrådet: Regionsindelad folkhögskolestatistik](#)

statistik längre tillbaka i tiden. Det är något som Folkbildningsrådet producerar och levererar efter beställning.

Det finns även statistik om individer med funktionsnedsättning på folkhögskolor. Den har samlats in sedan 2018 efter att SCB fick ansvar för den officiella statistiken för folkhögskolan, samt ett utökat ansvar att behandla känsliga personuppgifter för individer på folkhögskolan. Det som anges är vilken typ av funktionsnedsättning deltagaren har som påverkar studierna mest. Variabeln publiceras inte av etiska skäl.

Folkhögskoledatabasen är ett resultat av ett pågående forskningsprojekt. Där har man samlat folkhögskolornas nationella kurskataloger mellan åren 1952–2019.<sup>43</sup> Katalogerna innehåller beskrivningar av folkhögskolan som utbildningsform samt specifikationer över det kursutbud skolorna tillhandahållit. Det går att ladda ned alla digitala kataloger som OCR-inlästa pdf-filer. Databasen är öppen och fritt tillgänglig.

## 4.1 Statistikens tillgänglighet och trovärdighet

Som framgår publiceras statistiken av SCB i vissa fall kompletterade med sammanställningar som finns på Folkbildningsrådets webbplats. Det innebär att statistiken har hög tillgänglighet. Vad gäller trovärdigheten får den i huvudsak också anses hög med något undantag. Liksom inom skolområdet är sannolikt kostnadsstatistiken mindre trovärdig. Folkhögskolorna har inget gemensamt ekonomisystem eller gemensam kontoplan. Det medför att kostnadernas fördelning sannolikt inte redovisas likadant från alla omkring 150 folkhögskolor.

### 4.1.1 FAIR-principerna inom folkhögskolesektorn

Den statistik om folkhögskolorna som finns på nationell nivå tycks i hög grad följa FAIR-principerna. Data går att hitta (Findable), den är till största delen tillgänglig (Accessible), både data och metadata följer vedertagna standarder och data publiceras i filformat som stöds av olika operativsystem och kan öppnas i flera program. Slutligen beskrivs data med tillräckliga metadata som kan läsas av både människor och datorer, och att det finns tydliga upplysningar om i vilket sammanhang data samlades in och villkor för hur data får användas måste anges tydligt. Därmed uppfylls även kravet på återanvändbarhet (Reusable).

---

<sup>43</sup> [LiU Electronic Press: Folkhögskoledatabasen](#)

## 5 Beskrivning av datamängder inom yrkeshögskolan

Myndigheten för yrkeshögskolan (MYH) är förvaltningsmyndighet för yrkeshögskolan. Myndigheten har ansvar för eftergymnasial yrkesutbildning, som inte faller under högskolelagen. Den absolut största utbildningsformen är yrkeshögskoleutbildningarna. Där började 2022 omkring 19 000 personer en utbildning. Konst- och kulturutbildningar har omkring 3 000 deltagare per år och tolkutbildningarna omkring 300.<sup>44</sup>

### 5.1 Data om deltagare

Sedan 2018 är SCB ansvariga för officiell statistik om sökande, studerande och examinerade inom yrkeshögskolan samt om studerande och studerande som slutfört utbildning inom konst- och kulturutbildningar och utbildningar med endast tillsyn. Statistiken finns publicerad på SCB:s webbplats.<sup>45</sup> Merparten av övrig statistik i rapporten finns som tabeller på MYH:s statistiksida.<sup>46</sup>

För yrkeshögskolan redovisas statistiken i de flesta fall från 2007. Statistiken för konst- och kulturutbildningar samt utbildningar med endast tillsyn redovisas från det att utbildningsformen startade hösten 2015. För tolkutbildningarna inom folkbildningen redovisas statistiken huvudsakligen från 2012 då myndigheten tog över ansvaret för utbildningarna.

Det finns individdata (inkl. personnummer) på alla som studerat på yrkeshögskolan (YH) eller som det tidigare kallades kvalificerad yrkesutbildning (KY).

De uppgifter som samlas in handlar om utbildningsdeltagande och examen, individdata för studerandes prestationer, kursdeltagande och kursbetyg samt även individens start-, slut- och examensdatum. Sammantaget finns 41 variabler för studerande och examinerade. MYH kan efter prövning lämna ut mikrodata för forskningsändamål.

SCB erbjuder dokumentation om mikrodataregister för YH.<sup>47</sup> Variablerna i mikrodataregisterna kan vid beställningar kompletteras med till exempel den utbildningsinformation som SCB använt för publicering av statistik om YH och andra registervariabler som finns hos SCB. Mikrodata hos SCB är skyddad av

---

<sup>44</sup> [Myndigheten för yrkeshögskolan \(MYh\): Statistisk årsrapport 2023 \(pdf\)](#)

<sup>45</sup> [SCB: Yrkeshögskolan](#)

<sup>46</sup> [Myndigheten för yrkeshögskolan \(MYh\): Statistik](#)

<sup>47</sup> [SCB: Dokumentation av mikrodata](#)

sekretess men kan efter sekretessprövning lämnas ut i avidentifierad form för forskningsändamål.<sup>48</sup>

## 5.2 Data om utbildningar

MYH ansvarar för fyra typer av utbildningar: yrkeshögskolan, konst- och kulturutbildningar, tolkutbildning inom folkhögskolan samt utbildningar med endast tillsyn.

### Yrkeshögskoleutbildningar

Yrkeshögskoleutbildningar finns inom flera olika branscher över hela landet. Utbildningarna ska svara mot arbetslivets behov av yrkeskompetens och drivs i nära samverkan med företag och arbetsgivare.

Utbildningarna är högskoleförberedande, yrkesinriktade eller inom kulturarvsområdet och drivs av privata utbildningsanordnare. MYH publicerar statistik över utbildningsanordnare efter huvudmannatyp, utbildningsområde, län och arbetsmarknadsregion.<sup>49</sup> Här finns även uppgifter om examensgrad, sysselsättningsgrad för de studerande efter avslutad utbildning, nöjdhet med utbildningarna med mera.

### Konst- och kulturutbildningar

Inom konst- och kulturutbildningar finns det flera olika inriktningar som bildkonst, musik, teater, dans, mode, skrivande, arkitektur, silversmide och keramik för att nämna några. Utbildningarna är indelade i tre kategorier med olika syften och ska:

- förbereda för högskoleutbildningar som kan leda fram till konstnärliga examina,
- medverka till att utveckla ett kvalificerat yrkeskunnande inom det konstnärliga eller kulturella området, eller
- ha ett innehåll som syftar till att bevara eller utveckla kulturarvet.

Utbildningarna drivs av privata utbildningsanordnare och kan få statligt stöd genom statsbidrag och/eller genom att de studerande får rätt till studiemedel från CSN. Utbildningsanordnarna tar ut studerandeavgifter som varierar mellan utbildningarna.

MYH publicerar statistik över ansökningar om stöd, utbildningsanordnare, vilka utbildningar som givits och utbetalt statsstöd.<sup>50</sup> Även här finns uppgifter om antal studerande och hur stor andel som slutfört utbildningarna.

---

<sup>48</sup> Mer information om att beställa mikrodata från SCB och vilka mikrodata som SCB har finns här: [SCB: Mikrodata](#)

<sup>49</sup> [Myndigheten för yrkeshögskolan \(MYh\): Statistik: Utbildningsanordnare](#)

<sup>50</sup> [Myndigheten för yrkeshögskolan \(MYh\): Statistik: Utbildningar och platser](#)

## Tolkutbildningar

Tolkutbildning inom folkbildningen innefattar utbildningar till kontakttolk, teckenspråks- och dövblindtolk samt skrivtolk som drivs av folkhögskolor och studieförbund. Från och med 2019 finns kontakttolkutbildningar även inom yrkeshögskolan.

Motsvarande uppgifter om utbildningsanordnare och antal studerande med mera finns för de olika tolkutbildningarna.

## Statligt stöd i form av tillsyn

Fristående eftergymnasiala utbildningar inom olika områden kan ansöka om statligt stöd i form av tillsyn av MYH. Den typ av data om utbildningarna som samlas in är till exempel inriktning och studieform, utbildningsanordnare och utbetalda statliga medel. Från ansökningarna att få bedriva utbildningar finns också en del information som beskrivningar av arbetsmarknadens efterfrågan av yrkesrollen, information om arbetslivets medverkan i utbildningen, organisation och ledning, kostnadsbudget. Vilken information som ansökan ska innehålla kan dock variera något mellan åren så det är inte säkert att samma uppgift funnits över åren.

Vidare finns en del data i utbildningsplanerna som kursbeskrivningar, uppgifter om tillträdet till utbildningen (särskilda förkunskapskrav och urvalsgrunder), utbildningens utformning och organisation samt kvalitetssäkring av utbildningen.

MYH gör en del egna undersökningar och främst beskrivande rapporter.<sup>51</sup> Det kan handla om sådant som söktryck, examensgrad, sysselsättning efter genomförd utbildning, etableringsgrad, matchning eller utbildningsform (bunden respektive distans). Man har även gjort en del djupare studier, till exempel om flyttmönster samt avklarade poäng/kurser för ej examinerade.

## 5.3 Statistikens tillgänglighet och trovärdighet

Tillgängligheten av statistiken måste bedömas som god. Den har samma tillgänglighet som inom övriga utbildningssektorer. Vad gäller trovärdigheten uppger, som tidigare nämnts, myndigheten själv i en intervju att kvaliteten i statistiken är ”tillräckligt bra” för åren 2005–2006. Vidare menar man att det är ”bra kvalitet” 2007–2016 och ”mycket bra kvalitet” från och med 2017.<sup>52</sup>

Samtidigt bör det också påpekas att yrkeshögskolan står inför samma problem som skolan vad gäller inrapportering av statistik. Det är många, och ofta små, utbildningsanordnare som ska rapportera in sina uppgifter till myndigheten. Nya utbildningsanordnare tillkommer ofta och gamla försvinner – allt eftersom kursutbudet anpassas efter efterfrågan. Vidare handlar det om många olika utbildningar och på senare år också ett ökande antal korta kurser. Det är därför

<sup>51</sup> [Myndigheten för yrkeshögskolan \(MYH\): Publikationer](#)

<sup>52</sup> Intervju med myndigheten 2023-04-13

rimligt att anta att många utbildningsanordnare är ovana att rapportera och kanske saknar personal med adekvat kompetens. Men några intervjuer med utbildningsanordnare har inte gjorts så en självständig bedömning av trovärdigheten kan inte göras inom ramen för denna kartläggning. Därför får myndighetens beskrivning av trovärdigheten gälla.

### **5.3.1 FAIR-principerna inom yrkeshögskolesektorn**

Statistiken om yrkeshögskolan som finns på nationell nivå tycks i hög grad följa FAIR-principerna. Data går att hitta (Findable), den är till största delen tillgänglig (Accessable), både data och metadata följer vedertagna standarder och data publiceras i filformat som stöds av olika operativsystem och kan öppnas i flera program. Slutligen så beskrivs data med tillräckliga metadata som kan läsas av både människor och datorer, och att det finns tydliga upplysningar om i vilket sammanhang data samlades in och villkor för hur data får användas måste anges tydligt. Därmed uppfylls även kravet på återanvändbarhet (Reusable).



## 6 Beskrivning av datamängder inom högskolan

### 6.1 Universitetskanslersämbetet

Universitetskanslersämbetet (UKÄ) har i uppdrag av regeringen att ansvara för den officiella statistiken inom högskoleområdet. SCB samlar in uppgifter från universitet och högskolor på uppdrag av UKÄ.

Liksom inom övriga utbildningssektorn finns statistik dels om de studerande och dels om lärosätena. Vad gäller individstatistiken så publicerar UKÄ statistik över sökande, antagna och nybörjare i högskolan. Myndigheten beskriver själv på sin webbplats att detta är komplicerat.<sup>53</sup> Mycket händer mellan söktillfälle och terminsstarten. Alla som söker blir inte antagna och alla som antas påbörjar inte utbildningen. Det sker också efteranmälningar till utbildningar även långt efter terminsstart. Statistik som sammanställs vid en tidpunkt kommer därmed sannolikt skilja sig från statistik som sammanställs vid en annan tidpunkt på grund av den utgår från olika populationer. Det finns flera rapporter om hur arbetet går till och vilka avvägningar myndigheten gjort.<sup>54</sup>

Det finns en webbsida kallad Högskolan i siffror där UKÄ:s statistik publiceras.<sup>55</sup> Den är indelad i fem delar: Utbildning på grundnivå och avancerad nivå, Utbildning på forskarnivå, Ekonomi, Personal samt Forskning vid universitet och högskolor.

Vad gäller utbildning på grundnivå och avancerad nivå är syftet med uppgifterna att beskriva inflöde, omfattning och utflöde ur högskolans utbildningsverksamhet. Statistiken redovisar antalet individer eller helårsstudenter per läsår eller termin. Exempel på vanligt förekommande dimensioner är ämne eller program, lärosäte och kön. På webbsidan publiceras statistik om sökande och antagna, nybörjare, registrerade studenter, internationell studentmobilitet, examina och examinerade, helårsstudenter (HST) och helårsprestationer (HPR), poänggivande uppdragsutbildning samt etablering på arbetsmarknaden 1–1,5 år efter examen.

För forskarutbildningen är det nybörjare, registrerade och examina som beskrivs. I statistiken redovisas antalet doktorander, som individer eller helårspersoner, per läsår eller termin. Exempel på vanligt förekommande dimensioner är ämne, lärosäte och kön.

---

<sup>53</sup> [Universitetskanslersämbetet \(UKÄ\): Statistik](#)

<sup>54</sup> Se t.ex. [Universitetskanslersämbetet \(UKÄ\): Kvalitetsstudie av sökande- och antagningsstatistiken \(pdf\)](#) och [Beskrivning av statistik om sökande och antagna - UKÄs statistik om sökande och antagna på grundnivå och avancerad nivå \(pdf\)](#)

<sup>55</sup> [Universitetskanslersämbetet \(UKÄ\): Högskolan i siffror](#)

För forskningsverksamheten vid universitet och högskolor finns uppgifter om forskningspublikationer, internationell rekrytering av forskare och forskares mobilitet. I statistiken redovisas uppgifter både på riksnivå och per lärosäte. Några förekommande dimensioner är forskningsämnesområde, anställningskategori och kön.

Lärosätens ekonomi och finansiering beskriver samlade uppgifter från den ekonomiska redovisningen, så som resultat- och balansräkningar. Där finns också detaljerade uppgifter om lärosätens intäkter. Statistiken består framförallt av intäkter och kostnader per kalenderår. Några förekommande dimensioner är lärosäte, intäktsslag, finansiär och typ av kostnad.

Under personal redovisas bland annat hur många som finns anställda av universitet och högskolor, vilka anställningskategorier de tillhör och hur utvecklingen har sett ut över tid. I statistiken redovisas uppgifter om antal individer och antal omräknat till heltidsekvivalenter, både på riksnivå och per lärosäte. Några förekommande dimensioner är forskningsämnesområde, åldersgrupp och kön.

Webbsidan redovisar statistiken i diagramform men ger också möjlighet att visa den i tabellform. Uppgifterna finns också nedladdningsbara som csv-filer och Excel-filer.

## 6.2 Universitets- och högskolerådet

Universitets- och högskolerådet (UHR) ansvarar för uppgifter i fråga om service, samordning, främjande och utvecklande verksamhet samt internationellt samarbete och mobilitet inom utbildningsområdet. Man har inget uttalat ansvar att samla in statistik men ansvarar ändå för en del databaser.

### **Beda – betygsdatabas**

UHR ansvarar för den nationella betygsdatabasen Beda som samlar in examina och slutbetyg från gymnasieskolor och kommunal vuxenutbildning och vidarebefordrar dem till SCB och antagningssystemet NyA. Under de kommande åren kommer Beda att förberedas för att också användas i antagningen till kommunal vuxenutbildning, folkhögskolan och yrkeshögskolan.

Beda samlar in examen och slutbetyg från gymnasial utbildning och levererar dem till antagningssystemet NyA som används för antagningen till svenska universitet och högskolor. Beda levererar också betygsdata till SCB och till kommunerna för uppföljningen inom det kommunala aktivitetsansvaret. Variablerna i Beda är:

- Elevens personnummer
- Skola
- Studieväg
- Programmets omfattning
- Elevens studieplan

- Betyg och poäng per kurs
- Poäng för lokalt tillägg för elever med slutbetyg

UHR sköter både verksamhets- och teknikförvaltningen av Beda. Men UHR har inte rätt att använda den data som finns i systemet. Det har endast SCB. Beda har använts i mer än 20 år och registret används av SCB för att för Skolverkets räkning bearbeta uppgifter från den nationella betygsdatabasen.

För att beställa mikrodata från Beda krävs ansökan till SCB som skriver att efter särskild prövning är det möjligt för till exempel forskare att få tillgång till avidentifierade mikrodata för egna bearbetningar.<sup>56</sup>

### **Antagningssystemet NyA**

Statistiken över sökande och antagna har samlats in från och med höstterminen 1995. Uppgifter på individnivå sparas i tio år. De variabler som finns i NyA är:

- Universitet/högskola
- Kön
- Ålder
- Ej tidigare i högskolan
- Program
- Kurs
- År
- Termin
- Sökalternativ
- Prioritet
- Urvalsresultat
- Yrkesexamenskod
- Klassificering enligt Svensk utbildningsnomenklatur, SUN 2000
- Anmälningavgift och studieavgift
- Län
- Kommun

SCB skriver att de på uppdragsbasis kan göra specialstudier, genom till exempel sambearbetning med annan statistik eller skapande av tidsserier. Efter prövning finns möjlighet att för forskning få tillgång till mikrodata i avidentifierad form för egna bearbetningar. Registret kan också utgöra urvalsram för enkätundersökningar.<sup>57</sup>

### **Valda – ansöknings- och administrationssystem för validering**

Valda är ett verktyg för att underlätta validering av yrkeskompetens i antagningen till yrkeslärarutbildningen och Lärarlyftet II. Det pågår också ett projekt för att utveckla Valda till ett stöd för valideringen av reell kompetens.

---

<sup>56</sup> [SCB: Data från Universitets- och högskoleregistret – studenter och examinerade i högskoleutbildning på grundnivå och avancerad nivå](#)

<sup>57</sup> [SCB: Data från Universitets- och högskoleregistret – studenter och examinerade i högskoleutbildning på grundnivå och avancerad nivå](#)

Valda innehåller också en kunskapsbank med avslutade ärenden. Alla avslutade ärenden sparas i ett nationellt arkiv som fungerar som en kunskapsbank för alla högskolor som är anslutna till Valda. Våren 2023 innehöll Valda 320 ärenden från fyra lärosäten: Malmö universitet, Stockholms universitet, Karlstad universitet och Högskolan Väst. Enligt uppgift från UHR är data ur Valda ännu inte tillgängliga för forskare.

### 6.3 Högskolestatistikens tillgänglighet och trovärdighet

Den officiella statistiken om högskolesektorn måste sägas ha hög tillgänglighet även om det i krävs ansökningar till SCB för att få tillgång till mikrodata. Även trovärdigheten måste anses hög. Jämfört med skolektorn så är det betydligt färre och större enheter som rapporterar in data och datainsamling och rapportering görs av kompetent personal.

#### 6.3.1 FAIR

Den högskolestatistik som finns på nationell nivå tycks i hög grad följa FAIR-principerna. Data går att hitta (Findable), den är till största delen tillgänglig (Accessible), både data och metadata följer vedertagna standarder och data publiceras i filformat som stöds av olika operativsystem och kan öppnas i flera program. Slutligen så beskrivs data med tillräckliga metadata som kan läsas av både människor och datorer, och att det finns tydliga upplysningar om i vilket sammanhang data samlades in och villkor för hur data får användas måste anges tydligt. Därmed uppfylls även kravet på återanvändbarhet (Reusable).

### 6.4 Brister i högskolestatistiken

De brister inom högskolestatistiken som framkommit i undersökningen gäller främst brist på tillgänglighet och att viss statistik inte samlas in.

En synpunkt som framförts i forskarintervjuerna är att det varit svårt att få tillgång till data om högskolan. En forskare fick inte tillgång till betygdata ur UHR:s BEDA-register (som innehåller uppgifter om studenters avgångsbetyg från gymnasieskolan). Vidare hade han svårigheter att få ut data från sitt eget lärosäte – hans kommentar var att högskolan tycks ovan och obenägen att låta sig själv bli föremål för forskning. I en annan intervju bekräftas ovanstående bild. Denna forskare menar även att lärosätena ofta saknar datainfrastruktur för att effektivt kunna behandla och analysera eget data. Man kan genomföra fallstudier, men inte så mycket mer.

Flera forskare framhåller att man inte fått ut lärosätets egna data ur LADOK. LADOK-systemet ägs gemensamt av 40 universitet och högskolor samt CSN.<sup>58</sup> På LADOK:s hemsida finns heller ingen information om att forskare kan ansöka om att få ut data. Det tycks med andra ord vara en parallell till de verksamhetssystem skolan har där Ifous-projektet Datadriven skolförbättring

---

<sup>58</sup> [Ladok.se](http://Ladok.se)

påpekade att systemen är designade enbart för rapportering uppåt, inte för egen bearbetning och verksamhetsutveckling.<sup>59</sup>

En annan synpunkt är att samma begränsning som råder på skolsidan även råder inom högskolan, nämligen att man inte får koppla ihop elevgrupp med lärare vilket begränsar möjligheterna till analyser bland annat av effektiva undervisningsmetoder. Därmed försvåras ett systematiskt utforskande av effektiva undervisningsmetoder och möjlighet till utvecklingsarbete.

Frånvaron av gemensamma tentamina eller andra examineringsmetoder (motsvarande skolans nationella prov) gör det dessutom omöjligt att jämföra kvalitet och effektivitet mellan olika lärosäten, framför allt inom programutbildningarna. Det går inte att avgöra om sjuksköterskorna utexaminerade på Umeå universitet är skickligare än de som examinerats på Göteborgs universitet, eller om en jurist från Lund är skickligare än en från Stockholm. Inte heller om det ena eller andra lärosätet hittat metoder att vara mer effektiva men till samma kostnad som andra lärosäten.

Men betydligt färre forskare efterfrågar denna typ av data för högskolan jämfört med på skolområdet.

---

<sup>59</sup> [Ifous: Datadriven skolförbättring \(pdf\) \(sid 63f\)](#)

## 7 Övriga relevanta datamängder

Nedan beskrivs kortfattat ett antal databaser eller datamängder som kan vara av intresse för utbildningsvetenskaplig forskning. Men eftersom vad som är relevant avgörs från fall till fall så kan uppräkningsen inte ses som uttömmande.

### 7.1 Utbildningsregistret – UREG

Utbildningsregistret är ett individregister hos SCB som innehåller information om högsta utbildningsnivå, utbildningsinriktning, år för avslutad utbildning, kommun för avslutad utbildning samt demografiska variabler.<sup>60</sup>

- Personens högsta utbildningsnivå
- Utbildningsinriktning
- Utbildningsår (varierande täckningsgrad)
- Utbildningsort (varierande täckningsgrad)
- Kön
- Ålder
- Folkbokföringslän och -kommun
- Födelseland

### 7.2 Utvärdering Genom Uppföljning – UGU

UGU är en kohortsekventiell nationell infrastruktur som bland annat används för utvärdering och forskning inom skolan.<sup>61</sup> Den baseras på elva kohorter i årskurs 3 som har blivit utvalda av SCB. Dessa individer följs genom deras skolgång och urvalet är stratifierade och riksrepresentativa. Hittills har UGU data för elva kohorter mellan 1948 och 2010. Data består av enkätfrågor, kognitiva tester och registerdata. Centralt för registret är att enkätdata består av elevernas individuella svar.

Data är tillgängliga mot förfrågan för forskare, doktorander och studenter vid svenska och utländska universitet och högskolor.

För UGU gäller att det finns utförliga beskrivningar på hemsidan av hur insamlingen skett, hur man kan få information om variablerna, vem som kan använda sig av datamängden osv.

---

<sup>60</sup> [SCB: Mikrodata för Registret över befolkningens utbildning](#)

<sup>61</sup> [Göteborgs universitet: Utvärdering genom uppföljning \(UGU\)](#)

### 7.3 Gothenburg Educational Longitudinal Database – GOLD

GOLD-data innehåller avidentifierat individdata på tre miljoner svenskar och deras uppgifter från bland annat Centrala Studienämnden (CSN), Pliktverket och Riksarkivet.<sup>62</sup>

Datamängden är tillgänglig för disputerade forskare, forskare/affilierad forskare på Göteborgs universitet, personer som är verksamma i ett land/myndighet som hanterar personuppgifter i enlighet med dataskyddsförordningen GDPR. Även GOLD:s data är utförligt beskriven på hemsidan.

Den 1 april år 2001 genomfördes en lagändring som gör det möjligt för SCB att bevara identifikationsnyckeln för individer i databaser som annars är avidentifierade. Detta gör det möjligt för forskare att arbeta med omfattande individdatabaser utan att detta innebär hot mot den individuella integriteten, samtidigt som det är möjligt att påföra ytterligare uppgifter från uppföljningsundersökningar. I enlighet med detta regelverk har en databas konstruerats av SCB, vilken är möjlig att uppdatera.

### 7.4 COMPEAT

Comparative Analysis of Educational Achievement (COMPEAT)<sup>63</sup> är ett center vid Göteborgs universitet som har i uppgift skapa förutsättningar för analyser av data från de komparativa studierna genom att bygga upp och förbereda databaser för analys, samt genom att medverka till uppbyggnad av analyskompetens genom kurser, workshops och seminarier. Under de senaste 40 åren har ett stort antal internationella jämförelser genomförts av elevers kunskaper och färdigheter inom olika områden av organisationer som IEA och OECD.

### 7.5 ILSA Gateway till internationella studier

ILSA Gateway<sup>64</sup> är ingången till information om internationella storskaliga utvärderingar (ILSA) inom utbildning. Här finns omfattande information om varje studie, inklusive direkllänkar till data och dokument på de olika externa studiewebbplatserna, en databas med studierelaterade forskningsrapporter och meddelanden om senaste nytt och kommande evenemang. Gateway har utformats för en global publik, först och främst forskare, men den bör också vara ett användbart verktyg för politiska beslutsfattare, beslutsfattare och alla andra som är intresserade av utbildningsforskning.

---

<sup>62</sup> [Göteborgs universitet: GOLD](#)

<sup>63</sup> [Göteborgs universitet: Center for Comparative Analysis of Educational Achievement, COMPEAT](#)

<sup>64</sup> [International Association for the Evaluation of Educational Achievement \(IEA\): ILSA-Gateway](#)

## 7.6 Skolvalet

Myndigheten för ungdoms- och civilsamhällesfrågor (MUCF) har en databas över resultat i skolvalet 2010, 2014, 2018/19 och 2022.<sup>65</sup> På webbsidan finns tillgång dels till rapporter från de olika valen men också möjlighet att i Excel-format bryta ner resultaten på läns-, kommuns- och ortsnivå.

## 7.7 Ungidag

MUCF driver ungidag.se. Där samlas statistik om ungas levnadsvillkor från elva myndigheter.<sup>66</sup> Statistiken på ungidag presenteras i sex ämnen:

- Arbete & boende
- Ekonomisk och social utsatthet
- Fysisk och psykisk hälsa
- Inflytande och representation
- Kultur och fritid
- Utbildning

Varje år görs en fördjupad analys inom ett eller flera ämnen. Dessa redovisas i en rapportserie kallad just ”ung idag”.

Statistiken som redovisas på ungidag.se kommer från följande myndigheter:

- Arbetsförmedlingen
- Arbetsmiljöverket
- Brottsförebyggande rådet
- Centrala studiestödsnämnden
- Folkhälsomyndigheten
- Försäkringskassan
- Kronofogdemyndigheten
- Myndigheten för ungdoms- och civilsamhällesfrågor
- Statens skolverk
- Socialstyrelsen
- Statistiska centralbyrån

## 7.8 Barnen och internet

Barnen och internet är en del av den större undersökningen Svenskarna och internet som genomförs av Internetstiftelsen.<sup>67</sup> Svenskarna och internet är en årlig undersökning av svenska folkets internetvanor, hur användningen av internet utvecklas och digitaliseringen av samhället. Barnen och internet innehåller frågor om internetanvändning av olika slag både i skolan och på fritiden.

---

<sup>65</sup> [Myndigheten för ungdoms- och civilsamhällesfrågor \(MUCF\): Skolval vid allmänna val](#)

<sup>66</sup> [Myndigheten för ungdoms- och civilsamhällesfrågor \(MUCF\): Ung idag](#)

<sup>67</sup> [Internetstiftelsen: Svenskarna och internet](#)



## 7.9 Högskoleprovet

Umeå universitet förvaltar registret för högskoleprovsresultat på uppdrag av UHR.<sup>68</sup> Registret innehåller resultat för alla som skriver Högskoleprovet på individ- och uppgiftsnivå. Begäran om att få ut uppgifter ur Högskoleprovet sker hos UHR. Det anges tydligt på webbsidan att uppgifterna enbart lämnas ut i pappersformat. Inga undantag görs.

## 7.10 Tvillingregistret

Karolinska institutet förvaltar Tvillingregistret med information om cirka 87 000 tvillingpar med känd zygositet (en- eller tvåäggstvillingar).<sup>69</sup> Registret är världens största i sitt slag. Svenska Tvillingregistret upprättades på 1960-talet och idag pågår ca 30 projekt som baseras på registret. Tillgång till data ges efter ansökan. På webbplatsen finns också en aktuell prislista.

## 7.11 Folkhälsomyndigheten

Hos Folkhälsomyndigheten samlas data över barns och elevers hälsovanor.<sup>70</sup> Det gäller till exempel kroppsuppfattning, levnadsvanor, sexualitet, hälsa och sociala relationer. Det finns även en rubrik kallad Skolan, med underrubriker som gymnasievalet, krav i skolan, mobbing, trivsel och relationer till lärare och klasskamrater.

Folkhälsomyndigheten undersöker även hur skolbarns hälsa och levnadsvanor förändrats över tid. Finns det skillnader i hälsa och vanor mellan olika grupper av skolbarn? Det görs genom undersökningen Skolbarns hälsovanor.<sup>71</sup>

Undersökningen är internationell och genomförs vart fjärde år i ett femtiotal länder i samarbete med Världshälsoorganisationen (WHO).

## 7.12 Ungdomsbarometern

Ungdomsbarometern<sup>72</sup> publiceras av ett privat undersökningsföretag Cint AB. De genomför enkätundersökningar med paneler av ungdomar i olika frågor som trendrapporter, generationsrapporter och mycket annat. Det går att ta del av rapporterna men är sannolikt svårt att få ut mikrodata.

---

<sup>68</sup> [Umeå universitet: Utdrag ur Högskoleprovsregistret](#)

<sup>69</sup> [Karolinska Institutet \(KI\): Svenska Tvillingregistret](#)

<sup>70</sup> [Folkhälsomyndigheten: Folkhälsodata - Skolbarns hälsovanor, nationella och regionala resultat](#)

<sup>71</sup> [Folkhälsomyndigheten: Skolbarns hälsovanor](#)

<sup>72</sup> [Ungdomsbarometern: Rapport & Undersökningar](#)

### 7.13 Undersökning av levnadsförhållanden – ULF

I SCB:s statistikdatabas som byggts upp utifrån ULF finns även en del uppgifter om fritidslärande, det vill säga informellt lärande.<sup>73</sup>

### 7.14 IFAU

IFAU har tre forskningsdatabaser varav en, IFAU-U, innehåller uppgifter om vilka som deltar i utbildning, både barn och vuxna. Det finns vissa uppgifter om hur det gått (betyg) och hur studierna finansierats. Det finns också äldre uppgifter från inskrivningsprov (mönstring). Eleverna kan följas upp i vuxen ålder. IFAU-U innehåller därför uppgifter om samtliga svenskar oberoende av ålder.

Av sekretessöverenskommelsen mellan Statistiska Centralbyrån (SCB) och IFAU framgår att materialen endast får utnyttjas av forskare knutna till IFAU. Med detta avses anställd personal, forskare som får forskningsbidrag från IFAU, forskare eller andra som i ett IFAU-projekt samarbetar med anställd personal, samt affilierade forskare. En förteckning över affilierade forskare finns på IFAU:s webbplats.<sup>74</sup>

### 7.15 Citeringsdatabaser

En citeringsdatabas är ett register över vilka artiklar (exempelvis examensrapporter, konferenspublikationer, patent och journalartiklar) och författare som hänvisar till en viss artikel eller författare. Detta kan användas för citeringsanalys, vilket är en viktig metod inom bibliometri, exempelvis för att ranka hur betydelsefull en viss vetenskaplig artikel, forskare eller tidskrift är. De mest kända är förmodligen Web of Science<sup>75</sup>, Scopus<sup>76</sup> och Google Scholar Citations.

Som en hjälp att jämföra olika citeringsdatabaser har Finland byggt upp en webbplats kallad Tritonia.<sup>77</sup> Där beskrivs till exempel de olika citeringsdatabasernas inriktning och omfattning. Vissa är mer inriktade mot hårda vetenskaper, andra mot mer mjuka. Det finns också information om databaserna är fritt tillgängliga eller om det krävs avgifter för att nyttja tjänsterna.

---

<sup>73</sup> [SCB: Statistikdatabasen - Levnadsförhållanden](#)

<sup>74</sup> Se IFAU:s policy för forskningsdata: [Institutet för arbetsmarknads- och utbildningspolitisk utvärdering \(IFAU\): Forskningsdata](#)

<sup>75</sup> [Clarivate: Web of Science](#)

<sup>76</sup> [Elsevier: Scopus](#)

<sup>77</sup> [Tritonia: Utvärdering av vetenskapliga publikationer](#)

## 8 Nordisk utblick

### 8.1 Danmark

Danmarks Statistik, motsvarigheten till SCB, har tagit fram en manual för dansk utbildningsstatistik.<sup>78</sup> Manualen ger både en översikt och en detaljerad genomgång av de olika delarna i utbildningsstatistiken. Den vänder sig både till interna användare på Danmarks Statistik och till externa användare som i samband med forskning eller analyser behöver kunskap om utbildningsstatistiken och de olika register som den består av.

I manualen skriver man att syftet med utbildningsstatistiken är att skapa ett registerbaserat underlag för offentlig utbildningsstatistik och att publicera statistik om grundutbildning. Dessutom är syftet att skapa en omfattande datagrund för registerbaserad forskning om villkoren för utbildning.

Övergripande försöker utbildningsstatistiken täcka följande områden med registerdata:

1. Utbildningsverksamhet i Danmark inom det ordinarie utbildningssystemet samt vuxenutbildningssystemet i följande dimensioner:
  - a. Antal elever (personer/kursdeltagare/årselever) som ett uttryck för "belastningen" av utbildningssystemet.
  - b. Elevflöden: inflöde och utflöde (avslutade/avbrutna) och härledda mått på produktion/produktivitet på enskilda program, utbildningsområden eller i karriärvägar.
  - c. Studenters rörlighet över gränserna – både när det gäller hela utbildningsprogram och delar av utbildningsprogram (utbytesstudenter).
  - d. Betyg etc. inom specifika områden som ett mått på utbildningens kvalitet eller som ett mått på den personliga utbildningsstyrkan.
2. Den bofasta befolkningens utbildningskvalifikationer oavsett var kvalifikationerna erhöles.

I utbildningsstatistiken finns följande grundläggande register:

1. Elevregistret (UDD\_ELEV), som täcker utbildningsprogram inom det ordinarie utbildningssystemet. Registret omfattar både utbildningsaktiviteter och kvalifikationer och finns tillgängligt i en "rå" version (motsvarande rapporterna från institutionerna) samt en bearbetad version. Nedan följer betygsregister som täcker utvalda områden:
  - a. Det enskilda ämnesregistret (UDD\_KARAKTER\_GRUNDSKOLE) med information om ämnen och betyg i grundskolan och gymnasiet.

---

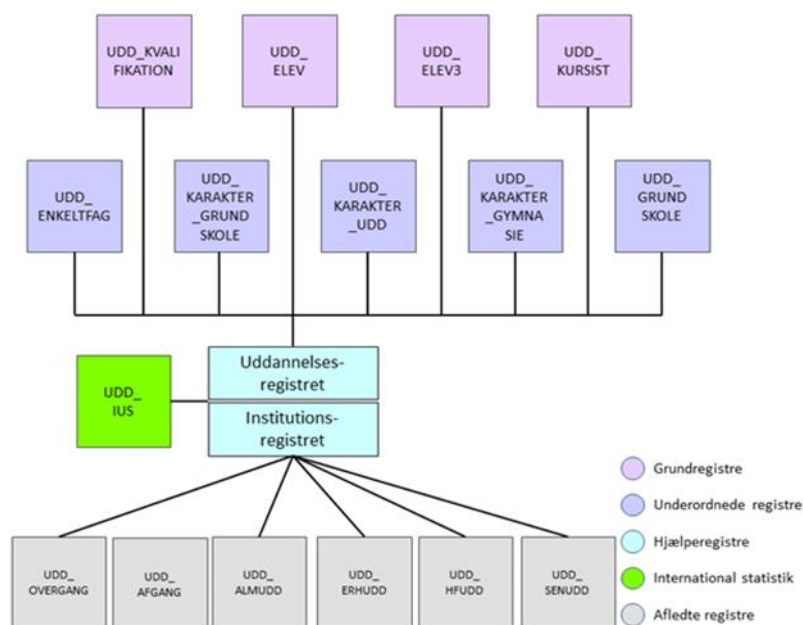
<sup>78</sup> [Danmarks statistik: Uddannelsesstatistikens manual](#)

- b. Ämnesregistret (UDD\_KARAKTER\_GYMNASIE) med information om ämnen och betyg i gymnasiet.
  - c. Ämnesregistret (UDD\_ENKELTFAG) med information om ämnen från grundskolan och gymnasiet utan betyg.
  - d. Betygsregistret (UDD\_KARAKTER\_UDD) med betygsmedelvärden för gymnasieutbildning, yrkeshögskola.
  - e. Grundskoleregistret (UDD\_KARAKTER\_GRUNDSKOLE) med information om ämnen och betyg i grundskolan.
  - f. Grundskoleregistret (UDD\_GRUNDSKOLE) med information om specialundervisning.
2. Elevregistret (UDD\_KURSIKT) som omfattar kursverksamhet inom vuxenutbildningssystemet.
  3. Kvalifikationsregistret (UDD\_KVALIFIKATION), som innehåller information om kvalifikationer som härrör från:
    - a. Folk- och bostadsräkningen 1970 (Hist).
    - b. Invandrades utbildning (IVU).
    - c. Kvalifikationer som erhållits inom vuxenutbildningssystemet.
    - d. Kvalifikationer från auktorisations- eller medlemsregister.
    - e. Kvalifikationer som erhållits utomlands och godkänts av Styrelsen för International Uddannelse.
    - f. Kvalifikationer från Grönland.
    - g. Kvalificerande program.

Utöver dessa register med primärdata finns det tre härledda register för vilka elevregistret och kvalifikationsregistret utgör grunden:

1. Befolkningens utbildning har 4 aspekter
  - a. Högsta allmänna utbildning (UDD\_ALMUDD)
  - b. Högsta yrkesutbildning (UDD\_ERHUDD)
  - c. Högsta avslutade utbildning (UDD\_HFUDD)
  - d. Senast avslutade utbildning (UDD\_SENUDD)
2. Övergångsregister
3. Avgångsregister

Det logiska förhållandet mellan de grundläggande registren illustreras i tabellöversikten nedan.



Datansamlingen organiseras som folkräkningar (eller insamlingar), där varje "folkräkning" täcker ett område inom det reguljära utbildningssystemet.

Insamlingarna är följande:

- Skolor, omfattar grundskoleutbildning.
- Gymnasier, omfattar allmän och yrkesinriktad gymnasieutbildning.
- Företag, omfattar yrkesskolor och skolor med yrkesinriktad gymnasieutbildning.
- Högskoleutbildning, omfattar yrkesakademier och högskolor.
- Universitet, omfattar kandidat- och masterprogram vid universitet.
- PhD, omfattar universitet och andra utbildningsinstitutioner som erbjuder doktorandprogram.

Det finns ett antal olika elev- (eller student-) administrativa system som är anpassade till behoven hos olika typer av utbildningsinstitutioner. I praktiken definieras en insamling av de olika elevadministrativa system från vilka insamlingen tar emot data, eftersom insamlingen har ingått avtal med systemleverantörerna inom insamlingsområdet om formatet och innehållet i elevrapporten. Utöver de elektroniska rapporterna får insamlingen också ett litet antal Excel-rapporter från institutioner som inte använder ett elevadministrativt system – vanligtvis små institutioner.

### 8.1.1 LER

Den utbildningsstatistiska manualen är inte uppdaterad med ett register, kallat LER, som nyligen tillgängliggjorts. LER-folkeskole syftar till att berika grundskolestatistiken med information på lektionsnivå.<sup>79</sup> Den information som kopplar elever till lärare och aktiviteter finns tillgänglig via de digitala lärplattformarna AULA och UniLogin, som används i stor utsträckning för schemaläggning, administration och pedagogiska tjänster i det danska

<sup>79</sup> [Danmarks statistik: LER grundskole](#)

skolsystemet. Dessa datakällor gör det möjligt att koppla samman elever i grupper, se vad grupperna undervisas i och vem som undervisar. Detta gör det möjligt att ge en närmare statistisk bild än någonsin tidigare av den undervisning som äger rum i danska grundskolor och ger nya möjligheter att undersöka de faktorer som påverkar betyg, vidareutbildning, avhopp från skolan osv.

Det som gör det möjligt att skapa ett register som LER är att utbildningsdepartementet ställer tydliga krav på leverantörerna av utbildningsadministrativa system. Skolorna inom ungdoms-, vuxen- och fortbildning väljer själva vilket elevhanteringssystem de vill använda, till exempel genom ett anbudsförfarande. Till skillnad från tidigare finns det inte ett enda centraliserat elevadministrativt system för ett utbildningsområde, till exempel yrkesutbildning. Istället kan varje skola eller lärosäte välja mellan de studentadministrativa system som finns på marknaden. Systemen erbjuds av privata leverantörer och stödjer var och en ett eller flera av de programområden som omfattas. Men departementet ställer bland annat krav på att de studieadministrativa systemen ska ha integrationer mellan de studieadministrativa systemen och de centrala systemen vid utbildningsdepartementet och några andra myndigheter.<sup>80</sup>

### 8.1.2 STIL

Danska utbildningsdepartementets avdelning för it och lärande (på danska Styrelsen for It og Læring – STIL) tillgängliggör även statistik och indikatorer om grundskolan. I grundskoleområdet kan man följa utvecklingen av elev- och skolresultat, elevernas välbefinnande och frånvaro.<sup>81</sup> Här finns även uppgifter om bl.a. antalet anställda, antalet elever per lärare, andelen av personalens tid som ägnas åt eleverna och kostnaden för grundskolan per elev. Tillgång till detaljerade data kräver emellanåt inloggning, men på aggregerad nivå är statistiken öppet tillgänglig.

### 8.1.3 Avslutande kommentar

Sammanfattningsvis har Danmark i flera avseenden ett ur forskningssynpunkt mer gynnsamt utgångsläge eftersom man samlar in och tillgängliggör data som gör det möjligt att koppla samman elever i grupper, se vad grupperna undervisas i och vem som undervisar. LER-registret ger unika möjligheter men är ännu så nytt att det sannolikt ännu inte resulterat i så mycket forskningsrapporter. Registret har varit möjligt att upprätta på grund av två omständigheter. Den första är bestämmelser som skapar skyldigheter för leverantörer av lärplattformar att tillgängliggöra sin data för staten. Den andra är att Danmark tycks göra en annan tolkning av GDPR och vad som kan anses vara integritetskänsliga uppgifter.

---

<sup>80</sup> [Børne- og Undervisningsministeriet - Styrelsen for It og Læring: Rammer for anvendelse af studieadministrative it-systemer](#)

<sup>81</sup> [Børne- og Undervisningsministeriet: Uddannelsesstatistik - Grundskole](#)

## 8.2 Finland

### 8.2.1 Vipunen

Vipunen är utbildningsförvaltningens statistiktjänst.<sup>82</sup> Undervisnings- och kulturministeriet och Utbildningsstyrelsen svarar gemensamt för dess innehåll. Statistiken grundar sig på Statistikcentralens, undervisnings- och kulturministeriets och Utbildningsstyrelsens uppgifter och register.

I tjänsten finns statistik- och indikatoruppgifter för utbildningen inom olika sektorer och arbete efter utbildningen, forskning som har bedrivits vid högskolorna samt befolkningens utbildningsstruktur och studerandenas socioekonomiska bakgrund.

Vipunens innehåll indelas i följande sektorer:

- Gemensamma för utbildningen
- Småbarnspedagogik
- Förskoleundervisning och grundläggande utbildning
- Gymnasieutbildning
- Yrkesutbildning
- Yrkehögskoleutbildning
- Universitetsutbildning
- Högskoleutbildning och FoU-verksamhet
- Fritt bildningsarbete
- Befolknings-, utbildnings- och yrkesstruktur

I Vipunen finns information om bland annat följande ämnen:

- sökande och antagna
- studerande och examen
- tidigare och senare utbildning
- studieförlopp
- placering efter avlagd examen
- nätverk av utbildningsanordnare
- personal

Vidare finns information om en del utbildningssektorer gällande följande teman:

- internationalisering
- mobilitet
- ekonomi
- utbildning som inte leder till examen

Gällande yrkehögskolor och universitet finns också statistik om studieprestationer, studeranderespons, publikationer samt forsknings- och utvecklingsarbete.

---

<sup>82</sup> [Utbildningsförvaltningens statistiktjänst: Vipunen](#)

Vipunen presenterar i huvudsak statistik från 2000-talet. Databasen innehåller också statistik om forskning som har bedrivits vid högskolorna, befolkningens utbildningsstruktur samt studerandenas socioekonomiska bakgrund. Statistiken baserar sig på uppgifter och register som Statistikcentralen har sammanställt eller på informationsresursen Koski som upprätthålls av Utbildningsstyrelsen och dit utbildningsanordnarna för in uppgifter.

Statistikuppgifterna som publiceras i tjänsten grundar sig på uppgifter och register som Statistikcentralen, Undervisnings- och kulturministeriet samt Utbildningsstyrelsen samlat in.<sup>83</sup>

### 8.2.2 Statistikcentralen

Statistikcentralen har samlat in uppgifter om studerande och examina vid utbildningsanordnarna direkt från läroanstalterna, datalagret Koski, högskolornas riksomfattande datalager (VIRTA) och studentexamensregistret. Statistikcentralen för ett examensregister över de examensuppgifter som de samlar in. Uppgifterna i examensregistret kompletteras av Statistikcentralen bland annat med befolkningsuppgifter från Myndigheten för digitalisering och befolkningsdata samt uppgifter i det egna sysselsättningsregistret. Sysselsättningsregistrets personpopulation utgörs av den befolkning som på årets sista dag är stadigvarande bosatta i landet. Uppgifterna bygger i regel på förvaltningsregister och statistiska material. Statistikcentralen har en webbsida där statistik om utbildning och forskning finns samlad.<sup>84</sup>

En del av Statistikcentralens material grundar sig på en kombination av ovan nämnda individbaserade material och register. I sådant material kan man till exempel granska genomströmning av utbildning (studiernas förlopp), avbrytande av utbildning eller arbetslivssituation för personer som avlagt examen. Dessutom samlar Statistikcentralen regelbundet in uppgifter om lärare på grundskolenivå och andra stadiet samt inom det fria bildningsarbetet och den grundläggande konstundervisningen.

### 8.2.3 Utbildningsstyrelsen

Utbildningsstyrelsen (motsvarigheten till Skolverket) samlar in de kostnads- och verksamhetsuppgifter som utgör grunden för statsandelen av kostnaderna för utbildning på grundskolenivå och andra stadiet av utbildningsanordnarna. Undantag är kommunernas kostnadsuppgifter för förskoleundervisning, grundläggande utbildning och gymnasieutbildning som samlas in av Statskontoret.

Utbildningsstyrelsen upprätthåller ett riksomfattande datalager för undervisning och utbildning (Koski)<sup>85</sup>, ett antagningsregister (utbildning och

---

<sup>83</sup> Tyvärr har frågor om mer exakt hur informationsinsamlingen går till, trots upprepade påminnelser, inte besvarats av finländska myndigheter. Det innebär att uppgifterna i texten är hämtade från olika webbsidor.

<sup>84</sup> [Statistikcentralen: Utbildning och forskning](#)

<sup>85</sup> [Utbildningsstyrelsen: KOSKI-tjänsten](#)



högskoleutbildning efter den grundläggande utbildningen) samt ett datalager för småbarnspedagogiken, motsvarande förskolan (Varda)<sup>86</sup>, vars uppgifter används i statistikföringen. Koski innehåller uppgifter i den personliga utvecklingsplanen för kunnandet (ePUK) som sparats i registret för studerande inom yrkesutbildningen.

Utbildningsstyrelsen samlar in uppgifter från högskolorna om studerandenas internationella utbytesperioder.

#### **8.2.4 Undervisnings- och kulturministeriet**

Undervisnings- och kulturministeriet samlar in uppgifter från yrkeshögskolorna och universiteten om deras verksamhet, personal och ekonomiska verksamhet.

Forskning.fi är en tjänst som tillhandahålls av undervisnings- och kulturministeriet och som samlar och sprider information om forskning i Finland.<sup>87</sup> Tjänsten förbättrar tillgången till forskningsinformation och experter samt ökar den finländska forskningens synlighet och samhällspåverkan. För närvarande innehåller tjänsten information om bl.a. det finländska forskningssystemet, finländska organisationers publikationer, projekt som finansieras av offentliga och privata forskningsfinansiärer, forskare som är verksamma i Finland och forskningsaktiviteter, samt statistik om utvecklingen av forskningsresurserna och forskningens genomslag.

I tjänsten uppdateras automatiskt uppgifter som redan finns i bl.a. universitetens, yrkeshögskolornas, forskningsinstitutens, universitetssjukhusens och forskningsfinansiärernas system. Tjänsten togs i bruk i juni 2020 och utvecklas stegvis.

#### ***Övriga***

Studentexamensnämnden samlar in vitsorden för studentexamen. Informationstjänsten för genomslag i utbildningen Arvo samlar in uppgifter om karriäruppföljning, studeranderespons och arbetslivsrespons.

Andra informationskällor som används i Vipunen är Företags- och organisationsdatasystemet (FODS), ETER (European Tertiary Education Register), OECD, Web of Science (biometri), Scopus (bibliometri), CORDIS (European Commission).

#### **8.2.5 Finlands samhällsvetenskapliga dataarkiv**

Finlands samhällsvetenskapliga dataarkiv (FSD) tillhandahåller information och vägledning för att stödja forskning, undervisning och studier.<sup>88</sup>

Informationstjänster och vägledning Arkivet erbjuder råd och vägledning i hur

---

<sup>86</sup> [Utbildningsstyrelsen: Informationsresursen inom småbarnspedagogiken Varda](#)

<sup>87</sup> [Undervisnings- och kulturministeriet: Forskning.fi](#)

<sup>88</sup> [Finnish Social Science Data Archive](#)

man hittar lämpliga forskningsdata för dina ändamål. Man tillhandahåller också information om datahantering och långsiktigt databevarande.

Riktlinjer för datahantering som tillhandahålls av arkivet ger tydliga och konkreta instruktioner om hur man skapar datahanteringsplaner, hanterar digitala forskningsdata och informerar forskningsdeltagare.<sup>89</sup>

FSD utgör en portal för ett brett utbud av digitala forskningsdata för lärande, undervisning och forskningsändamål. Användare av tjänsterna är främst samhällsvetare, men man tillhandahåller också tjänster och data till forskare, lärare och studenter inom humaniora och hälsovetenskap. Alla grundläggande tjänster är kostnadsfria. Arkivet är en separat enhet vid Tammerfors universitet men tillhandahåller tjänster till forskare, lärare och studenter i hela Finland samt utomlands. Man deltar också aktivt i internationellt samarbete. Finlands Akademi rekommenderar att forskare deponerar sina forskningsdata i ett lämpligt arkiv och nämner FSD som ett sådant arkiv. Många andra forskningsfinansiärer har redan följt Akademiens exempel.

Det finns mer än tusen kvantitativa dataset och cirka tvåhundra kvalitativa dataset i FSD:s samlingar, och antalet ökar hela tiden. Många kvantitativa undersökningar har redan översatts till engelska (frågor och svarsalternativ). Fler översätts på begäran, kostnadsfritt, för användare som inte talar finska.

Populära undersökningsserier på Aila:

- Undersökningar om utvecklingssamarbete
- Finländska nationella valundersökningar
- EVA-undersökningar om finländska värderingar och attityder
- Finländska ungdomsundersökningar
- Polisbarometrar
- Undersökningar om åsikter om säkerhetspolitik
- Profil för skolvälbefinnande

Det finns mer än 1 800 dataset i FSD:s samlingar, som alla är tillgängliga för forskningsändamål. Data Management Guidelines innehåller användbar information om forskningsplanering, datainsamling och hantering av avtal under forskning samt konkreta tips för datahantering i olika skeden av datans livscykel.

Aila Data Service innehåller alla datamängder i FSD: s innehav och detaljerade studiebeskrivningar finns både på finska och på engelska. I Ailas sökgränssnitt kan man söka i beskrivningar av studier samt frågor och variabler.

### 8.2.6 Avslutande kommentar

Statistikinsamlingen i Finland tycks likna den i Sverige. De statliga myndigheterna samlar in uppgifter från den kommunala nivån respektive från lärosäten, och sammanställer statistiken. Statistikutbudet och tillgängligheten verkar också vara ungefär likartad som här.

---

<sup>89</sup> [Finnish Social Science Data Archive: Data Management Guidelines](#)

## 8.3 Norge

### 8.3.1 Nationella utbildningsdatabasen

Norge har en central databas kallad Nationella utbildningsdatabasen (NUBD). Här samlas all statistik från grundskolan till avslutad forskarutbildning i en och samma databas. Statistiken är på individnivå och går tillbaka till 1970. En innehållslista visar att det enbart är personstatistik i NUBD.<sup>90</sup>

Här finns register som omfattar demografiska data samt befolkningens utbildningsnivå. Vidare alla kurser som har pågått eller slutförts vid utbildningsinstitutioner i Norge. Här inkluderas även all avslutad utbildning som registrerats i folkräkningen och andra källor till information om utbildningsnivå. Här finns resultat från nationella prov, betygsdata från avslutad grundskola och examensbevis samt ämnesbetyg från gymnasiet.

Utdanningsdirektoratet, som ungefär motsvarar Skolverket, har även statistik över kostnader och personal samt förskolan.<sup>91</sup> Vidare anges elevers frånvaro i årskurs 10, en uppgift som även förs in i elevernas examensbevis. Här finns uppgifter om antal elever per kommun, per stadium och antal elever med specialundervisning. Lärartäthet anges också per kommun och stadium. Vidare finns uppgifter om antal lärare per kommun, stadium och ämne som ansökt om respektive genomfört kompetensutveckling samt om de är godkända av Utdanningsdirektoratet. Här finns också statistik över elevers betyg nedbrutet per län, kommun och skolenhet samt resultat på nationella prov i årskurs 5, 8 och 9 med mera.

Insamlingsmetoderna motsvarar de i Sverige, det vill säga att skolhuvudmännen ansvarar för att rapportera in uppgifter om sina skolor, lärare och elever. Det gäller både för skolnivån, yrkeshögskolorna och högskolorna. Se mer information om detta nedan.

### 8.3.2 Database for statistikk om høyere utdanning

Vidare finns en databas för statistik om högre utbildning kallad DBH, som delvis överlappar NUBD.<sup>92</sup> DBH innehåller ett brett utbud av information om högre utbildning i Norge, inkluderande universitet, högskolor och specialiserade skolor. Databasen etablerades 1991 på initiativ av Utbildnings- och forskningsdepartementet, med syftet att samla in och tillgängliggöra data om högre utbildning och forskningssektorn i Norge. Idag innehåller DBH uppgifter om studenter, utbildning, forskning, personal, museer, lokaler, ekonomi och ägande av företag. Uppgifterna är öppet tillgängliga för forskning inom sektorn, för planering både centralt i departementen och lokalt på varje enskild institution.

---

<sup>90</sup> [Statistisk sentralbyrå: NUDB - innholdsliste](#)

<sup>91</sup> [Utdanningsdirektoratet: Statistikk](#)

<sup>92</sup> [Direktoratet for høyere utdanning og kompetanse: Database for statistikk om høyere utdanning](#)

Databasen innehåller både uppgifter på personnivå om studenters ansökningar till högre utbildning och deras meritpoäng, kurser de genomfört, examina de erhållit, lämplighetsbedömningar för utbildning till exempel till sjuksköterska, grundskollärare, förskollärare, socialt arbete med mera. Det finns också register över samtliga personer som deltar i forskarutbildning och avlagda doktorsexamina. Vidare finns institutionella data om lärosätens ekonomi och lokalsituation, vetenskaplig publicering, tillsatta tjänster med mera.

DBH-F är en nationell databas för statistik om yrkeshögskoleutbildning som omfattar alla yrkeshögskolor i Norge.<sup>93</sup> Även den innehåller data både på individnivå och organisatoriska data. På motsvarande sätt som vad gäller högskolan finns statistik om sökande och antagna studenter, vilka utbildningar de går, vilken typ av studiefinansiering de har, studiepoäng och examensdata med mera. På organisatorisk nivå finns uppgifter om hela studieutbudet och möjlighet att jämföra utbildningarna med avseende på utbildningsform, finansieringsform, studieavgifter, studielån etc. Vidare finns skolornas ekonomiska redovisningar och årsberättelser redovisade. Det finns även ett register över samtliga huvudmän där alla utbildningar de anordnar anges.

### 8.3.3 Grundskola och gymnasium

Uppgifter om grundskoleutbildning samlas in årligen med 1 oktober som insamlingsdatum. Uppgifterna samlas in via informationssystemet för grundskolor (GSI)<sup>94</sup>, som ägs av det norska Utdanningsdirektoratet (Udir), vilka även ansvarar för insamlingen. Den enskilda skolan och kommunen är gemensamt ansvariga för att samla in data för sitt område. I regel matar skolorna in uppgifterna först och därefter kontrollerar kommunen uppgifterna. Den statliga administratören (tidigare fylkesmannen/landshövdingen) ansvarar för att samla in uppgifter från privata, länsvisa och statliga skolor. Efter att skolorna och kommunerna har fyllt i och kontrollerat sina uppgifter tar statsadministratören över uppgifterna och genomför sina kontroller. När den statliga förvaltaren är klar tar Udir och Statistisk sentralbyrå (SSB) över uppgifterna och utför sina egna kontroller. SSB och Udir publicerar uppgifterna samtidigt. SSB publicerar ett urval av uppgifterna, medan Udir publicerar mer detaljerade uppgifter. SSB använder också grundskoledata i publiceringen av KOSTRA<sup>95</sup>, där relevantalet är kopplat till kommunernas redovisningssiffror. Mer information om grundskolestatistiken finns under "Om statistiken" på statistiksidan.<sup>96</sup>

Data från GSI är aggregerade på skol-/kommunalnivå och inte på individnivå. SSB publicerar dock även betyg i slutet av grundskolan, samt resultat från nationella prov i grundskolan, som är tillgängliga på individnivå. Rapporteringen av dessa uppgifter görs av skolor som rapporterar betyg till fylkeskommunernas

<sup>93</sup> [Direktoratet for høyere utdanning og kompetanse: DBH Fagskolestatistikk](#)

<sup>94</sup> [Utdanningsdirektoratet: Grunnskolens Informasjonssystem \(GSI\)](#)

<sup>95</sup> [Statistisk sentralbyrå: KOSTRA](#)

<sup>96</sup> [Statistisk sentralbyrå: Elevar i grunnskolen](#)

administrativa datasystem för antagning till gymnasieutbildning (VIGO). De rapporterar också resultat från nationella prov genom testadministrationssystemet (PAS), som ägs av det norska direktoratet för utbildning och yrkesutbildning.<sup>97</sup>

### 8.3.4 Gymnasieutbildning

Dataunderlaget för elever, lärlingar, lärlingskandidater och sökande i gymnasieskolan är registerdata som hämtas från fylkeskommunernas intagningsystem VIGO, i vissa fall även separata registreringsystem för vuxna i gymnasieskolan. VIGO:s huvudsakliga funktion är att administrera antagningen till de olika kursnivåerna inom gymnasieskolan. VIGO-systemet integrerar också uppgifter om elever vid fristående skolor och statliga skolor som drivs enligt skollagen genom separata insamlingsförfaranden. VIGO omfattar även yrkesutbildning, dvs. personer som lär sig i ett företag (lärlingar/lärlingskandidater) och som gör ett yrkesprov/kompetensprov. En översikt över lärlingar/lärlingskandidater och yrkesprov/kompetensprov skickas från yrkesutbildningskontoret i länet.<sup>98</sup>

### 8.3.5 Datainsamling på högre nivå (universitet och högskolor, yrkeshögskolor)

Datainsamlingen organiseras av Database for statistikk om høyere utdanning (DBH), som tar emot data från varje lärosäte.<sup>99</sup> DBH har en separat avdelning för yrkeshögskoledata respektive universitets- och högskoledata. SSB samlar in uppgifter från båda avdelningarna vid DBH. Uppgifter från universitet och högskolor och uppgifter från yrkesskolor behandlas också separat hos SSB.

#### *Databehandling/kontroller*

Maskinella kontroller utförs på utbildningsdata efter mottagandet i SSB. Dessa omfattar borttagning av identiska dubletter, kontroll av variationer mellan årets och förra årets datamaterial och test av om observationerna har giltiga värden för variablerna i förhållande till kravspecifikationen. Födelse- och personnummer kontrolleras för giltighet och SSB lägger till identifierare från andra register, t.ex. bosättningsinformation och invandrarinformation. Kontroller utförs också under och efter revideringen och tillägget av nya variabler för att säkerställa att de tillagda variablerna är korrekt placerade och att inga fel uppstår i det statistiska underlaget som ett resultat av den faktiska bearbetningen.

---

<sup>97</sup> Man kan läsa mer på SSBs statistiksida för resultat i grundskolan [Statistisk sentralbyrå: Karakterer og nasjonale prøver i grunnskolen](#)

<sup>98</sup> För mer information om datainsamling, se SSB:s statistiksida [Videregående opplæring og annen videregående utdanning](#)

<sup>99</sup> [Direktoratet for høyere utdanning og kompetanse: Database for statistikk om høyere utdanning](#)

### ***Nationell utbildningsdatabas***

Alla uppgifter från grundskola till universitet och högskolor läggs slutligen in i en gemensam datafil som också läggs in i den nationella utbildningsdatabasen (NUBD).<sup>100</sup>

NUDB är en longitudinell databas som gör det möjligt att sammanställa siffror från många år. Detta gör det möjligt att genomföra bredare och större analyser inom utbildningsområdet. NUDB möjliggör kompletteringsanalyser av t.ex. längd och val av studieprogram genom hela utbildningsvägen.<sup>101</sup>

### ***Datas tillgänglighet***

Uppgifterna görs tillgängliga på SSB:s statistiksidor genom tabeller samt artiklar, analyser och rapporter. Uppgifterna kan också beställas i tabellform som betalda uppdrag om kunderna vill ha en sammanställning av uppgifter som inte publiceras på webbplatsen. SSB lagrar datafiler för långtidsförvaring och tillhandahåller data för forskning.<sup>102</sup>

Forskare, studenter och anställda i departement och direktorat kan använda microdata.no, en lösning som gör mikrodata tillgängliga för forskning utan att man behöver ansöka om tillgång.<sup>103</sup> Flera av utbildningsvariablerna är tillgängliga via microdata.no, och målet är att göra ännu fler data tillgängliga i denna lösning i framtiden.

### **8.3.6 Avslutande kommentar**

Även den norska insamlingen av utbildningsstatistik verkar likna den svenska. Samma typer av data samlas in och tillgängliggörs.

## **8.4 Sammanfattning av nordisk utblick**

Sammanfattningsvis kan det sägas att utblicken över de övriga nordiska länderna ger intrycket att det såväl i Danmark som Finland och Norge råder större tydlighet och att det där finns bättre möjlighet till överblick över vilka datamängder som finns tillgängliga för forskning. Man tycks också mer angelägen att tillgodose forskares önskemål om tillgång till data.

Inte minst Danmark förtjänar att lyftas fram som ett föredöme. Trots att man har ett lika decentraliserat utbildningssystem som det svenska, finns det tydliga riktlinjer från departementet om vilka krav som ställs på olika aktörer ifråga om datahantering. Det svenska systemet känns mer präglad av en ad hoc-uppbyggd verksamhet där ramar och riktlinjer är otydligare.

---

<sup>100</sup> [Statistisk sentralbyrå: Tema- og variabelliste for NUDB, med kodelister, 6. november 2023](#)

<sup>101</sup> Se även mer information på [Statistisk sentralbyrå: Utdanning](#)

<sup>102</sup> [Statistisk sentralbyrå: Tilgang til data fra SSB](#)

<sup>103</sup> [Statistisk sentralbyrå: microdata.no – registerdata uten å søke](#)

## 9 Behov av data och infrastruktur

### 9.1 Behov av ytterligare datamängder

#### **Öka tillgängligheten genom bättre överblick**

Vad gäller tillgängligheten på nationell nivå så är den tekniska tillgängligheten god, det vill säga att data vanligen är publicerade i format som gör dem möjliga att bearbeta. Den fysiska tillgängligheten däremot skulle kunna bli bättre. Det finns brister i överskådlighet, det är inte alltid lätt att veta vilka datamängder som finns och vem som ansvarar för dem. Tillgängligheten till data skulle öka betydligt för den som inte känner till vilka datamängder som existerar om det fanns en beskrivning över strukturen och innehållet i svensk utbildningsstatistik. Rekommendationen är att öka tillgängligheten genom att ta fram en manual för svensk utbildningsstatistik efter förebild från Danmark (se avsnitt 8.1). Bilaga 1 och 2 är ett första försök att öka överblicken över tillgängliga datamängder.

#### **Möjlighet att koppla undervisning till lärare**

Trots mängden utbildningsdata saknas viss typ av data som potentiellt skulle kunna öka värdet på tillgängliga datamängder. Det är framför allt data som gör det möjligt att koppla samman undervisningsgrupper, och kanske enskilda studerande, med lärare. Elevrapporteringssystemen som huvudmännen har grundas ofta i schemaläggning där det framgår vilka elever och lärare som möts i olika lektioner. Det är ett rikt material som inte är tillgängligt för forskning just nu. Med tillgång till sådan data skulle forskare kunna använda data om studieresultat, kostnader, kompetensutbildning och lokaler med mera på ett betydligt intressantare sätt.

Här förefaller den danska databasen LER vara ett lysande undantag som samlar in uppgifter som gör det möjligt att koppla samman elever i grupper, se vad grupperna undervisas i och vem som undervisar. Som framhölls tidigare gör dessa data det möjligt att ge en närmare statistisk bild än någonsin tidigare av den undervisning som äger rum i danska grundskolor och ger nya möjligheter att undersöka de faktorer som påverkar betyg, vidareutbildning, avhopp från skolan osv.

Motsvarande svenska data på alla utbildningsnivåer vore oerhört värdefulla. Men förutom svårigheter att faktiskt samla denna data ur svenska skolors verksamhetssystem utan att först ha standarder och överenskommelser på plats för hur man definierar olika begrepp, i vilket format uppgifterna ska samlas in och exporteras och att detta ska ske betydligt oftare än dagens datainsamling (som vanligen sker 1–2 gånger per år), så föreligger också skilda tolkningar av GDPR. I Danmark och i viss mån även i Norge, tycks man ha snävare tolkningar av var gränserna för personlig integritet går – det tycks vara enklare att samla in och bearbeta data – förfaranden som sannolikt inte skulle tillåtas i Sverige.

Slutsatsen blir att förutom att skapa förutsättningar för att samla in data som möjliggör att koppla ihop undervisningsgrupper med lärare, så behövs också diskussioner – gärna på nordisk eller europeisk nivå – om hur GDPR ska tolkas.

### **Data om elevers frånvaro**

Som redan påpekats är det en påtaglig brist att det för närvarande saknas en definition i skollagen av giltig och ogiltig frånvaro och att det inte förs nationell statistik om ogiltig frånvaro. Det är den enskilda skolan, och sannolikt ofta den enskilda läraren, som bedömer om en elev är frånvarande eller inte och om frånvaron är giltig eller inte. Det finns svårigheter att bygga upp en statistik runt detta, men området är så angeläget att man borde försöka. Detta är primärt inte en fråga för Vetenskapsrådet men myndigheten skulle kunna bidra till att frågan aktualiseras ytterligare.

### **Data från nationella prov**

Skolverkets redovisning av nationella prov bör redovisa resultat, inte bara betyg. Uppgifterna finns men redovisas inte. Därmed kan forskare inte se hur elever klarar sig mellan betygsstegen. Här skulle en enkel åtgärd kunna betyda mycket.

Skolverket redovisar inte heller resultat från så kallade ersättningsprov, det vill säga nationella prov som används när uppgifter från det planerade provet läckt ut. Enligt uppgift i intervjuerna kan detta innebära att det något år är så få som 20 procent av resultaten som redovisas.

Nationella proven i årskurs 3 i grundskolan är inte anpassat för att mäta kunskaper över hela populationen, utan för att identifiera lågpresterande elever. De enda betygsnivåer som finns är godkänt och underkänt. Eftersom i stort sett alla elever får godkänt ger provet ytterst begränsad information, trots att det är en stor och kostsam process att genomföra nationella prov.

Ytterligare ett önskemål när det gäller data från nationella prov avser det faktum att det för närvarande är oklart när en elev skrev det nationella provet på en viss kurs och när eleven fick betyg på samma kurs. En elev kan med andra ord ha skrivit provet VT2022 men inte betygssatts förrän exempelvis VT2023 men det går inte att se i datan. Det vore därför önskvärt om registren för kursbetyg från gymnasiet innehåller en anteckning om provdatum respektive betygsdatum (eller i varje fall termin).

### **Möjlighet att koppla register på mikronivå till internationella undersökningar**

Det finns ett stort behov av att kunna koppla svenska register på mikro- eller elevnivå med de internationella undersökningar som görs inom skolområdet, till exempel PISA, PIRLS, TIMSS, ICCS, ICILS, TALIS etc. Detsamma gäller de nationella undersökningar som görs av antingen urval av eller hela populationer, nationella prov på uppgiftsnivå, utvärderingar av olika slag. Idag saknas information på uppgifts- eller ämnesnivå för de nationella proven, och data från enkätundersökningar och andra kunskapsundersökningar som Skolverket gör är



inte automatiskt tillgängliga för forskare. Att få tillgång till data på uppgiftsnivå jämte de aggregerade mått som skapats i efterhand av Skolverket är nödvändigt av flera skäl, dels för att kunna undersöka uppgifternas mätegenskaper men också för att forskare ska kunna skapa egna aggregerade mått. En sådan tillgång kräver en prövning från databärande myndighet om någon sekretess skulle hindra ett utlämnande av data när den är på individnivå.

### **Ökad transparens från Skolverket**

Det finns från forskarhåll en viss oro över utvecklingen att Skolverket får allt större ansvar för statistiken på skolområdet. Denna oro gäller det faktum att det inte längre är expertmyndigheten på statistik och databaser (SCB) som ansvarar för databasernas uppbyggnad och skötsel. Oron gäller också att Skolverket historiskt varit mindre transparenta och konsekventa i vilka data som lämnas ut till forskning och vilka som inte gör det, jämfört med SCB. Det finns också en undran om Skolverket inte kommer att uppleva att den egna verksamheten granskas av forskarna och att man för att skydda sin egen verksamhet blir mindre generös än vad den oberoende statistikmyndigheten SCB varit, i att lämna ut data. Vetenskapsrådet uppmanas att inleda en dialog med Skolverket i dessa frågor.

### **Möjlighet att jämföra lärosäten**

Som framhållits tidigare finns det endast begränsade möjligheter att jämföra högskoleutbildningar vid olika lärosäten. Mest intressant vore det att kunna jämföra olika programutbildningar, som ju har gemensamma utbildningsmål men där undervisningspraxis och organisation sannolikt skiljer sig en del mellan utbildningsorterna. Motsvarande jämförelser sker återkommande inom skelsektorn. Där har det till och med tidvis upprättats rankinglistor över bästa och sämsta skolkommun.

Det är viktigt att framhålla att systematiska jämförelser inte är till för att värdera, utan för att kunna vidareutvecklas – som enskild aktör och tillsammans. Högskolepedagogik är förvisso ett forskningsfält, men den forskningen rör sig oftast i relationen lärare, student och stoff. Organisatoriska och ekonomiska frågor vägs sällan in.

Med mer gemensamma examinationsförfarande, eller examinationer som genomförs på samma sätt, skulle högskolan öppna för sådana systematiska jämförelser. Det är påtagligt att sådana önskemål ytterst sällan framförs på högskolenivå, vilket möjligen bottnar i att betydligt färre forskare forskar om högskolan och att högskolan i sig nog är ovan och ovillig att se sig som ett forskningsobjekt.

## 9.2 Behov av infrastrukturella insatser

I sin guide till forskningsinfrastrukturen 2023 definierar Vetenskapsrådet vad man anser vara ”forskningsinfrastruktur av nationellt intresse”.<sup>104</sup> Definitionen kompletteras med en beskrivning av vad som kan vara infrastruktur inom samhällsvetenskaplig forskning, som sägs kunna se mycket olika ut. Det kan exempelvis handla om stora frågeundersökningar över levnadsvillkor, attityder, och politiskt handlande, men kan också utgöras av registerdata som samlats in och sammanställts av olika myndigheter, men också breda samlingar av kontextuella data som beskriver hur våra samhällsliga institutioner ser ut och utvecklas över tid.<sup>105</sup>

Den breda användningen av begreppet medför att det väcker olika associationer beroende på vem som tillfrågas. Nedan följer ett urval av frågor och behov som lyfts fram under intervjuerna.

### Register och databaser

- Ökade möjligheter att som forskare få tillgång till mikrodata hos Skolverket och SCB.
- Möjligheter att statistiskt koppla samman elev med lärare och skola, som i Danmark.
- Databaser med tillgång till data från äldre forskningsprojekt (se vidare behov av portal). Den tidigare nämnda portalen ILSA Gateway som ger tillgång till stora internationella undersökningar är ett gott exempel. Göteborgs universitets COMPEATE är ett annat.
- Tillgång till kostnader på skolnivå, inte bara huvudmannanivå. Huvudmännen samlar in kostnadsdata från respektive skolenhet men aggregerar dem innan de rapporteras. Om rapporteringens istället skedde på skolenhetsnivå så skulle denna data kunna bli tillgänglig.

### GDPR-relaterade frågor

- I intervjuerna med forskare har det framkommit att man uppfattar att det är oklara regelverk (GDPR) om lagring framför allt av kvalitativa data. Regelverket tolkas olika mellan lärosätena och man har olika tekniska plattformar för bearbetning. Det skapar svårigheter att samarbeta med andra forskare – var ska bearbetning ske av kvalitativa data? Även här skiljer sig tolkningarna av GDPR mellan lärosäten.
- Olika tolkningar av GDPR på nationell nivå. Svenska forskare får inte tillgängliggöra svenska data till forskare som befinner sig utanför EU/EES, vilket försvårar samarbeten med utländska forskare. Inte heller svenska forskare – som vanligtvis har tillgång till data – ges datatillgång utanför EU/EES, vilket försvårar internationella utbyten. Exempelvis gör Danmark inte samma tolkning vad gäller tillgång till danska data utanför EU/EES, inte heller av vilken data man kan samla in, lagra och bearbeta om elevgrupper och undervisande personal (jämför LER).

---

<sup>104</sup> [Vetenskapsrådet: Vetenskapsrådets guide till forskningsinfrastrukturen 2023 \(pdf\)](#)

<sup>105</sup> Ibid, sid 48f

## Behov av kompetensutveckling

- Behov av kompetensutvecklingsinsatser för forskare runt kvantitativa metoder och databashantering. Även om ytterligare resurser läggs på att bygga portaler för att underlätta för forskare att hitta utbildningsdata, så kommer forskarna även fortsättningsvis behöva lära sig att hitta de datamängder de behöver och att beställa data från SCB och Skolverket.
- Behov av kompetensutveckling i datahantering (insamling, lagring och bearbetning) för skolhuvudmännens administrativa personal, och i viss mån för lärare och skolledare. Detta är inte heller primärt en fråga för Vetenskapsrådet men Vetenskapsrådet kan påverka frågan genom att ta upp en diskussion med skolmyndigheterna.

## Portal till utbildningsdata

- Vidareutveckla Svensk Nationell Datatjänst. Idag innehåller tjänsten forskningsdata från 25 utbildningsvetenskapliga projekt. Sannolikt finns ett stort antal äldre projekt som vore värdefulla att ha tillgängliga. En portal skulle också ge ökad överblick över vilka register som finns av relevans för utbildningsforskning (förhoppningsvis utgör Bilaga 1 och 2 ett steg på vägen mot bättre överblick).

## Utveckla kvalitetsregister

En forskningsgrupp som under ett par år drev ett Vinnovafinansierat projekt runt utbildningsdata, lyfte i en debattartikel i Dagens samhälle 2019 behovet av att inrätta kvalitetsregister för utbildningsdata motsvarande de som finns inom hälso- och sjukvården.<sup>106</sup> I sin slutrapport skriver forskningsgruppen följande:

”Organisationen Nationella Kvalitetsregister<sup>107</sup> finansieras av Sveriges regioner och staten för verksamheten, med 100 miljoner från vardera sida för 2019. Registren har klara riktlinjer för hur de byggs upp och drivs från etablering som registerkandidat upp till slutlig certifieringsnivå 1. Syftet med de nationella kvalitetsregistren är att stödja såväl kliniskt förbättringsarbete som forskning. Det publiceras rapporter där kvalitetsregister använts i vårdens förbättringsarbete och mera övergripande rapporter. Registerhållaren kan även ta fram rapporter på beställning. Den organisation som lämnar data till registret kan se egna data, men inte andras. Det är dock möjligt att jämföra egna data mot mått från alla data exempelvis medel- och medianvärden. Registren samlar in data från alla relevanta organisationer oavsett verksamhetsform: offentliga såväl som privata vårdgivare.

I mötena runt möjligheten att etablera en verksamhet för att förbättra utbildningar av alla slag baserat på data, det vill säga att bedriva evidensbaserad utveckling, framstår alternativet

<sup>106</sup> [Dagens Samhälle: Inrätta kvalitetsregister — och ge inte eleverna datorer!](#)

<sup>107</sup> [Nationella kvalitetsregister](#)

kvalitetsregister som mest lämpligt. Kvalitetsregister stödjer forskning men etableras inte utifrån forskningens prioriteter utan ifrån sektorns förbättringsarbete. Vidare konstaterar vi att persondata från register kan lämnas i aggregerad form genom öppna API:er för tredjeparts användningar. Varje registerhållare är en professionell organisation. Systemmässigt går det förlägga samtliga register till en datacentral med expertis för att driva dem tillförlitligt och säkert samt för att erbjuda digital suveränitet till uppgiftslämnarna.

Fördelar med kvalitetsregister är att konceptet är väl etablerat och bevisat nyttigt inom vårdsektorn; verksamhetsformer, certifiering och validering av register finns etablerade och utgör ett bra utgångsläge för kvalitetsregister för utbildningssektorn. Data samlas in av en trovärdig part med expertis att skydda dem och med tydligt syfte att förbättra utbildningsverksamhet i alla former.

Nackdelar är att det behövs ett långsiktigt arbete för att etablera register med en tydlig huvudman och finansieringsform. Uppgifter av relevans kan vara svåra att representera, såsom pedagogisk metod och olika former av åtgärder på student-, klass och skolnivåer. Insamling av data kan leda till merarbete hos skolor och utbildningsföretag vilket kan leda till lågt deltagande och dålig täckning.”<sup>108</sup>

Vidare skriver man i den tidigare refererade debattartikeln i Dagens samhälle att:

”förslag på register från vårt projekt är nationell sammanställning av frånvarodata med insatta åtgärder i syfte att förhindra studieavbrott. Vidare skulle resultat från nationella prov kunna kompletteras med frågor till elever, föräldrar, lärare och huvudmän för att årligen ta fram och följa framgångsfaktorer för svenska skolor och gymnasier.”

Ett avslutande förslag är att Vetenskapsrådets Utbildningsvetenskapliga kommitté arbetar vidare på förslaget om att upprätta ett kvalitetsregister för utbildningsdata.

---

<sup>108</sup> Slutrapporten är inte offentlig men har gjorts tillgänglig för mig av projektledaren professor Gunnar Karlsson, KTH.

## Bilaga 1 - Översikt över utbildningsstatistiska register<sup>109</sup>

En översikt över de register som är de mest använda i mikrodatauppdrag finns här [Individregister och mikrodata \(scb.se\)](#).

Här följer en översikt med länkar till sidorna som också nås via översiktssidan ovan samt statistikprodukter med register som saknas i översikten ovan. Om det finns flera registervarianter anges, utan länk, för vilka år de finns.

### Skolan

Det är Skolverket som ansvarar för den officiella statistiken om skolan. Statistiken finns här [Sök statistik om förskola, skola och vuxenutbildning - Skolverket](#). Mikrodata beställs från SCB.

### Förskolan

- [Förskola: Barn och personal per 15 oktober \(scb.se\)](#), från 2014
- [Annan pedagogisk verksamhet: Barn och personal per 15 oktober \(scb.se\)](#), från 2014

### Grundskolan

- [Grundskolan: elever per 15 oktober \(scb.se\)](#), från 2008
- [Nationella prov årskurs 3: resultat \(scb.se\)](#), från 2003
- Nationella prov årskurs 5, 2008-2010
- [Nationella prov årskurs 6: resultat \(scb.se\)](#), från 2012
- [Nationella prov årskurs 9: resultat \(scb.se\)](#), från 2003
- [Grundskolan: betyg årskurs 6 \(scb.se\)](#), från 2013
- [Grundskolan: slutbetyg \(scb.se\)](#), från 1987/88

### Specialskolan

- [Specialskolan: elever per 15 oktober \(scb.se\)](#)
- Elever, från 2019
- Betyg åk 10, från 2005/06

---

<sup>109</sup> Bilaga 1 är framtagen av Paula Kossack, Analytiker  
Avdelning för social statistik och analys, Arbetsmarknad och utbildning på  
SCB, statistikmyndigheten.

## Fritidshem

- [Fritidshem: Elever och personal per 15 oktober \(scb.se\)](#)
- Grundskola, från 2008
- Anpassad grundskola, från 2019
- Specialskola, från 2019

## Gymnasieskolan

- [Gymnasieskolan: Sökande och antagna \(scb.se\)](#)
- Sökande och Antagna, 1985–1993
- Sökande, från 1994
- Antagna, 1994–2016
- Bakgrund elev, 1994–2016
- Antagna och Bakgrund - elev, från 2017
- [Gymnasieskolan: elever per 15 oktober \(scb.se\)](#), från 1992
- [Nationella prov gymnasieskolan: resultat \(scb.se\)](#), 2011–2019
- [Gymnasieskolans betygsuppgifter \(scb.se\)](#)
- Avgångna från Waldorf och IB, 1997–2004
- Avgångna, sent inkomna uppgifter, 1997–2002
- Samlat betygsdokument, 2011–2004
- Slutbetyg, från 1972/73
- Kursbetyg, från 1984/85
- Examens- och studiebevis T4, , från 2011/12
- Gymnasieintyg, 2011/12–2019/20
- Gymnasieintyg kurser, 2011/12–2019/20
- Grundskolebetyg utfärdat i gymnasieskola, 2011/12–2013/14
- [Etablering på arbetsmarknaden och fortsatta studier efter gymnasieskolan \(scb.se\)](#) Den här statistiken är registerbaserad, dvs. den baseras på bearbetningar av befintliga register på SCB. Den innehåller därför inget eget mikrodataregister.
- Anpassad grund- och gymnasieskola: elever per 15 oktober
- [Anpassad grundskola: elever per 15 oktober \(scb.se\)](#), från 2019
- [Anpassad gymnasieskola: elever per 15 oktober \(scb.se\)](#), från 2019 För tidigare årgångar finns uppgifter på gruppnivå.

## Kommunal vuxenutbildning inkl. svenska för invandrare

- [Komvux: elever, kursdeltagare och utbildningsresultat \(scb.se\)](#), se också översiktssidan [Komvuxregistret \(scb.se\)](#)
- Elever, från 1996
- Elever med slutbetyg/gymnasieexamen, från 2011
- Kursdeltagare, från 1979
- Kursdeltagare med slutbetyg/gymnasieexamen, från 2011
- Nationella prov, från 2013
- [Utbildning i svenska för invandrare: elever, kursdeltagare och utbildningsresultat \(scb.se\)](#)
- Kursdeltagare och resultat, från 1993/94

- Nationella prov: resultat, från 2009

## Anpassad utbildning för vuxna

- [Komvux som särskild utbildning: elever per 15 oktober \(scb.se\)](#), från 2019.  
För tidigare årgångar finns uppgifter på gruppnivå.
- Statens skolor för vuxna (SSV), nedlagda
- Elever 1997/98–2001/02
- Kursdeltagare 1997–2002

## Lärrarregistret

- [Pedagogisk personal \(lärrarregistret\) \(scb.se\)](#), se även översiktssidan [Skolverkets register över pedagogisk personal \(Lärrarregistret\) \(scb.se\)](#)
- Personregistret, från 1995/96
- Tjänsteregistret, från 1978/79
- Svenska utlandsskolor. Uppgifterna är antingen på gruppnivå eller saknar personnummer
- Distansundervisning, elever
- Grundskolan, elever
- Gymnasieskolan, elever
- Kompletterande undervisning, elever
- Lärare
- Skolledning och undervisningsvolym
- Skolor

## Folkhögskolan

- [Folkhögskolestatistik \(scb.se\)](#)
- Deltagarregister från 2005
- Deltagare i sfi sedan 2018
- Pedagogisk personal från 2009

## Studieförbundsstatistik

- [Statistikdatabasen - Välj tabell \(scb.se\)](#)
- Deltagare i studieförbundens verksamhet, från 2012
- Studieförbund, 2012–2019
- Studieförbund, från 2020
- Det finns också äldre data från 2002 som baseras på gruppuppgifter.

## Eftergymnasial yrkesutbildning

- [Yrkeshögskolan \(scb.se\)](#), se även översiktssidan [Mikrodata för yrkeshögskolans sökande, studerande och examinerade \(scb.se\)](#)
- Sökande, från 2014

- Studerande och examinerade, från 2005
- Kvalificerad yrkesutbildning (föregångaren till yrkeshögskolan), 1996–2002
- [Konst- och kulturutbildningar och vissa andra utbildningar \(scb.se\)](#), från 2015
- Kompletterande utbildningar (föregångaren till Konst- och kulturutbildningar)
- Deltagare som fullföljt utbildning, 2001–2011
- Deltagare, 2011–2011
- Deltagare, 2012–2015

## Högskolan

Det är Universitetskanslersämbetet (UKÄ) som ansvarar för den officiella statistiken. Mikrodata beställs från SCB.

Här finns alla statistikprodukter för högskolan: [Utbildning och forskning \(scb.se\)](#)

Här finns översiktlig information om de olika registren:

- [Universitets- och högskoleregistret – sökande och antagna till högskoleutbildning på grundnivå och avancerad nivå \(scb.se\)](#)
- [Universitets- och högskoleregistret – studenter och examinerade i högskoleutbildning på grundnivå och avancerad nivå \(scb.se\)](#)
- [Universitets- och högskoleregistret – doktorander och examina i högskoleutbildningen på forskarnivå \(scb.se\)](#)
- [Universitets- och högskoleregistret – personal vid universitet och högskolor \(scb.se\)](#), se [Personal vid universitet och högskolor \(scb.se\)](#) för länk till mikrodata

## Befolkningens utbildning och studiedeltagande

Det är SCB som ansvarar för den officiella statistiken.

- [Befolkningens utbildning \(scb.se\)](#), se även översiktssidan [Mikrodata för Registret över befolkningens utbildning \(scb.se\)](#)
- Gymnasieintyg, ackumulerat register samt summerade poäng, från 2012
- Komvux, från 1988
- Personens samtliga utbildningar, från 1997
- Samlat betygsdokument/studiebevis, ackumulerat register, från 2001
- Personens högsta utbildning, från 2008
- Personens högsta utbildning, 16–74 år, 1997–2007
- Ut- och invandrades utbildning, från 1990
- [Befolkningens studiedeltagande \(scb.se\)](#)
- Studiedeltagande, från 1995
- Studiemedel, från 2001
- [Den internationella undersökningen av vuxnas färdigheter \(scb.se\)](#), 2012
- [Undersökningen om vuxnas deltagande i utbildning \(scb.se\)](#), 2005/06, 2011/12, 2016



- [Personalutbildning i företagen \(scb.se\)](https://scb.se) och [Personalutbildningsstatistik \(scb.se\)](https://scb.se), olika frekvens 1995–2010,

## Övergångar mellan utbildningsformer och till arbetsmarknaden

Undersökningar som SCB ansvarar för:

- [Gymnasieungdomars studieintresse \(scb.se\)](https://scb.se), intermittenta enkätundersökningar sedan 2003/04
- [Inträdet på arbetsmarknaden \(scb.se\)](https://scb.se) intermittenta enkätundersökningar sedan 1997/97
- Examinerade från yrkeshögskolan 2019, 2022
- Gymnasieavgångna/högskoleexaminerade 2000, 2002, 2004
- Högskoleexaminerade 2006, 2008, 2010
- Högskolestuderande 2013, 2017, 2021
- Gymnasieavgångna, intermittent 2006–2023
- [Sysselsättning och inkomster efter avslutad utbildning \(scb.se\)](https://scb.se) Den här statistiken är registerbaserad, dvs. den baseras på bearbetningar av befintliga register på SCB. Den innehåller därför inget eget mikrodataregister.
- [Övergång gymnasieskola-eftergymnasiala utbildningar \(scb.se\)](https://scb.se) Den här statistiken är registerbaserad, dvs. den baseras på bearbetningar av befintliga register på SCB. Den innehåller därför inget eget mikrodataregister.

## Studiestöd

Det är Centrala studiestödsnämnden CSN som ansvarar för den officiella statistiken. Mikrodata om studiestöd finns hos SCB (se produkten Befolkningens studiedeltagande), men mer data kan finnas hos CSN.

Mikrodata om återbetalning finns hos CSN. [Officiell statistik om studiestödet - CSN](#)

## Övriga undersökningar med mikrodata som SCB ansvarar för

- [Högutbildade utrikes födda \(scb.se\)](https://scb.se) Har varit enkätundersökningar (2004, 2009, 2015, 2018). Den senaste undersökningen 2023 är en registerbearbetning.
- [Lärare utanför yrket \(scb.se\)](https://scb.se) Enkätundersökningar 2016 och 2022/23. Ska kompletteras med registerbaserad statistik.
- Här finns flera engångsundersökningar: [Analyser och statistik om befolkningens utbildning \(scb.se\)](https://scb.se) De analyser som baseras på enkäter har tillhörande mikrodataregister.
- [Elevpaneler för longitudinella studier \(scb.se\)](https://scb.se) 2022–2020

## Bilaga 2 - Översikt över tillgänglig statistik

Myndighet och motsvarande	Skolform	Statistikområde
Skolverket	Pedagogisk omsorg	<a href="#">Barn och grupper, personal och kostnader</a>
Skolverket	Öppen förskola	<a href="#">Verksamheter, personal och kostnader</a>
Skolverket	Förskola	<a href="#">Barn och grupper</a>
		<a href="#">Kostnader</a>
		<a href="#">Personal</a>
Skolverket	Omsorg på obekväma tider	<a href="#">Barn och grupper</a>
Skolverket	Förskoleklass	<a href="#">Elever</a>
		<a href="#">Kostnader</a>
		<a href="#">Personal</a>
Skolverket	Öppen fritidsverksamhet	<a href="#">Verksamheter, personal och kostnader</a>
Skolverket	Fritidshem	<a href="#">Elever och grupper</a>
		<a href="#">Kostnader</a>
		<a href="#">Personal</a>
Skolverket	Grundskola	<a href="#">Betyg årskurs 6</a>
		<a href="#">Betyg årskurs 9</a>
Skolverket		<a href="#">Kostnader</a>
Skolverket		<a href="#">Nationella prov</a>
Skolverket		<a href="#">Personal</a>
Skolverket		<a href="#">Skolor och elever</a>

<b>Myndighet och motsvarande</b>	<b>Skolform</b>	<b>Statistikområde</b>
Skolverket	Grundsärskola	<a href="#">Elever</a>
		<a href="#">Kostnader</a>
		<a href="#">Personal</a>
Skolverket	Specialskola	<a href="#">Betyg</a>
		<a href="#">Elever</a>
		<a href="#">Kostnader</a>
Skolverket	Gymnasieskola	<a href="#">Personal</a>
		<a href="#">Betyg och studieresultat</a>
		<a href="#">Efter gymnasieskolan</a>
Skolverket	Gymnasiesärskola	<a href="#">Kommunernas aktivitetsansvar</a>
		<a href="#">Kostnader</a>
		<a href="#">Nationella prov</a>
Skolverket	Gymnasiesärskola	<a href="#">Personal</a>
		<a href="#">Skolor och elever</a>
		<a href="#">Elever</a>
Skolverket	Komvux	<a href="#">Kostnader</a>
		<a href="#">Personal</a>
		<a href="#">Betyg och studieresultat</a>
Skolverket	Komvux	<a href="#">Elever och kursdeltagare</a>
		<a href="#">Etableringsstatus</a>
		<a href="#">Kostnader</a>
Skolverket	Komvux	<a href="#">Personal</a>
		<a href="#">Kostnader</a>

<b>Myndighet och motsvarande</b>	<b>Skolform</b>	<b>Statistikområde</b>	
Skolverket	Svensk utbildning i utlandet	<a href="#">Betyg</a>	
		<a href="#">Elever</a>	
		<a href="#">Kostnader</a>	
		<a href="#">Personal</a>	
		Skolinspektionen	<a href="#">Skolenkäten</a>
			<a href="#">Statistik om regelbunden (planerad) tillsyn</a>
			<a href="#">Statistik om individärenden</a>
			<a href="#">Statistik om regelbunden kvalitetsgranskning</a>
			<a href="#">Statistik över riktad tillsyn</a>
			<a href="#">Statistik tillstånd fristående skolor</a>
<a href="#">Statistik över viten</a>			
<a href="#">Statistik om jämställdhet</a>			
<a href="#">Etableringskontroller av fristående skolor</a>			
<a href="#">Ombedömning av nationella prov</a>			
<a href="#">Anmälningar till Lärarnas ansvarsnämnd</a>			
SCB	Folkhögskola	<a href="#">Unika deltagare</a>	
Folkbildningsrådet	Folkhögskola	<a href="#">Regionindelad folkhögskolestatistik</a>	
Linköpings universitet	Folkhögskola	<a href="#">Folkhögskoledatabasen (nationella kurskataloger)</a>	
SCB	Yrkeshögskola	<a href="#">Sökande, studerande och examinerade</a>	

Myndighet och motsvarande	Skolform	Statistikområde
Myndigheten för yrkeshögskolan	Yrkeshögskola	<a href="#">Program (YH-utbildn. från 100 poäng)</a>
		<a href="#">Kurser (YH-utbildn. upp till 99 poäng)</a>
		<a href="#">Behörighetsgivande förutbildning</a>
		<a href="#">Uppdragsutbildningar</a>
		<a href="#">Kurs på tom plats</a>
		<a href="#">Konst- och kulturutbildningar</a>
		<a href="#">Tolkutbildningar inom folkbildningen</a>
		<a href="#">Utbildningar med endast tillsyn</a>
Universitets- kanslersämbetet	Högskola	<a href="#">Utbildning på grund- och avancerad nivå</a>
		<a href="#">Utbildning på forskarnivå</a>
		<a href="#">Ekonomi</a>
		<a href="#">Personal</a>
		<a href="#">Forskning vid universitet och högskolor</a>
SCB	Högskola	<a href="#">Registret över slutbetyg från gymnasieskolan</a>
		<a href="#">Universitets- och högskoleregistret – sökande och antagna till högskoleutbildning på grundnivå och avancerad nivå</a>
SCB		<a href="#">Utbildningsregistret</a>

<b>Myndighet och motsvarande</b>	<b>Skolform</b>	<b>Statistikområde</b>
Göteborgs universitet		<a href="#">Utvärdering genom uppföljning - UGU</a>
		<a href="#">Gothenburg Educational Longitudinal Database – GOLD</a>
		<a href="#">Center for Comparative Analysis of Educational Achievement, COMPEAT</a>
Myndigheten för ungdoms- och civilsamhällesfrågor	Grundskola, gymnasium	<a href="#">Skolvalet</a>
		<a href="#">Ung idag</a>
Folkhälsomyndigheten	Grundskola, gymnasium	<a href="#">Skolbarns hälsovanor, nationella och regionala resultat</a>
		<a href="#">Skolbarns hälsovanor</a>
Ungdomsbarometern		<a href="#">Rapporter och undersökningar</a>
Internetstiftelsen		<a href="#">Svenskarna och internet</a>
Umeå universitet	Högskola	<a href="#">Högskoleprovet</a>
ILSA		<a href="#">ILSA Gateway till internationella studier</a>
Tritonia	Citeringsdatabaser	<a href="#">Utvärdering av vetenskapliga publikationer</a>