

Utmaningar för tillgång och utnyttjande av data

Utbildningsdata kan användas för att utveckla såväl undervisning och läromedel som ledning, program och miljöer för utbildning. Rapporten från AP2¹ visar flera exempel på nyttan av dataanalys för utbildningsändamål.

I detta arbetspaket har vi hållit arbetsmöten för att samla in, sammanställa och kritiskt värdera behov av data och dataanalys för att få fram beskrivningar av datamängder som går använda operativt för att utveckla utbildningar. Vi har även tagit fram kravspecifikationer i fråga om nödvändiga datamängder och analysmetoder samt frågor runt integritetsskydd och datasäkerhet samt villkor för företag och organisationer att dela med sig av data. Här följer en kortfattad rapport av det vi kommit fram till.

Det finns två olika perspektiv på tillgång till data. Det ena kan vi kalla dataförberedelse² och den andra för tillitshantering.

Dataförberedelse

Begreppet teknikmognad (eng. technology readiness³) avser en given tekniks utveckling från grundläggande egenskaper visade genom forskning på nivå 1 och vidare upp till nivå 9 där tekniken prövats operativt. På liknande sätt behöver vi resonera om data. Data kan vara tillgängliga på en låg nivå exempelvis genom att det är styrkt att de finns men är ännu inte insamlade och data kan finnas tillgängliga i standardiserad form med kunskap om begränsningar utifrån en given uppgift.

Ett konkret förslag till förberedelsenivå för dataanalys kommer från Neil Lawrence vid Amazon Research Cambridge och University of Sheffield⁴. Han grupperar nivåerna i band från A till C med numrerade undernivåer; band C är den mest grundläggande nivån. Detta är en kort summering⁵.

Band C innebär att data görs tillgängliga i standardiserad form beaktat legala, etiska och integritetsmässiga begränsningar. Frågor som måste besvaras för att uppfylla kraven i band C innefattar:

- Finns data, var finns de och går det att få tillgång till dem?
- Vilket format har de (MS Word, PDF eller inskannad handskrift för text; representation för numeriska digitala dataformat)?
- Vilka etiska aspekter samt legalitets- och integritetsaspekter finns och går de uppfylla för datatillgången?
- Vilka tekniska svårigheter finns för tillgången till data? Är de spridda över flera system?

När vi hör talas om data då befinner vi oss på nivå C4 och när man uppnått förberedelsenivå C1 så är datamängden redo att köras i analysprogramvara, eller erbjudas till andra⁶.

Band B av förberedelsenivåer gäller att bedöma den trovärdighet man kan ha till datamängden om vad den faktiskt representerar. Det innebär att data som är redo på nivå C1 ska nu bedömas utifrån begränsningar och felaktigheter. Det kan gälla felkällor vid rapportering och insamling, saknade

¹ Olga Viberg, *How to implement data-driven analysis of education: opportunities and challenges*

² <http://data-readiness.org/>

³ https://en.wikipedia.org/wiki/Technology_readiness_level

⁴ Neil D. Lawrence, Data Readiness Levels, <https://arxiv.org/abs/1705.02245> (tillgänglig 2019-05-13)

⁵ <http://data-readiness.org/>

⁶ Se exempel på datamängder på [OpenML](#)

datavärden och förväxlingar (ex. måttenheter, rad och kolumn i tabeller) och avvikelser (såsom tidstämplar med klockavvikelser eller fel inställt datum). När data är förberedda till nivå B1 så kan en expert veta hur väl datamängden svarar mot den beskrivning som givits för den och man kan resonera om användningen av datamängden.

Band A gäller sedan data i ett sammanhang. Hur väl ämnad är datamängden för att besvara en given frågeställning med en vald analysmetod? Det går enbart att uttala sig om förberedelsenivåerna i band A för en given frågeställning. En datamängd kan därför vara användbar på nivå A1 för en frågeställning och samtidigt inte vara klar för att besvara en annan frågeställning, vilket innebär nivå B1.

Kompetenskrav för att förbereda data för en given frågeställning: För band C krävs administrativt och eventuellt juridiskt stöd för att få tillgång till data om det finns utanför organisationen samt för att hantera integritetsaspekter. För att hantera och sammanställa datamängder rent tekniskt krävs god programmeringsfärdighet, systemkunnande (speciellt vid stora datamängder och för strömmande data) och kunskap om datasäkerhet. För band B krävs god kunskap i statistik, dataanalys och visualisering av data. Band A utförs av den som ska göra analysen och som då måste behärska analysverktygen, exempelvis olika former av maskininlärning. Lawrence pekar att det är resurskrävande att förbereda data för analys. Det band som han anser kräver mest är band B då trovärdigheten och lämpligheten av data ska säkerställas. Även arbetet på band C underskattas och undervärderas ofta i projekt.

Dataförberedelse inom utbildningsdata

Vi använder nu dessa tre breda band av förberedelsenivåer för att bedöma några exempel med utbildningsdata.

Frånvarodata i gymnasiet: Frågeställningen är om det går förutse studenter som riskerar att avbryta gymnasiestudierna eller avsluta dem utan fullständiga betyg. Vid vårt besök på regionsförbundet GR fick vi en genomgång av projektet PlugInnovation och efterföljande projekt, PlugIn och Plug Innan (för åk 5 till 9). Insamling av data görs av de lokala projektledarna på varje skola för varje deltagande elev och rapporteras i Excel-ark. De anser att det är mycket information för en lokal projektledare att fylla i och på grund av den administrativa bördan har information endast samlats in för en begränsad andel av deltagarna. Det finns även brister i underlaget till följd av bristande system och rutiner:

- Vad är giltig och ogiltig frånvaro? Individuella studieplaner påverkar exempelvis detta.
- Avseende antal timmars insats: Är det inrapportering av genomsnittligt per vecka eller per termin.
- Vad är kvaliteten på insatsen?

Det som framkom är att datainsamling och bearbetning som underlag för ett löpande arbete med att förhindra avhopp ännu inte nått upp till den förberedelsenivå som behövs. Data finns och kan sammanställas manuellt, men format och definition är inte entydigt bestämda. Så förberedelsenivån är någonstans mellan C2 och C3. För att vara en del av skolors vanliga, löpande uppföljningsarbete av studiegången och grund för beslut om kontakter och insatser bör data vara tillgängliga och förberedda för utvalda analysverktyg på nivå A1: data och analys är klara att använda.

Antagningsdata för högskolan: Studenter antas från svenska gymnasieprogram till högskolan i tre grupper: en tredjedel antas på meritvärde (summan av meritpoäng och medelbetyg), en tredjedel genom resultat från högskoleprovet (förutsatt att de är behöriga) och en tredjedel kan antas i lokal antagning. Lokal antagning är ovanlig och för de flesta utbildningar antas den kvoten på meritvärde. KTH och Chalmers använder dock ett matematik-fysikprov som urvalsgrund för lokal antagning. Syftet med studien var att följa hur väl antagningsgrunden förutsäger framgång med studierna. För

detta behövs individuella data med meritvärde och resultat från högskoleprovet alternativt matematik-fysikprovet, om något av dem var antagningsgrund.

Planen var att undersöka studenter på civilingenjörsprogrammet i elektroteknik. En variabel i studien är definitionen av framgång i studierna och resultat från studien skulle behöva undersöka känsligheten i den variabeln. En enkel definition skulle vara att minst N antal högskolepoäng skulle vara avklarade efter tre års studier. Studien har inte genomförts av följande skäl. Den databas, Beda, som innehåller alla antagningsdata finns hos Universitets- och högskolerådet (UHR) och vi har inte fått tillgång till den. Gymnasiebetygen behövs för att vissa ämnesbetyg kan vara indikativa för hur väl det går i studierna, särskilt betygen i matematik. Medelbetyget döljer därmed viktig information. En kvalificering av antagningen kunde vara ett resultat av studien, exempelvis att matematikämnet är en bättre indikator för framgång än meritvärdet eller poäng från högskoleprovet.

Vidare är studenterna inte skyldiga att anmäla studieavbrott och uppfattningen hos vår programansvarige är att det inte görs i allmänhet. Det innebär att ytterligare data behövs för att bedöma om en student varit aktiv, exempelvis blivit registrerad på en kurs eller gjort examinationer för befintliga ej avklarade kurser. Detta innebär också att data måste justeras för eventuell deltidsstudietakt. Alla betygsdata finns i LADOK som lyder under offentlighetsprincipen. Det innebär inte att data lämnas ut digitalt för analys. Nu var studien tänkt att ske på KTH så det är möjligt att data varit tillgängliga.

Vi konstaterar att denna datadrivna studie inte gått att genomföra. Data finns och är kompletta för samtliga studenter och formatet är känt. Men eftersom villkoren för åtkomst inte är klara så bedömer vi att förberedelsenivån är på C1-nivån för LADOK-data och på C2-nivån för data från UHRs Beda eftersom villkoren för tillgång inte angivits. Det fortsatta arbetet på band B skulle gälla osäkerhet i studieaktivitet för varje individ och bedömning av studenter som blivit antagna men som lämnat under första perioden (kom de in på annan utbildning som de hellre ville gå, eller insåg de att de inte skulle klara programmet?). Förberedelsenivå i band A och B skulle skett i projektet.

Textdata för examensarbeten: RISE SICS har gjort en studie av examensarbeten med hjälp av naturlig språkbehandling: 59 461 rapporter har laddats ned för analys från Digitalt vetenskapligt arkiv, DiVA, för öppen publicering⁷. Rapporterna publiceras i PDF-format med meta-data. Detta är den datamängd som vi i projektet har kommit åt för en studie. Det innebär dock inte att DiVA erbjuder data för textanalys på en B1-nivå! Meta-data har stora brister och luckor. Till exempel går det inte entydigt att koppla en rapport till ett utbildningsprogram och det går inte alltid bedöma om arbetet gjordes hos en extern part eller på högskolan. Vidare är rensningen av PDF-formatet inte tillräckligt bra för att veta om formateringskod från PDF påverkar analysen av den faktiska texten. Detta innebär att rapporter för examensarbeten som finns tillgänglig i DiVA inte är redo för vidare förberedelse i band A av detta ramverk, utan förberedelsenivån ligger i band C. DiVA används inte av alla svenska högskolor så datamängden räcker inte för en analys om nationella förhållanden; då behövs ytterligare insamling från lärosätenas respektive databaser vilket också är förberedelse inom band C med eventuell hantering i band B av data som inte finns tillgängliga alls.

Sammanfattning av dataförberedelse av utbildningsdata

Vi konstaterar att det saknas systematiskt arbete med att definiera och samla in data för analys. Den troliga orsaken är att dataanalys inte har varit ett krav när insamlingen har påbörjats och att möjligheterna och nyttan som finns med dataanalys inte är tillräckligt kända för att utveckla arbetet

⁷ Fredrik Olsson och Magnus Sahlgren, *Analys av uppsatser i DiVA: En fallstudie kring öppna data och utfallet av högre utbildning i Sverige*, arbetspaket 5.

med förberedelse av data. Exempelvis för DiVA hade varje rapport kunnat finnas i ren oformaterad text för analys och meta-data hade kunnat fyllas i automatiskt för examensarbeten genom koppling till LADOK för studenten. Vi ser därför ett stort behov av fortsatt arbete med att samla in och hantera data för analys inom utbildningssektorn.

Nedan beskriver vi tre verksamhetsformer med tillhörande krav på dataförberedelse för att utbildningssektorn ska kunna bli datadriven.

Tillitshantering

Sedan fredagen den 25 maj 2018 gäller den europeiska dataskyddsförordningen som avser skydda oss från att bli datamässigt utnyttjade. Utöver det allmänna skyddet av personuppgifter⁸ definieras vissa uppgifter som *känsliga* personuppgifter med huvudregeln att det är förbjudet att behandla dem. Detta anges i artikel 9:1:

Behandling av personuppgifter som avslöjar ras eller etniskt ursprung, politiska åsikter, religiös eller filosofisk övertygelse eller medlemskap i fackförening och behandling av genetiska uppgifter, biometriska uppgifter för att entydigt identifiera en fysisk person, uppgifter om hälsa eller uppgifter om en fysisk persons sexualliv eller sexuella läggning ska vara förbjuden.

Utbildningsdata innefattar intellektuella data vilka visar förmågor för inläring och tillämpning av kunskap för att producera lösningar, skrifter och konstverk; de visar personens arbetsförmåga, kommunikationsfärdigheter, stresstålighet och samarbetsförmåga. De är data som kan påverka en persons livschanser när de används vid beslut om anställning, befordringar, antagning till utbildningar och kanske även vid val av livspartner och socialt umgänge, om de blir synliga. Trots detta anses inte intellektuella personuppgifter vara känsliga av förordningen⁹.

Vi har synen att data om en persons intellektuella kapacitet är känsliga och måste hanteras som sådana. Det finns dock undantag från förbudet att behandla känsliga persondata. De anges i artikel 9.2; bland andra skäl kan personer samtycka till databehandling.

Vidare finns det andra former av utbildningsdata som inte är personuppgifter såsom storskaligt aggregerade data och data som inte har med personer att göra, exempelvis matsedlar från skolbespisningar och utlåningsdata för boktitlar från skolbibliotek. Vidare finns begreppet pseudonymiserade data¹⁰ (begreppet anonymisering används inte). Dessa omfattas inte av förordningen. De viktigaste data för att förbättra utbildningar och undervisning är dock persondata så en verksamhet för datadriven utveckling av undervisningssektorn måste hantera de mest känsliga data vi kan samla in om personer. Det innebär en hög grad av systemkompetens för att skydda åtkomsten till data samt rutiner för insamling och hantering av data som garanterar att förordningen uppfylls.

Det finns även behov att skydda identiteten på uppgiftslämnaren, exempelvis den utbildningsanordnare som delar med sig av data. Detta särskilt tydligt inom kommunal vuxenutbildning där utbildningsföretag konkurrerar om att få bedriva utbildning. Det finns därför

⁸ Personuppgifter är all slags information som kan knytas till en levande person (källa Datainspektionen).

⁹ Gunnar Karlsson, [Sverige bör satsa på globala kunskapsprov](#), Svenska Dagbladet, 7 juni 2018 (tillgänglig 2019-07-04)

¹⁰ Dataskyddsförordningen Art. 4 §5. *pseudonymisering*: behandling av personuppgifter på ett sätt som innebär att personuppgifterna inte längre kan tillskrivas en specifik registrerad utan att kompletterande uppgifter används, under förutsättning att dessa kompletterande uppgifter förvaras separat och är föremål för tekniska och organisatoriska åtgärder som säkerställer att personuppgifterna inte tillskrivs en identifierad eller identifierbar fysisk person.

farhågor att data från en aktör kan användas mot den vid exempelvis upphandlingar. Kombinationen av risken med att dela data och en oklar nytta leder till att utbildningsföretag och andra huvudmän inte är tveksamma till att delta i datadrivet utvecklingsarbete utanför den egna organisationen (och man saknar kanske resurser och kunnande för att göra det inom organisationen, eller genom ombud).

Inom ramen för Industrie 4.0 i Tyskland utvecklas system, referensmodeller och begrepp för datautbyte mellan företag, kunder, myndigheter och andra parter. Ett intressant och viktigt begrepp är *digital suveränitet*, det vill säga möjligheten att ha kontroll över data man äger även när man delar dem med någon. Hela arkitekturen är omfattande och kan inte belysas i denna rapport¹¹. Det är dock uppenbart att utbildningssektorns utmaningar med att dela, använda och hantera data inte är skilda från andra sektors behov. Därför kan vi utgå ifrån att de lösningar och begränsningar som etableras inom denna *Industrial Data Space* kan utnyttjas av utbildningssektorn: i detta avseende är den bara en tjänsteindustri bland många andra.

Av detta går det sluta sig till att insamling, förberedelse och tillhandahållande av data för analys förutsätter flera specialistkompetenser och det krävs system som klarar industriella krav på datahantering för att uppfylla dataskyddsförordningen och kunna ge nytta för utbildningssektorn. I nästa avsnitt beskrivs olika verksamhetsformer som har behandlats i projektet med avseende på dessa krav.

Verksamhetsformer för dataanvändning för utbildning

Som vi beskrivit ovan består data för utbildningssektorn av känsliga persondata för de som utbildas samt för deras instruktörer, lärare och skolläda. Vidare finns data som inte går förknippa med individer men som kan vara konkurrenspåverkande mellan utbildningsaktörer och därmed känsliga. Det finns även data som reservationslöst går dela helt öppet.

I projektet har vi gått igenom tre olika förslag till verksamhet för att samla, förbereda och tillhandahålla data för analys eller tillhandahålla resultat från analyser. De presenteras kortfattat nedan.

Datalabb

Datalabb är ett begrepp som Vinnova använder för en organisation som ger tillgång till data och hjälper aktörer att använda data för innovation. Ett exempel är Trafiklab¹² som öppet erbjuder data från kollektivtrafiken i Sveriges regioner genom väldefinierade API:er.

Ett datalabb för utbildning måste skydda persondata och tillhandahålla alla data med förberedelsenivå B1, redo för analys. Som vi konstaterat är det oklart i vilken mån data finns för olika ändamål och det är osäkert i vilken utsträckning och under vilka villkor som dataägare kan tänkas lämna över data. För persondata gäller att ett datalabb fungerar som personuppgiftsbiträde¹³ åt den som lämnar data. Uppgiften för biträdet kan vara att sammanställa eller pseudonymisera persondata för vidare användning av uppgiftslämnaren eller externa parter. Uppgiftslämnaren kan själv pseudonymisera eller sammanställa persondata så att dataskyddsförordningen inte längre är tillämplig. Då är frågan om det går att säkerställa att data är förberedda enligt de krav som ställs i

¹¹ International Data Spaces Association, Reference Architecture Model 3.0
<https://www.internationaldataspaces.org/ressource-hub/publications-ids/> (tillgänglig 2019-05-13).

¹² <https://www.trafiklab.se/om-trafiklab>

¹³ <https://www.datainspektionen.se/lagar--regler/dataskyddsförordningen/personuppgiftsansvariga-och-personuppgiftsbitraden/>

band B: Hur vet datalabbet exempelvis att uppgiftslämnaren har hanterat data korrekt? Och om vem tar ansvaret för att det inte går att urskilja personer i uppgifterna?

Fördelen med verksamhetsformen datalabb är att tredje part får tillgång till data för tjänster och data-drivna system för utbildningssektorn. Det möjliggör en hög grad av innovation eftersom den som utvecklar för sektorn inte behöver ha direkt tillgång till data på egen hand, något som i många fall inte är möjligt. **Nackdelen** är att datalabbet enbart kan lämna ut pseudonymiserade persondata (om det ens går säkerställa att de inte går koppla till individer) samt data som inte går att hänföra till en individ såsom aggregerade persondata och data utan koppling till personer. Följden är att det är oklart vilken nytta som dessa data kan ge och vidare hur en verksamhet kan motiveras och i förlängningen finansieras om det inte går visa på en tydlig nytta utifrån en tredjeparts affärsmodell (se rapport för AP1 för problem med tillgång till data).

Föresatsen att planera för ett datalabb för utbildningssektorn har inte gått att genomföra därför att nyttan inte gått påvisa, dataägare har inte visat öppenhet för att lämna ut data samt huvudmannaskapet och finansieringen inte gått att identifiera. Vi har inte funnit någon datamängd för en användning som kunde tjäna som labbets "killer application" och som därmed hade kunna vara grunden för att etablera ett datalabb.

Forskningsdata

Forskningen kan hantera persondata och vetenskapliga frågeställningar är ofta en god motivation för att få tillgång till data från de parter som äger dem. Vi har därför beaktat möjligheten att stödja utbildningssektorns datadrivna utvecklingsarbete i form av en forskningsdatabas. I slutbetänkandet *Rätt att forska: Långsiktig reglering av forskningsdatabaser* (SOU 2018:36) förslås en förändrad lagstiftning för att säkerställa att forskningen har tillgång till persondata när så är nödvändigt.

Det finns en stor mängd databaser för forskning av olika karaktär. I nästa stycke beskrivs kvalitetsregister som stödjer forskning såväl som verksamhetsutveckling. Här snävar vi in definitionen till datasamlingar för rena forskningsfrågor som knyts de till högskolor och institut där verksamheten bedrivs. Livscykeln för databasen blir därmed beroende på forskningens längd och finansieringsform.

Fördelar med forskningsdata för utbildning är att i den mån det går att få tillgång till persondata så har organisationen rätt att behandla dem (eventuellt behövs etikprövning). Resultaten publiceras och sprids även på andra sätt och kan sedan komma utbildningssektorn till del. **Nackdelar** är att datasamlingen görs utifrån forskningsfrågeställningar, inte utifrån verksamhetens behov och utvecklingsplaner. Vidare upphör samlingen i samband med att forskningen är avslutad även om den skulle kunna utnyttjas för andra ändamål. Och slutligen kan steget från forskningresultat till förslag som går att förverkliga i en verksamhet vara väldigt tid- och resurskrävande.

Datasamlingar för forskning innebär något annat än att tillhandahålla data för innovation inom utbildningssektorn. Forskningen behöver data men det behovet skiljer sig från verksamhetens behov av stöd för utveckling av nya tjänster, system och utbildningsprocesser.

Kvalitetsregister

Det är av högsta vikt att utbildningssektorn bygger upp evidens för löpande förbättringsarbete¹⁴. För detta har vi tagit intryck av de nationella kvalitetsregister som finns för hälso- och sjukvården¹⁵.

¹⁴ Läs gärna Nils Karlsson, [Evidens – Socialstyrelsens definition och alla andras](#), Dagens samhälle 2019-01-04 (tillgänglig 2019-07-04)

¹⁵ <http://www.kvalitetsregister.se>

Organisationen *Nationella Kvalitetsregister*¹⁶ 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 förbättringsarbete¹⁷ som forskning¹⁸. Det publiceras rapporter där kvalitetsregister använts i vårdens förbättringsarbete och mera övergripande rapporter¹⁹. 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 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 skolvivåer. Insamling av data kan leda till merarbete hos skolor och utbildningsföretag vilket kan leda till lågt deltagande och dålig täckning.

Sammanfattning

Utbildningssektorn arbetar med digitalisering inom grundskola och gymnasium²⁰, högskola och yrkeshögskola samt inom vuxenutbildning och fort- och vidareutbildning. Data-driven verksamhetsutveckling, det vill säga att graden av måluppfyllnad är mätbar, förutsätter naturligtvis data och som vi konstaterar är tillgången på data otillfredsställande. Ofta fanns data inte insamlade, eller det är oklart vad de representerar och organisationen vet inte hur man ska använda dem men vågar inte heller lämna dem till ett personuppgiftsbiträde. Sällan har man resurser att ens arbeta med frågan.

Om data ska användas inom den egna organisationen behöver den själv eller med hjälp av ett biträde förbereda data till A1 nivå för att analysresultaten ska gå att använda för att förbättra verksamheten. Som visat kräver det många olika expertkunskaper och datasäkrade system. Vårt förslag är därför att data samlas i kvalitetsregister som har rätten och expertisen att hantera personuppgifter. Registrens uppgift är att ge stöd för förbättringar av utbildningar av allehanda slag samt, i andra hand, stödja

¹⁶ Dataskyddsförordningen Art. 4 §6. *register*: en strukturerad samling av personuppgifter som är tillgänglig enligt särskilda kriterier, oavsett om samlingen är centraliserad, decentraliserad eller spridd på grundval av funktionella eller geografiska förhållanden.

¹⁷ [Guide: Kvalitetsregister som stöd i förbättringsarbete](#) (tillgänglig 2019-07-04)

¹⁸ <http://www.kvalitetsregister.se/forskning.1814.html>

¹⁹ Exempelvis [Öppna jämförelser: hälso- och sjukvård 2019](#) (tillgänglig 2019-07-04)

²⁰ <http://skoldigiplan.se/>

forskning om utbildning och tillhandahålla öppna data för innovation. En möjlig början är att låta alla parter inom utbildningssektorn lämna de uppgifter som de enligt lag samlar in och lämnar till Statistiska centralbyrån. Det innebär därmed inget merarbete av någon. Nästa steg är då att ta fram analyser av nytta som går utöver de rapporter som SCB sammanställer.

Utkast

”Kartlägga behov för de viktigaste utmaningarna vad gäller tillgång och utnyttjande av data.”

Genomgång av fallstudier från WP2 för att kunna genomföra dem och bedöma behov av infrastruktur och specialistkompetens. Framtagning och sammanställning av ytterligare, egna förslag på användningar av data för förbättrad undervisning.

Resultat. Sammanställning av användningar och tillhörande system-, data- och analyskrav, behov av infrastrukturer och specialistkompetens. Dokumentation som sprids och används i WP1 för att definiera branschstandarder.

Min plan

- Summera olika användningsområden (från AP2)
- Tillgång till data: dataskyddsförordning, administrativt
- Beskrivning och analys av tre verksamhetsformer:
 - Datalabb – data suveränitet, anonymisering, aggregering
 - Kvalitetsregister
 - Forskningsdatabas
- Summering av möjliga verksamheter med för- och nackdelar samt behov av vidare arbete