

Semantikprojektet

Slutrapport, 31 januari 2023



Skolverket



Vetenskapsrådet

© Arbetsförmedlingen
Författare: Myndighetsrepresentanter inom Semantikprojektet
Datum: 2023-01-31
Diarienummer: Vinnova-2021-04628; Af-2021/00935272

Innehåll

Sammanfattning	4
1 Inledning.....	7
1.2 Frågor som vi sökt svar på i projektet.....	9
2 Identifierade förutsättningar.....	14
2.1 Verksamhetsspår som gett underlag till slutsatser	14
2.2 Teknik och statistik som kan tydliggöra kopplingar mellan existerande begreppsstrukturer	21
2.3 Finess krävs för översättning mellan aspekter på kunnande	26
2.4 På väg mot en värdering av gemensamma begrepp och/eller översättningsnycklar.....	29
3 Vad får det för följder i fortsatt förvaltning?.....	36
3.1 Motivering till projekt som kommer drivas under 2023	37
3.2 Förslag på utveckling som bör få förankring under 2023	39
Bilaga 1 – Begreppslista och bilder	40
Bilaga 2 – Kartläggning med hjälp av case	46
Bilaga 3 – Överblick av SKOS	48
Bilaga 4 – Underlag om Matchningsindikatorspåret	49
Bilaga 4 – Underlag om intervjuer	55
Bilaga 5 – UHR och semantikdelprojektet	56

Sammanfattning

Det är komplicerat.

Det är inte bara komplicerat att uttala det kortfattade syftet med projektet: att förbättra den semantiska interoperabiliteten. Det är också den första reaktionen, hos de flesta som börjar sätta sig in i vad det innebär, att utreda förutsättningarna att ta fram gemensamma begrepp eller översättningsnycklar inom domänen kompetensförsörjning och livslångt lärande.

Det räcker dock inte att stanna där. Gemensamma begrepp eller översättningsnycklar berör centrala delar av möjligheter till verksamhetsutveckling inom domänen, särskilt eftersom verksamhetsutveckling så ofta nuförtiden är mer eller mindre liktydigt med digital utveckling. Eller som digitaliseringsrådet beskriver avsaknaden av semantisk interoperabilitet i sin lägesbild från 2019: ”.. data från olika källor är strukturerade på olika sätt med olika namn, klassificeringar och definitioner vilket gör att uppgifter på ett enkelt sätt inte går att dela eller bearbeta maskinellt och rutinmässigt. Denna avsaknad av standardiserad informationsklassificering och hantering skapar stora inläsningseffekter.”

Det är också i linje med EU:s datastrategi som syftar till ett alltmer datadrivet samhälle. Tanken är att man i större utsträckning ska kunna nyttja data för modellbaserade och maskinella skattningar, inklusive avancerade statistiska beräkningar i neurala nätverk (det som ofta går under de magiska orden AI eller maskininläring). Datainfrastrukturens hävstångseffekter och de datadrivna förmågorna bottenar dock ofta i kvaliteten på de inputdata som används. Vilket sätter fingret på frågor om hur domänexperter behöver involveras för att kvalitetssäkra data och sammanlänkandet av dessa data.

Det fungerar alltså inte heller att famla efter potentiella värden långt borta. Projektets utredning av förutsättningar kan bekräfta att de värden som står på spel för individer, företag och samhälle potentiellt är mycket stora. Kopplingar mellan information om utbildningar och information om arbetsmarknad och näringsverksamhet är värdefulla i många av domänens verksamheter. Samtidigt visar utredningen att de största värdena troligen är förknippade med de största kostnaderna. Den här rapporten tar fasta på och försöker förmedla, vad vi i semantikprojektet sett att den balansen, mellan värden och kostnader, handlar om i praktiken.

De viktigaste lärdomarna som samlats i projektets olika verksamhetsspår finns sammanställda på kommande två sidor. För bättre förståelse av dessa och övriga lärdomar finns dock ingen annan genväg än att läsa även övriga kapitel, som i sig är sammanfattningar av mycket kunskapsutbyte och många utforskande samtal. Inte minst gäller det den begreppslista och de bilder som återfinns först bland bilagorna.

Ett sammandrag av huvudsakliga lärdomar

1. Att tillgängliggöra de olika terminologierna enligt SKOS-standarden underlättar informationsutbyte mellan myndigheter. Det blir lättare att konsumera ett begrepp från en annan myndighet när alla följer samma standard för att beskriva strukturen i sina terminologier.
2. Det finns begränsningar i SKOS som kan vara problematiska för vissa ändamål men det är tillräckligt uttrycksfullt för att representera de väsentliga delarna av de terminologier som kan vara aktuella att publicera på Sveriges dataportal.
3. Vid manuella bedömningar märks stora skillnader i tillvägagångssätt och hur bedömningar görs. Det behövs mer arbete för att prata sig samman för att göra mer likartade bedömningar
4. Slutsatsen från utvärdering genom case är att det är svårt att jämföra läranderesultat som kan härledas ur gymnasieskolans styrdokument med de kompetenskrav som ställs upp av arbetslivet. Det är inte meningsfullt att jämföra dem på så aggregerad nivå som examensmål, men det är också svårt att jämföra dem på ämnes- och kursnivå eftersom kompetenskraven inte är formulerade med samma skärning.
5. I skolväsendets styrdokument återfinns bland annat beskrivningar av syftet med respektive ämne. Där formuleras mål för undervisningen, vilket har vissa likheter med avsedda lärandemål (eng. *intended learning outcomes*). Men skillnaden är att målen siktar in sig på lärarens planering av undervisningen och inte på bedömningen av elevernas kunskaper. Det bidrar till att översättning från information om arbete till information om utbildning bör göras i två steg:
 - a. från/till **yrke** till/från **läranderesultat**. Eller egentligen mer specifikt översättning mellan arbetsuppgifter och det som inom EQF på engelska betecknas *intended learning outcomes*.
 - b. från/till **läranderesultat** till/från **styrdokument** för utbildningar. Där exempelvis en erfaren yrkeslärare kan omvandla skolväsendets innehåll i termer av ämnets syfte, centralt innehåll och bedömningskriterier, till läranderesultat i form av *intended learning outcomes*.
6. Oavsett om översättningsnycklar tas fram med hjälp av manuella redaktioner eller algoritmer för maskininlärning, behövs domänexperter för att bedöma kvalitet på framtagna översättningsnycklar.
7. För att kunna använda gjorda kopplingar på ett värdefullt sätt är det viktigt att beskriva och vara transparent med varför och hur en koppling gjorts. Det gäller oavsett hur väl bedömningen stämmer med verkliga jobbmöjligheter.

8. En rikare beskrivning, med mer information på mer detaljerad nivå, leder till mer underbyggda beslut om kopplingar och färre gissningar, samtidigt som den också ställer högre krav på förvaltning i och med att den totala informationsmängden blir större.
9. Potentiella konsekvenser av att ta fram gemensamma begrepp eller översättningsnycklar varierar mellan olika verksamheter och processer men från gjorda intervjuer med insatta nyckelpersoner framgår att flera ser betydligt större värde av kopplingar på finare detaljnivå än kopplingar mellan existerande begreppsstrukturer.

Det som framträder efter semantikprojektets djupdykning i förutsättningarna att ta fram gemensamma begrepp eller översättningsnycklar är föga förvånande behovet av mer arbete. Ett arbete som systematiskt kan ta ett steg i taget med de olika delar som frågans komplexitet spänner över.

Om första kapitlet beskriver vilka frågor som projektet fokuserat på, och andra kapitlet beskriver arbetet bakom de huvudsakliga lärdomarna ovan, så tar tredje kapitlet fasta på utvecklingen framåt. Det kapitlet beskriver både verksamheter som redan påbörjats och förslag på lämpliga steg som ännu inte fått någon organisatorisk förankring.

De begreppsstrukturer, begrepp och översättningsnycklar som berörs i rapporten är inte de enda data som är relevanta för utveckling av en sammanhållen datainfrastruktur. Förmågan att samverka digitalt, interoperabiliteten, kräver därutöver även annan utveckling. Sammantaget ger lärdomarna från semantikprojektet dock en rik och konkret bild av vilka data som behöver hanteras gemensamt, för att datainfrastrukturen för kompetensförsörjning och livslångt lärande ska vara sammanhållen.

1 Inledning

Jag måste använda ord när jag talar till er!

*Tänk er kaptenen
på en galeas, en hukarejakt,
en skonare, en brigantin,
en fregatt, en korvett, en snau
med molnet av segelduk över sig:
gaffel- och bomsegel,
toppsegel, stagesegel,
skothorn, fallhorn, halshorn.*

*När det blåser till storm
kan han väl inte bara peka
med handen
och ropa: Ta ner dom där!
Det skulle se ut.*

*Snickaren har en låda med fackord
Muraren sina
och den som skall styra en stat
kan väl inte bara gasta i stormen:
Dom där sitter fel! Reva dom där!*

*Jag måste använda ord
när jag talar till er
Ni måste lära er ord.*

- Lars Forsell, ur Oktoberdikter

Lars Forsells dikt förmår på ett typiskt poetiskt sätt fånga många av de utmaningar som vi i projektet tampats med drygt 50 år efter dikten skrevs. För att nämna några:

- Ord och termer är helt centrala objekt för att skapa en funktionell, hållbar och estetiskt tillgänglig informationsarkitektur. Arkitektur som i vårt fall behöver inrymma information kopplat till livshändelser i alla hörn av domänen för kompetensförsörjning och livslångt lärande (hädanefter KLL-domänen).
- Fackspråk förändras med tiden och kan ligga bortom det som är nåbart med resurser som tillsätts för att ta fram gemensamma begreppslistor. Vilket inte hindrar värdet i att utveckla teknik och lösningar som underlättar både arbetet med och tillgängligheten till sådana begreppslistor.
- Värdet av att ta fram översättningar mellan utbildningar och arbetsliv är svår att mäta och varierar mycket från fall till fall. Det är exempelvis i många sammanhang inte värt besväret att lära sig skilja på bomsegel och stagesegel, men det finns livshändelser inom KLL-domänen då det från flera perspektiv kan vara en fråga om liv och död.

- Fackspråk utgör ett, för varje expertområde specifikt, gränsland där verksamheter möter datainfrastrukturer. Datainfrastrukturer som sedan förväntas förmedla så mycket ursprunglig betydelse att informationen blir användbar i andra delar av samhället.
- Det krävs ibland både mänsklig och teknisk finess för att på ett effektivt sätt förmedla information till ett sammanhang som inte känner till den ursprungliga kontext där de lokala dialekterna frodas och gror.
- Det är alltid en fråga om perspektiv och ibland en fråga om maktutövning att bestämma när ett begrepp är synonymt eller inte med ett annat begrepp.

Så långt är det poetiska uttrycket ett skarpare vapen än många utsökt inredda utredningar. Udden försvinner dock snabbt när svar sökes på hur hörnstenar ska läggas till den sammanhållna datainfrastrukturen för kompetensförsörjning och livslångt lärande. Där blir det poetiska koncentratet aldrig mer än en inledning på arbetet med att formulera skarpa frågor och lämpliga avgränsningar.

1.1.1 Kontext och avgränsningar

Semantikprojektet (hädanefter projektet) är en del av regeringsuppdraget (hädanefter uppdraget) att utveckla en sammanhållen datainfrastruktur för kompetensförsörjning och livslångt lärande.

Uppdraget består av flera delar, i den del som beskrivs först står: ”Myndigheterna ska gemensamt utreda förutsättningarna för och påbörja arbetet med att utveckla gemensamma begrepp eller översättningsnycklar mellan existerande begreppsstrukturer inom utbildnings- och arbetsmarknadsområdet,”

Målsättningen med *projektet* är medvetet mer modest formulerat och utlovar inte att arbetet kommer att påbörjas. Allt för att kunna fokusera på det förarbete som krävs för att utreda förutsättningarna. Även om *projektet* inte haft som målsättning att påbörja arbetet har vi dock valt som metod att utreda förutsättningarna genom att testa vad som händer då vi ändå påbörjar ett arbete med att göra gemensamma begrepp eller översättningsnycklar tillgängliga.

Syftet med gemensamma begrepp eller översättningsnycklar är att underlätta utvecklingen av tjänster inom KLL-domänen. I *uppdraget* nämns tjänster för validering, matchning och vägledning. I *projektet* har dessa använts som utgångspunkt snarare än inramning, då det inom domänen har identifierats fler verksamheter där utveckling av digitala stöd skulle underlättas av gemensamma begrepp eller översättningsnycklar.

Denna slutrapport för projektet, lämnas 31 januari 2023 till Vinnova. Arbetet fortsätter dock med att förmedla lärdomar från *projektet* till *uppdragets* fortsättning. Rapporten avslutas följaktligen (i kapitel 3) med beskrivningar av de olika verksamheter som utifrån testverksamhet, identifierade förutsättningar och lärdomar, har fått eller bör få en fortsättning.

1.2 Frågor som vi sökt svar på i projektet

Vilka förutsättningar som behöver utredas var inte givet. Det ingick därmed i projektet att precisera och gruppera vilka frågor som behövde besvaras och/eller utredas vidare. Efter inledande kunskapsutbyte landade projektet, från ingående myndigheters perspektiv, i att dessa var de mest centrala frågorna:

1.2.1 Vilka begrepp och begreppsstrukturer är relevanta?

För att stärka utvecklingen behöver datainfrastrukturen underlätta digital samverkan mellan aktörer i olika delar av KLL-domänen. Det inkluderar den aspekt som av bland annat Digitaliseringsrådet benämnts semantisk interoperabilitet,¹ dvs. förmågan att förstå den information som behandlas digitalt och hur den kan användas.

Förmågan att förstå varandra över sektorsgränser och mellan verksamhetssilos innehåller flera delar där respektive område har ett kunskapsöversläge. Det kräver arbete men är relativt okomplicerat att i de sammanhangen vara transparent med definitioner, förklaringar och exempel som tydliggör vad information man delar med sig av betyder.

Därutöver finns även en del där information från olika områden egentligen syftar på samma sak. Eftersom områdena närmat sig dessa delar från olika perspektiv och med olika tradition finns det ingen konsensus om vilka övergripande begrepp som ska användas för att peka på informationen. Inom KLL-domänen finns ett typexempel på en sådan del med många perspektiv på det som egentligen syftar på samma sak.

Projektet avgränsades tidigt till att i huvudsak fokusera på att utreda förutsättningarna för gemensamma begrepp eller översättningsnycklar inom denna del av domänens informationsmängder. Från utbildningshåll är traditionen att det övergripande begreppet är *kunskap* medan det övergripande begrepp som brukar användas från arbetsmarknadshåll är *kompetens*. I slutbetänkandet från senaste valideringsdelegationen (SOU 2019:69) föreslogs att *kunnande* skulle kunna fungera som ett neutralt övergripande begrepp.² I projektet har vi inte tagit ställning till vilka begrepp som är lämpligast att använda på de övergripande nivåerna utan bara konstaterat att de troligen syftar på samma fenomen i verkligheten. Om inget annat anges ska därför *kunskap*, *kompetens* och *kunnande* tolkas som synonymer då de fortsatt förekommer i rapporten. Även om de alltså samtidigt kan vittna om ett visst perspektiv på frågan. Oavsett vilka övergripande begrepp vi använder så är arbetsgruppens erfarenhet att behovet av semantisk interoperabilitet är störst här.

Det finns inget projektgemensamt svar på frågan om vilka begreppsstrukturer och begrepp som är relevanta. Å ena sidan kan det tyckas enkelt. Det finns ett antal begreppsstrukturer som nämns i uppdraget (eg. SSYK, SUN, SNI och SeQF) och gemensamma begrepp eller översättningsnycklar ska underlätta för tjänster som gör kopplingar mellan dessa. Å andra sidan förekommer, inom KLL-domänen, betydligt fler begrepp än de som ryms i dessa begreppsstrukturer. Begrepp som exempelvis används för att förmedla information om vilka kunskaper man lär sig i en viss

¹ [Begrepp - Digitaliseringsrådet \(digitaliseringsradet.se\)](https://digitaliseringsradet.se)

² Se även tabell i Bilaga 1 med rubriken: Övergripande begrepp – olika perspektiv, samma referenter.

utbildning eller vilka kompetenskraven är i ett yrke. Gemensamma begrepp eller översättningsnycklar kan vara nog så viktiga för att förmedla också den typen av information. I projektets olika verksamhetsspår har vi därför kommit fram till olika sätt att precisera frågan om vilka begrepp och begreppsstrukturer som är relevanta.

1.2.2 Vilka är förutsättningarna att tillgängliggöra relevanta begreppsstrukturer?

I uppdraget framgår att myndigheterna ska verka för att tillgängliggöra, dela och nyttiggöra data i enlighet med Diggs riktlinjer. För ändamålet förvaltar Digg tjänsten Sveriges dataportal som ett samlat sätt att tillgängliggöra myndigheternas öppna data. De i uppdraget nämnda begreppsstrukturerna är typexempel på öppna metadata där Sveriges dataportal kan underlätta vidareanvändning.

Frågan om förutsättningar att tillgängliggöra relevanta begreppsstrukturer preciseras därför i projektet bland annat till frågor om hur nämnda begreppsstrukturer bör tillgängliggöras i Sveriges dataportal.

1.2.3 Vad är gemensamma begrepp eller översättningsnycklar i det sammanhanget?

För att komma åt värdet med att tillgängliggöra data finns det distinktioner som Tim Bernard Lee har formulerat i termer av fem nivåer av öppna data.³ Att tillgängliggöra data via Sveriges dataportal innebär bland annat att enskilda begrepp i noterade begreppsstrukturer får egna adresser, sk. uniform resource identifiers (URI:s), vilket motsvarar nivå 4 av 5, dvs. länkad data, i föreslagen nivåskala.

En av fördelarna med länkade data är att det gör det enklare att sammanlänka redan befintlig information genom att skapa kopplingar mellan länkade data.

Sammanlänkade länkade data är det som i föreslagen nivåskala motsvarar nivå 5. Där länkad data svarar på *var maskiner behöver leta för att hämta data*, svarar sammanlänkade data på *vilka parvisa kopplingar som finns* mellan sådana data, med eller utan beskrivning av dessa kopplingar.

Två länkade data, data med unika adresser, är sammanlänkade då det finns en angiven koppling mellan dem. Beskrivningen av den kopplingen, relationen mellan data, är en datalogisk definition av det som gemensamma begrepp eller översättningsnycklar syftar på. Som framgår av den här rapporten finns många sätt att beskriva relationen mellan data som kommer från olika delar av KLL-domänen.

En viktig dimension för att förstå skillnaden mellan olika relationsbeskrivningar är att tydliggöra vilken detaljnivå som länkade begrepp befinner sig på inom respektive område.

1.2.4 Hur kan relationer mellan strukturerna skapas?

Kopplingar och relationer mellan de i uppdraget angivna begreppsstrukturerna förekommer redan på många håll i samhället, men det finns ingen konsensus om vilka kopplingar som är bra, eller vilka kriterier som ska användas för att avgöra vilka kopplingar som är bra. De flesta skapar sina egna kopplingar utifrån sina behov.

³ [5-star Open Data \(5stardata.info\)](http://5staropen.org/)

Skillnaderna kan i praktiken vara små men utan ett gemensamt forum där skillnader och användningsråd diskuteras kommer kopplingarna inte att bli gemensamma. Som exempel kan nämnas kopplingar mellan SSKYK och SUN. Där det närmaste ett brett samförstånd som skapats är de matchningsbedömningar som gjorts inom ramen för de regionala matchningsindikatorerna.⁴

Projektet har konstaterat att det går att peka ut en aktör som får ansvar att ta fram och beskriva kopplingar mellan olika begreppsstrukturer och/eller olika begrepp. Hur denna aktör ska kunna ta ansvar för kvalitetssäkringen av sådana kopplingar är dock en större fråga, eftersom kvaliteten på kopplingen är helt beroende av vad den ska användas till.

1.2.5 Kan vi beskriva hur existerande begreppsstrukturer hänger ihop?

Ett antal verksamheter inom projektet tog sikte på denna fråga. Arbetet resulterade snabbt i ett konstaterande att de angivna begreppsstrukturerna inte går att kopplas samman på ett entydigt sätt.

Detaljnivån spelar här en stor roll, det är lika viktigt som att kunna beskriva måttenheter för att kunna mäta en sträcka. För att göra en karta över Sverige kan kilometermåtta vara användbart medan inte ens millimeter är tillräckligt för att ange de mått som behövs för att ta fram passande skruvar och muttrar. En stor utmaning är att det i samhället inte finns några motsvarande måttenheter för att formulera hur väl en viss kunskap motsvarar ett visst kompetensbehov. En del av utmaningen är att människors potential och beteende påverkar matchningen mellan uttryckta kunskaper och uttryckta behov. Måttet på matchning kan därmed inte uttryckas i absoluta termer utan uttrycks i bästa fall, om man ska vara noggrann, som en sannolikhet att en individ i framtiden kan utföra en beskriven aktivitet givet det sammanhang som gällde då mätningen gjordes.

Det svåra är att beskriva vilka kvalitetskriterier som behöver vara uppfyllda för att användare ska ha nytta av kopplingen en viss beskriven kunskap och ett visst beskrivet kompetensbehov. Där olika användare har olika behov. Följaktligen finns det stora utmaningar med att peka ut vilka system eller institutioner som behöver skapas för att kunna ta ansvar för kvalitetssäkringen av kopplingar med avseende på olika ändamål.

I projektet har vi sett fyra metoder som används i olika sammanhang för att identifiera om ett begrepp från arbetsmarknadsområdet ska kopplas ihop med ett annat begrepp från utbildningsområdet:

- En möjlighet är att använda historiska registerdata som "facit". Exempelvis beräkna hur stor andel av de individer som har en viss yrkeskod (SSKYK) som har genomgått en viss utbildning, med tillhörande kod (SUN). Större andel betyder i det fallet en starkare koppling mellan två angivna koder.
- En annan möjlighet är att använda individers sökmönster som "facit". Exempelvis hur stor andel av de som har en liknande utbildningsbakgrund

⁴ Se sida 24 i [Regionala matchningsindikatorer Fördjupad analys.pdf \(tillvaxtverket.se\)](#)

som söker efter jobbmöjligheter inom ett visst yrke. Återigen betyder större andel i det fallet en starkare koppling mellan identifierad utbildningsbakgrund och sökt yrke.

- En tredje möjlighet är att använda språkliga likheter som "facit". Exempelvis jämföra hur ett begrepp förekommer i jobbbannonser till ett visst yrke med hur samma ord förekommer i beskrivningar av en viss utbildning. Likheter mellan hur begrepp förekommer i det ena och det andra sammanhanget kan mätas på olika sätt, men oavsett mätmetod så avgör graden av likhet vad stark koppling betyder i det fallet.
- En fjärde möjlighet är att använda experters utlåtande som "facit". Exempelvis hur en yrkeslärare eller en grupp branschrepresentanter beskriver kopplingen mellan ett begrepp som används i en utbildning och ett begrepp som används i ett yrke. Styrkan på kopplingen avgörs i det fallet av hur trovärdigt det är att experten känner till hur begreppen används i de två sammanhang – utbildning och yrke – som beskrivs.

De olika metoderna ger olika möjlighet att använda maskinellt stöd. Hur mycket maskinellt stöd som används påverkar hur stora kostnaderna blir att hitta och beskriva kopplingar mellan begreppen. I slutändan avgör valet av avgränsning och metod vilka resurser som krävs för att beskriva hur existerande begreppsstrukturer hänger ihop.

1.2.6 Vilka är användarna av dessa gemensamma begrepp eller översättningsnycklar?

På motsvarande sätt som metod och avgränsning avgör kostnaden så avgör användare och deras användningsområden värdet av gemensamma begrepp eller översättningsnycklar.

KLL-domänen, de verksamheter som är del av samhällets insatser och stöd till kompetensförsörjning och livslångt lärande, har i uppdraget inte fått någon skarp definition. Projektet har tagit fasta på att möjligheten att räkna antalet verksamheter beror på vilka avgränsningar och benämningar som används. Beroende på att målsättningar och sammanhang skiljer sig åt i olika delar av domänen förekommer både överlappande verksamheter och verksamheter med samma namn men olika innehåll.

För att undvika att börja i någon specifik myndighets perspektiv utredde projektet frågan om användare och användningsområden genom att formulera ett antal case. Case som hade formen av tankeexperiment med utgångspunkt i ett urval av persona, livshändelser och ett gemensamt område av kunnande. I arbetet preciserades frågan successivt och resulterade i en metod att utreda var och hur begrepp och begreppsstrukturer möts.

1.2.7 Vilka nyttor och risker finns med gemensamma begrepp eller översättningsnycklar?

Målet med frågan om nyttor och risker var att kunna avgöra om och när det är lämpligt att utveckla en sammanhållen semantik.

Erfarenhet från andra domäner implicerar att det kan finnas stora nätverkseffekter med att göra gemensamma begrepp eller översättningsnycklar explicita, tillgängliga och eventuellt standardiserade. Nätverkseffekter som visat sig kunna öka värdet av data både för dataproducenter och datakonsumenter.

Nätverkseffekterna behöver dock inte alltid vara positiva. Att presentera en fritt vald koppling som fakta kan vara lika riskfyllt som att låta vem som helst presentera sin information på ens websida. Försiktighet, förtroende och sunt förnuft är minst lika viktigt som i andra sammanhang.

I projektet används termen ”översättningsnycklar” som ett sätt att associera den semantiska interoperabiliteten med vår förståelse för vad det innebär att göra en översättning från ett språk till ett annat.

I samhället har vi börjat få en förståelse för när det är lämpligt att använda en ”google-översättning” eller inte, och att gränsen förändras succesivt, dvs. att det är värt att hålla ögonen på pågående utveckling av digital översättningsförmåga. Liksom vid översättningar handlar det för datainfrastrukturen i slutändan om att ge digitala lösningar möjlighet att föra över tillräckligt mycket innebörd från en del av KLL-domänen till en annan del.

Projektet har inte haft möjlighet att ta arbetet hela vägen fram till beräkningar av vilka värden olika användare skulle få del av, om de skulle få tillgång till gemensamma begrepp eller översättningsnycklar. Det finns förslag på verksamhet och metod för att ta den centrala frågan vidare (se avsnitt 3.3). Men framför allt finns en förhoppning om att sprida förmågor att identifiera och beskriva vilka värden som står på spel.

1.2.8 Vilka är utvecklingsbehoven?

I myndigheternas delredovisning av uppdraget 2023 formuleras behovet av en koordinerande grupp för det informationsområde där grunddatadomäner för arbetsmarknad och utbildning överlappar varandra.⁵ Projektet bekräftar, och beskriver i mer detalj, behovet att utveckla samverkansförmågan kring gemensamma begrepp eller översättningsnycklar.

Även om semantikprojektet avslutades vid årsskiftet finns stora utvecklingsbehov kvar. I kapitel 3 beskrivs de verksamheter som projektets lärdomar ger särskild anledning att fortsätta med. Beskrivna verksamheter handlar både om redan startade projekt och förslag på verksamheter där den organisatoriska förankringen ännu inte formerats fullt ut.

⁵ Arbetsförmedlingen, *Uppdrag att utveckla en sammanhållen datainfrastruktur för kompetensförsörjning och livslångt lärande*, Af-2021/0053 0693

2 Identifierade förutsättningar

Utifrån identifierade frågor och kluster av frågor gick projektet vidare med tre verksamhetsspår. Avsnitt 2.1 beskriver innehållet i de olika verksamhetsspåren. De lärdomar som drogs i respektive spår sammanställdes och gav upphov till de tre kluster av lärdomar som beskrivs i avsnitt 2.2 till 2.4.

2.1 Verksamhetsspår som gett underlag till slutsatser

2.1.1 Case-spåret

Verksamheten i case-spåret har tagit sikte på att kartlägga förutsättningar genom att utgå från konkreta tankeexperiment med persona, livshändelser och ett gemensamt område av kunnande. Det första case som undersöktes närmre är en tänkt avhoppare från gymnasieskolans el- och energiprogram som vill vidareutbilda sig till fordonsmekaniker i en bransch under snabb och stor omvandling, framförallt med anledning av elektrifieringen. Den fiktiva personens livsresa har resonerats fram och vilka personer och funktioner som personen möter på sin väg från studier till arbetslivet, som exempelvis yrkeslärare, studie- och yrkesvägledare, vägledare inom arbetsförmedling, arbetsgivare, m fl.

Frågeställningen kan tyckas vara enkel: vilka livshändelser påverkar individen och vilka funktioner och institutioner möter individen på sin resa? Det kan, i fallet med en avhoppad elev från el- och energiprogrammet som ska arbeta som el-fordonsmekaniker, förefalla rimligt att anta att individen behöver delar av utbildningen från el- och energiprogrammet samt delar av utbildningen från fordonsprogrammet. Det är därför lätt att tänka att en avhoppad elev från gymnasieskolans el- och energiprogram som kompletterar sin utbildning med kurser från fordonsprogrammet på komvux skulle passa för yrkesrollen. Det har dock visat sig att det inte finns enkla svar på en till synes enkel fråga. Utbildningssystemet är flexibelt men samtidigt komplext. Komplexiteten gör att information och individuell studievägledning liksom individuell bedömning av individens kunnande i relation till branschens kompetenskrav blir mycket viktigt.

I projektet har vi analyserat innehållet i kurser som ingår i el- och energiprogrammet och relaterat dessa till de mål som finns inom kurser för utbildning till bussmekaniker och till läranderesultat som är framskrivna för den icke-formella men nivåbeslutade kvalifikationen bussmekaniker.

En elev som hoppat av ett gymnasieprogram efter halva tiden har sällan läst kurser som ger de kunskaper som behövs för arbete med eldrivna fordon. Den till synes enkla kombinationen av en halv el-utbildning inom gymnasieskolan och en halv fordonsutbildning inom komvux som skulle kunna leda till en examen från komvux på gymnasial nivå, skulle sannolikt inte vara tillräcklig för att möta de kompetenskrav som ställs för att arbeta som mekaniker på eldrivna bussmotorer

Ett andra case bestod av en nyanländ arbetssökande med erfarenhet som sjuksköterska i annat land och en erfaren undersköterska som möter en arbetsgivares behov av kompetensförsörjning. Arbetsgivaren har i exemplet behov av både en undersköterska och en specialistundersköterska. Exemplet var bra för att fånga hur mycket hemvisten för de olika stödprocesserna skiftade beroende på detaljer i de arbetssökandes bakgrund. En genomgående erfarenhet från casespåret är att arbetssättet varit användbart men också är resurskrävande.

För att påbörja arbete även med frågor om värdet av att ta fram gemensamma begrepp och/eller översättningsnycklar genomfördes även 12 intervjuer om nyttor och risker med gemensamma begrepp eller översättningsnycklar.

De som intervjuats valdes på klassiskt marknadsundersöknings manér så att de representerade personer som kan marknaden, som har god kännedom om användare, samt känner till vad som krävs för att få till den tekniska utveckling som krävs. Med bland de intervjuade fanns därmed såväl VD:ar och seniora HR-konsulter som produktansvariga och erfarna utredare.

Med ett undantag tog vardera intervju en timme och utgick från samma, i förväg utskickade, frågor. Intervjun började med en fråga om hur de intervjuade tolkade innebörden av påstått behov av ett gemensamt språk för utbildningsvärlden och arbetslivet. För att sedan följa upp med specifika frågor om vilka potentiella eller konstaterade konsekvenser de intervjuade såg av att få tillgång till två exempel på gemensamma begrepp eller översättningsnycklar. Skillnaden mellan exemplen var framför allt en fråga om detaljnivå.⁶ Resultat från intervjuerna beskrivs närmare i avsnitt 2.4.4.

2.1.2 Dataportal-spåret

Enligt regeringsuppdraget ska de deltagande myndigheternas för uppdraget relevanta datakällor finnas representerade i den av Digg förvaltade tjänsten Sveriges dataportal. I uppdraget nämns också ett antal olika terminologier som exempel på var det behövs relationer som brygger mellan olika klassificeringar för att beskriva kompetenser och kompetenskrav. De som nämns är Sveriges referensram för kvalifikationer (SeQF), Standard för svensk yrkesklassificering (SSYK2012), Svensk utbildningsnomenklatur (SUN), Standard för svensk näringsgrensindelning (SNI) och den europeiska klassificeringen av färdigheter, kvalifikationer och yrken (ESCO).

Sveriges dataportal är en nationell infrastruktur för att kunna hitta datamängder, API:er, specifikationer och begrepp/terminologier. Begreppsdelens av Sveriges dataportal bygger på standarden SKOS.⁷

I dataportalsspåret har de deltagande myndigheterna med stöd av Digg påbörjat arbetet med att identifiera för uppdraget relevanta terminologier med utgångspunkt i de som nämns i

SKOS (Simple Knowledge Organization System) är en standard för att uttrycka begreppsstrukturer. Många system för att organisera kunskap (ex. ordlistor, taxonomier och klassifikationer) delar liknande strukturer. SKOS fångar många av de likheterna genom att explicit definiera grundläggande koncept och relationer mellan begrepp. Se bilaga eller fotnot för längre beskrivning.

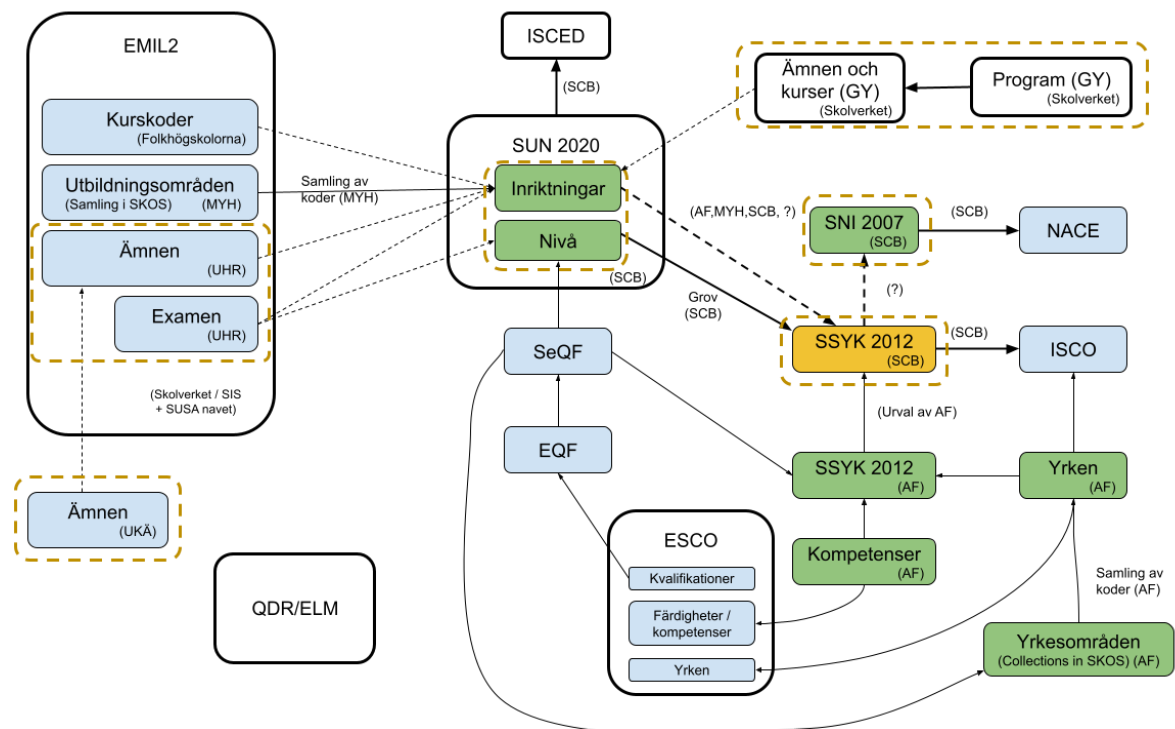
⁶ Se bilaga 4 för specifikation av metod och vilka som intervjuats. Notera att metod och urval innebär viktiga reservationer mot att dra generella slutsatser från resultaten.

⁷ SKOS Primer. <https://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>

uppdraget. Arbetet har därefter fortsatt med att titta på relationerna mellan de olika terminologierna, vem som ansvar för respektive terminologi, vilket format den är tillgängliggjord i och var den finns att hämta. Därefter har arbetet med att publicera myndigheternas terminologier på Sveriges dataportal påbörjats.

Genom detta arbete tillgängliggörs terminologier på ett enhetligt format så att allmänheten och andra myndigheterna lättare kan konsumera dem. Parallellt med detta har projektet testat att skapa relationer mellan begrepp på Sveriges dataportal. Detta har ökat förståelsen för de möjligheter och begränsningar som standarden SKOS innebär.

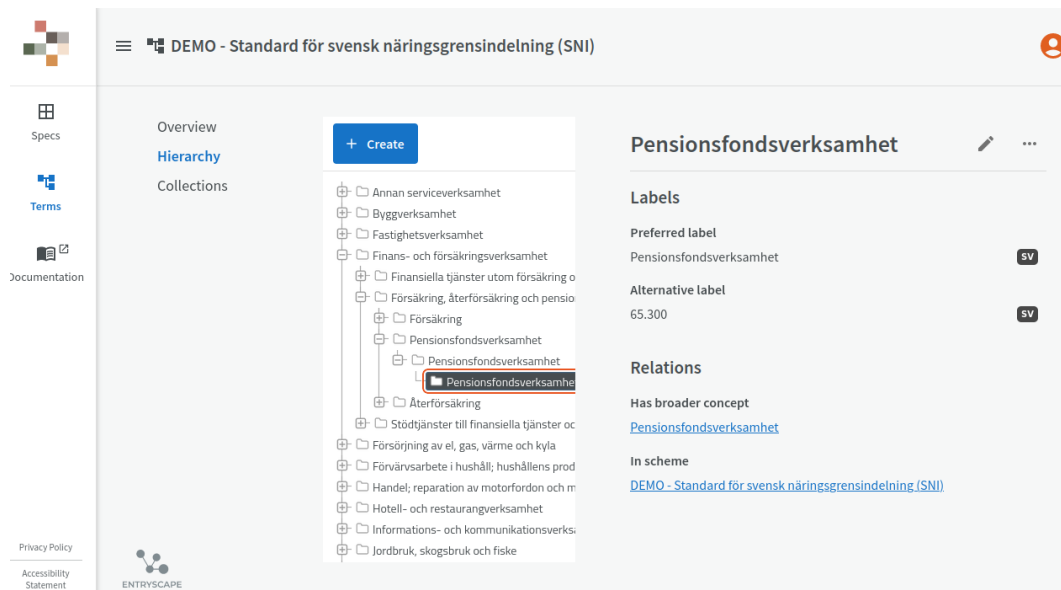
Publicering av begrepp på Sveriges dataportal



Figur 1: Diagram som ger en överblick över de terminologier som kan vara aktuella att publicera på Sveriges dataportal.

De olika bubblorna i blått, grönt, orange och vitt representerar terminologier. Pilarna mellan bubblorna representerar befintliga och möjliga kopplingar mellan begrepp ur de olika terminologierna. De terminologier som har streckade ramar kring sig har under projektet publicerats på dataportalen.

På Sveriges dataportal finns ett webbaserat verktyg för att redigera och utforska terminologier, begrepp och samlingar av begrepp baserade på SKOS-standard. Följande skärmbild är ett exempel på hur gränssnittet kan se ut när man redigerar terminologin *Standard för svensk näringsgrensindelning (SNI)*. I bilden åskådliggörs begreppen som en trädstruktur och vi kan till exempel se detaljer om begreppet *Pensionsfondsverksamhet*.



Figur 2: Bild av användargränssnitt för att redigera begreppsstrukturer i Sveriges dataportal

Publicering av begrepp

Inom **Arbetsförmedlingen** används begreppsstrukturen JobTech Taxonomy. Ett viktigt användningsområde är att kategorisera den information som finns i platsannonser på Platsbanken. Begreppsstrukturen består främst av begrepp som är centrala för arbetsmarknaden, däribland cirka 3700 yrkesbegrepp och cirka 5800 kompetensbegrepp. Dessa två grupper av begrepp exponeras som två olika SKOS-terminologier till Sveriges dataportal. Yrkesbegreppen består delvis av begrepp från SSYK men har kompletterats med ytterligare en nivå av *yrkesbenämningar* som inordnas under den lägsta SSYK-nivån. JobTech Taxonomy innehåller också begrepp för andra typer av information som kan förekomma i platsannonser, exempelvis geografiska platser och körkortstyper, men dessa har vi valt att utelämnas från publicering på dataportalen.

Från **Skolverket** har ämnen och program på gymnasiet importerats som två separata terminologier. Här är det viktigt att notera att man är tvungen att göra vissa förenklande antaganden för att kunna göra en översättning till SKOS. Exempelvis representeras bara det centrala innehållet för varje kurs, inte kunskapskraven. Vidare har kurser lagts till under respektive ämne – vilket visserligen ger en bra översikt – men man kan diskutera om det är lämpligt att en enskild terminologi bör innefatta begrepp som har väsentligt olika karaktär (i detta fall kurser och ämnen).

UHR ansvarar för ämnen för att kategorisera utbildningserbjudanden. Dessa ämnen ingår som en del i EMIL-standard och har nu exponerats som en terminologi. Alla ämnen har namn och en del har också synonymer som representeras inom SKOS som skos:prefLabel och skos:altLabel respektive.

Terminologin **Svensk indelning av utbildningsämnen** etablerades 2011 och förvaltas av **UKÄ (Universitetskanslerämbetet)**. Terminologin består av tre nivåer av namngivna ämnen, ibland med kortare förtydliganden om vad som ingår. Vid import tilldelades varje ämne en unik URI baserad på dess kod och ämnens

inbördes hierarki representerades med hjälp av skos:broader och förtydliganden med hjälp av skos:scopeNote. Terminologin finns på både svenska och engelska vilket bevarades enligt gängse sed med upprepade properties och språktaggar vid importen till SKOS.

Från **SCB** har tre olika terminologier importerats, SNI, SSYK och SUN. Då alla tre kommer från samma underliggande system (Metaplus) har överföringen hanterats på ett likvärdigt sätt. Det mesta av det som finns i systemet som benämningar, alternativa benämningar, allmänna noter och vad som ingår och inte ingår går att överföra direkt till SKOS. Endast begränsningar i giltighet saknar motsvarighet, men inget av de terminologier som importerats använder dessa.

2.1.3 Matchningsindikator-spåret

Innehållsbeskrivning & syfte

Arbetet med ”Matchningsindikatorer - SUN/SNI/SSYK”, även kallat ”Matchningsindikatorspåret”, kan delas in i de två huvudaktiviteterna:

- 1 ”Etablerande av matriser och matchningsindikatorer”
(genomfört apr 2022 - maj 2022) samt
- 2 ”Logisk identifikation av samhörighet mellan klassifikationer samt utvärdering”
(genomfört aug 2022 – okt 2022).

Matchningsindikatorspåret har genomförts för att gemensamt utreda förutsättningarna att utveckla begreppsstrukturer inom SCB:s klassifikationer för utbildnings- och arbetsmarknadsområdet. Aktiviteten syftar mer specifikt även till att inhämta kunskap och lärande om hur väl koderna i SSYK (Standard för svensk yrkesklassificering), SUN (Svensk utbildningsnomenklatur) och SNI (Svensk Näringsgrensindelning) är sammanlänkade på olika aggregeringsnivåer.

De versioner av klassifikationer som användes i detta arbete var de senast framtagna klassifikationsversionerna:

- SNI2007 (Bransch),
- SSYK2012 (Yrke) och
- SUN2020 (Utbildningsnivå och -inriktning).

Även klassifikationsversionen SUN2000 är indirekt inbegripen i arbetet.

Under det första halvåret 2022 tog SCB fram ett bakgrundsunderlag inom ramen för huvudaktiviteten ”Etablerande av matriser och matchningsindikatorer”, bestående av delaktiviteterna ”Informationsmatriser” och ”Matchningsindikatorer”. Dessutom genomförde SCB en intern workshop med betoning på kvalitetsfrågor med utpekade klassifikationsansvariga (KLASA) och vissa experter inom register-, mät- och metodområdet.

Under hösten 2022 genomfördes sedan en form av analysarbete inom ramen för huvudaktiviteten ”Logisk identifikation av samhörighet mellan klassifikationer samt utvärdering”, bestående av delaktiviteterna ”Logisk identifikation av samhörighet i klassifikationer” och ”Utvärdering av identifikation av samhörighet i klassifikationer”.

1) Etablerande av matriser och matchningsindikatorer

De data som ligger till grund för de nivåspecifika matriserna avser avidentifierade individdata från SCB:s register UREG (Utbildningsregistret), YREG (Yrkesregistret) samt FDB (Företagsdatabasen) för årgångarna 2015, 2016, 2017, 2018, 2019, 2020.

De sex separata tabellerna för årspopulationerna 2015-2020 utgår från nödvändiga data i UREG som matchas genom variabeln avidentifierad individ med nödvändiga data i YREG och matchas därefter med data från anställda i FDB.

Figur 3 nedan visar antalet befintliga kombinationer mellan de tre klassifikationerna, där varje kombination består av minst 10 individer. Notera att vi för SUN-kod enbart använder SUN-inriktning och bortser från indelning baserat på olika utbildningsnivå. Nivåbenämningen "DETALJ" avser 4-ställig nivå för SUN_Inr2020 och SSYK2012 samt 5-ställig för SNI2007.

NIVÅ	<u>MATRIS A:</u> SUN2020 SSYK2012	<u>MATRIS B:</u> SUN2020 SNI2007	<u>MATRIS C:</u> SSYK2012 SNI2007	<u>MATRIS D:</u> SUN2020 SSYK2012 SNI2007
1-ställig	81	180	178	1 490
2-ställig	941	1 753	2 836	19 150
3-ställig	9 702	13 951	12 712	45 274
DETALJ	32 284	43 284	30 443	55 254

Matristyp A: Kombinationer mellan utbildningsinriktning och yrke

Matristyp B: Kombinationer mellan utbildningsinriktning och bransch

Matristyp C: Kombinationer mellan yrke och bransch

Matristyp D: Kombinationer mellan utbildningsinriktning, yrke och bransch.

Figur 3: Antal faktiskt förekommande kombinationer mellan klassifikationer i 2020 års populationsdata

2) Logisk identifikation av samhörighet mellan klassifikationer samt utvärdering

Som ett andra steg i huvudspåret om matchningsindikatorer genomfördes en analys av hur väl klassifikationerna bedöms "hänga ihop" och är kopplade till varandra på olika nivåer, utifrån manuella logiska bedömningar av textbeskrivningar för respektive klassifikation och nivå. För den 1-ställiga nivån gjordes således jämförelser för följande sex kombinationer: SNI1_SSYK1 och SSYK1_SNI1, SNI1_SUN1 och SUN1_SNI1, samt SSYK1_SUN1 och SUN1_SSYK1.

De manuella analyserna gjordes i särskilt framtagna mallar. Mallen innehöll en tabell per kombination av två klassifikationer, exempelvis en tabell för SNI1_SSYK1. I tabellerna angavs för varje kod i den första klassifikationen vilka koder på samma nivå i den andra klassifikationen som bedömdes vara kopplade.

I ett nästkommande steg gjordes sedan en jämförelse mellan de kombinationer som manuellt ansågs hänga ihop, och de framtagna matchningsindikatorerna, vi kallar denna jämförelse för "Utvärdering av identifikation av samhörighet". En utvärdering görs enbart av de kombinationer som utpekats i den manuella delen. I utvärderingen jämförs dessa med matchningsindikatorerna (beräknade populationsandelar).

Tanken här är att denna övning med manuell logisk identifikation av samhörighet + utvärdering av identifikation av samhörighet ska generera insikter om hur väl bedömd samhörighet mellan kodkombinationerna kan återspegla den faktiska samhörigheten (betraktat ur ett populationsregister perspektiv).

Vi går vidare genom att studera resultatet från utvärderingen. Tillvägagångssättet är ovetenskapligt och ganska subjektivt, men tjänar ändå till att ge en uppfattning om hur väl man manuellt prickar in kodkombinationer som sett till textbeskrivningarna av klassifikationskoderna överensstämmer med de kombinationer som har större andelar i matchningsindikatorerna, som då i detta fall avser 2020 års registerpopulation.

För att få fram en sorts tröskelreferensvärde utgår vi här från antalet koder som ska fördelas på populationsandelarna. Om vi exempelvis har $n=10$ koder på 1-ställig nivå så anser vi här att varje kod teoretiskt förväntas gälla för $1/n = 10\%$ av populationen (förväntat enhetsmått), så att alla koder inom en nivå, i.e. $10 \cdot 10\%$, förväntas motsvara 100% av populationen. Vårt tröskelvärde blir då $0,1$.

Om sedan en utpekad kodkombination i den manuella bedömningen visar sig ha en populationsandel som överstiger $0,1$ för den relevanta kombinationen, så anses kombinationen avse en korrekt bedömning, jämfört med faktiskt registrerade koder i populationen, i annat fall inte.

I en del av de slutförda excelfilerna med manuella bedömningar har noterats inte bara om det anses finnas en koppling mellan de två klassifikationskoderna utan också om kopplingen är stark. För dessa kombinationer har en separat utvärdering gjorts, där tröskelvärdet för utvärdering om korrekt bedömd kombination gentemot matchningsindikator är det dubbla gentemot att det finns en koppling, dvs $2/n$.

2.1.4 Sekretariatet och andra myndigheter

Utöver de tre verksamhetsspåren har projektet i olika konstellationer genomfört möten och samarbeten med sekretariatet och andra intressenter som nämnts i regeringsuppdraget.

Det mesta av projektets verksamhet har handlat om att hitta beskrivningar av och metoder för att tydliggöra hur olika perspektiv möts. Möten och samarbeten med andra grupperingar har därför ofta känts som att börja om på nytt för varje tillfälle. Samtidigt har behov av underlag till dessa samarbeten i flera fall gjort att arbetsgruppen haft anledning att formulera delresultat, var för sig eller i ett gemensamt dokument.

Den typiska reflektionen vid dessa möten är att ”det här är komplicerat”. En av de största utmaningarna i projektet har varit att inte stanna där, utan hitta vägar förbi den överväldigande mängd av sammanhang och perspektiv som berörs. Möten och samarbeten med sekretariatet och andra intressenter har därmed bidragit till kontinuerligt arbete med att slipa på formuleringar av lärdomar, slutsatser och förslag.

Samarbetet med sekretariatet har bjudit på flera milstolpar i det avseendet. Först med presentation av lärdomar och delresultat till uppdragsgemensam workshop den 15 juni. Därefter många avstämningar inom respektive myndighet i arbetet med att ta fram projektförslag inför 2023. Ett gemensamt möte hölls 5 oktober då projektet hade en heldag för att formulera lärdomar att ta med till fortsatt förvaltning. Under hösten fanns även många varianter av myndighetsseparata avstämningar mellan projekt och sekretariat. Bland annat i samband med sekretariatets workshopar 12 oktober och 23 november samt genom medskick till och synpunkter på sekretariatets *Utredning av förslag till: en gemensam förvaltning och utveckling av en sammanhållen datainfrastruktur för kompetensförsörjning och livslångt lärande*, samt inför återrapport *Uppdrag att utveckla en sammanhållen datainfrastruktur för kompetensförsörjning och livslångt lärande – Delredovisning 31 januari 2023* (båda del av diarienummer Af-2021/0053 0693).

Flera intressenter nämnda i regeringsuppdraget hade liksom projektet deltagare på de Open Space som genomfördes 16-17 mars samt 1-2 juni. Projektledaren deltog därutöver på Yrkeshögskolans *Hearing – Generella kompetenser* 19 november samt Diggs Seminarium *Sverige i Europas dataekosystem*, 15 december.

Vid fyra tillfällen har projektet mött representanter från lärosäten i deras arbete med projektet *Högskolans digitala ingång för livslångt lärande*. De senare mötena var med deras arbetsgrupp för taxonomi. Att även det projektet har haft en specifik arbetsgrupp för taxonomin är ett av flera exempel på hur de frågor som semantikprojektet fokuserat på återkommer i många sammanhang. Slående är att det, i många projekt och verksamheter med fokus på dessa taxonomier, är svårt att få till den bredd och kontinuitet av resurser som behöver samordnas för att det ska leda till utveckling på önskad detaljnivå.

2.2 Teknik och statistik som kan tydliggöra kopplingar mellan existerande begreppsstrukturer

Processen att skapa relationer mellan begrepp i skilda begreppsstrukturer går i den internationella litteraturen under benämningar som "ontology matching", "ontology alignment" och "ontology mapping". Skapandet av relationer kan ske för hand eller med hjälp av matchningsalgoritmer. Otero-Cerdeira et al.⁸ ger en överblick över olika tekniker för att skapa relationer algoritmiskt och delar in dessa i olika klasser. Ett exempel på en sådan uppdelning kan vara om algoritmen i första hand är innehållsbaserad eller kontextbaserad. En innehållsbaserad algoritm använder i första hand innehållet i de begreppsstrukturer som ska matchas, exempelvis begreppsdefinitioner. En kontextbaserad algoritm använder däremot data utanför begreppsstrukturerna, exempelvis hur begreppen används i olika sammanhang. De två huvudaktiviteterna inom Matchningsindikatorspåret är utforskande studier av hur begrepp mellan olika begreppsstrukturer hänger samman snarare än algoritmer för att skapa översättningsnycklar. Men om man ändå ska försöka ordna in dem i tidigare klassificering, kan aktiviteten "Etablerande av matriser och

⁸ Otero-Cerdeira, L., Rodríguez-Martínez, F. J., & Gómez-Rodríguez, A. (2015). Ontology matching: A literature review. *Expert Systems with Applications*, 42(2), 949-971.

matchningsindikatorer” sägas vara kontextbaserad eftersom den utgår från hur begrepp används, bland annat baserat på SCB:s register med individdata. På samma sätt kan aktiviteten ”Logisk identifikation av samhörighet mellan klassifikationer samt utvärdering” sägas vara innehållsbaserad eftersom den utgår från begreppens textbeskrivningar.

För att jämföra hur bra olika algoritmer fungerar finns Ontology Alignment Evaluation Initiative (OAEI)⁹ som varje år publicerar en jämförelse över hur bra olika algoritmer presterar i ett antal tester. Att tillämpa algoritmer, som presterar särskilt bra i OAEI, på begreppsstrukturerna som Semantikprojektet berör skulle vara ett intressant fortsatt arbete.

Det tekniska stödet för begrepp och översättningsnycklar mellan begrepp utgörs av standarder och mjukvara implementerad i enlighet med en sådan standard. Sveriges dataportal stöder idag terminologier som följer standarden SKOS: Simple Knowledge Organization System.¹⁰ Inom Dataportalspåret (avsnitt 2.1.2) har vi undersökt möjligheten att publicera befintliga begreppsstrukturer på Sveriges dataportal. Den främsta lärdomen av detta arbete är att SKOS är tillräckligt uttrycksfullt för att representera de väsentliga delarna av de terminologier som kan vara aktuella att publicera på Sveriges dataportal, men att vissa detaljer inte kan uttryckas, som exempelvis speciella attribut som är kopplade till begrepp men som inte har någon motsvarighet i SKOS.

I enlighet med SKOS har Sveriges dataportal idag också stöd för att koppla ihop två begrepp med någon av relationerna `broadMatch`, `narrowMatch`, `closeMatch` eller `exactMatch`.¹¹ Huruvida dessa relationer är tillräckligt uttrycksfulla för att representera de översättningsnycklar som är aktuella inom Semantikprojektet återstår att se. En uppenbar begränsning är att det i SKOS inte går att koppla extra data till en relation: ett exempel på sådana data skulle kunna vara relationens användningsområde eller någon form av mått på hur stark en relation är.

En annan fråga är hur man ska hantera alternativa översättningsnycklar: man skulle kunna tänka sig att det mellan två begreppsstrukturer finns olika översättningsnycklar beroende på ändamål. Eftersom SKOS uttrycker begrepp och relationer mellan begrepp genom listor av så kallade tripletter (subjekt-predikat-objekt) kan alternativa översättningsnycklar uttryckas som olika uppsättningar av sådana tripletter.

Genom att tillämpa Data Science (DS) med vetenskapliga metoder, processer, algoritmer och system för att extrahera eller extrapolera kunskap och insikter, och tillämpa kunskap från data över domäner kan statistiken koordineras och förenas mellan existerande begreppsstrukturer. Inom DS inbegrips dataanalys, informatik och relaterade metoder för att förstå och analysera faktiska fenomen med data. Genom att extrahera kunskapen från stora datamängder och i ett nästa steg tillämpa kunskapen och insikterna från dessa data ges större möjligheter att lösa problem

⁹ <http://oaei.ontologymatching.org/>

¹⁰ <https://www.w3.org/2004/02/skos/>

¹¹ <https://www.w3.org/TR/skos-primer/#secmapping>

inom ett brett spektrum av tillämpningsdomäner, däribland områden relaterade till ekosystemets intressenter.

En viktig del i DS är maskininlärning (ML) vilket i sin tur är en typ av artificiell intelligens (AI) som tillåter mjukvaruapplikationer att bli mer exakta när det gäller att förutsäga resultat, utan att vara explicit programmerad att göra detta. ML-algoritmer använder historiska data som indata för att förutsäga nya utdatavärden, och ger god kunskap om trender och mönster i data. Många av dagens ledande företag, som exempelvis Facebook, Google och Uber använder ML inom en central del av sin verksamhet, och har genom detta skapat betydande konkurrensfördelar mot många företag.

En ML-algoritm inom språkbehandling som har fått utbredd praktisk användning är Bidirectional Encoder Representations for Transformers (BERT)¹². Denna algoritm tränas först på stora mängder text för att erhålla en generell språkförståelse. En färdigtränad BERT-modell tar som indata en kortare eller längre text och producerar som utdata numeriska värden. Hur dessa numeriska värden sedan används beror på ändamålet som BERT-modellen tränades för.

Sentence-BERT¹³ är en utökning av BERT för ändamålet att koda om en text till en numerisk vektor, det vill säga en lista med ett bestämt antal numeriska värden, som fångar innebörden av texten. Detta är användbart i ett datasystem för att kunna hitta texter som betyder samma sak även om de använder helt skilda ord och formuleringar.

Det görs genom att de texter som ska jämföras kodas om till numeriska vektorer vilket gör det möjligt att hitta texter med liknande betydelse som närliggande vektorer i en vektorrymd. Det här kan vara användbart för att hitta kopplingar mellan begrepp från två olika begreppsstrukturer. Sentence-BERT har exempelvis använts för att hitta kopplingar mellan de yrkesrelaterade taxonomierna ESCO och O*Net,¹⁴ genom att vektorer beräknas för begreppens textuella definitioner. Ett annat exempel på en mer komplex algoritm som bygger på BERT, men inte Sentence-BERT, är BERTMap¹⁵ som kombinerar strängmatchning med en BERT-baserad klassificerare samt lokala kopplingar mellan närliggande begrepp för att matcha begrepp mellan olika begreppsstrukturer.

En fråga som dyker upp i samband med användningen av maskininlärning och språkbehandling på arbetsmarknadsrelaterade begrepp är hur bra algoritmerna förstår fackuttryck inom arbetsmarknaden och utbildningsvärlden. I allmänhet begränsas en maskininlärningsalgoritms prestation av kvaliteten på den data som den tränas på. För svenska språket har Kungliga biblioteket tränat och publicerat en

¹² Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

¹³ Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

¹⁴ Neutel, S., & de Boer, M. H. (2021). Towards Automatic Ontology Alignment using BERT. I *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*.

¹⁵ He, Y., Chen, J., Antonyrajah, D., & Horrocks, I. (2022). BERTMap: a BERT-based ontology alignment system. I *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 5, pp. 5684-5691).

BERT-modell på en generell textmassa.¹⁶ Denna modell kan emellertid finjusteras för specifika ändamål genom att ytterligare träna den på text som är representativ för dess användningsområde.

För ändamålet att träna modellen att känna igen begrepp så som yrkesbenämningar skulle de historiska platsannonser som JobTech har publicerat¹⁷ med tillhörande yrkesbenämningar och SSYK-koder nivå 4 kunna vara användbara. Däremot skulle dessa platsannonser troligtvis vara otillräckliga för att modellen ska lära sig begrepp från andra områden som exempelvis utbildningsvärlden.

Nyligen utvecklade Europeiska kommissionen, som ansvarar för ESCO, BERT-baserad teknik som kan relatera text från flera olika språk till en vektorrepresentation.^{18, 19} Den här BERT-modellen kan användas både för att klassificera platsannonser till yrkesbegrepp och att föreslå kopplingar mellan nationella yrkesbegrepp och ESCOs yrkesbegrepp. De data som algoritmen är tränad på är bland annat platsannonser. Därmed får algoritmen en särskilt god förståelse för de yrkesrelaterade begrepp som återfinns i dessa. Däremot kan det vara otillräckligt för att den ska lära sig fackuttryck och yrkesrelaterade detaljer som inte uttrycks i platsannonser.

Arbetet med att publicera de olika terminologierna i SKOS-format på Sveriges dataportal har bidragit till regeringsuppdragets leverans och underlättar fortsatt publicering genom att det har skapats ett tekniskt ramverk som gör det lättare att konvertera begrepp från olika källor till SKOS. Det praktiska arbetet som skett i samband med ramverket har visat på de stora likheter som finns mellan olika terminologier och visar på de möjligheter och begränsningar som finns med SKOS-standarden.

Det webbaserade verktyget på dataportalen har vidareutvecklats för att möjliggöra en bättre separation mellan det som importeras och den ytterligare information som man lägger till i verktyget, så kallade *förstärkningar*. Exempel på förstärkningar, kan vara översättningar till andra språk, förtydliganden av definitioner och nya typer av relationer. Denna separation gör att man inte förlorar information om man behöver uppdatera terminologer från externa källor efter att man redigerat dem. Detta skulle annars vara ett stort hinder för underhåll av förstärkningar, särskilt relationer i ett längre perspektiv. Verktyget har också fått stöd för att skapa samlingar av begrepp från flera olika terminologier.

En begränsning i SKOS-standarden är att det inte finns något etablerat sätt att koppla extra information till relationerna mellan begreppen, till exempel att uttrycka en sannolikhet att två begrepp hör ihop. Inom projektet har vi använt skos:editorialNote som ett alternativt arbetssätt. Detta arbetssätt är dock oprecist, särskilt om man har många relationer utgående från samma begrepp. En mer fullödig lösning kräver ett

¹⁶ Malmsten, M., Börjeson, L., & Haffenden, C. (2020). Playing with Words at the National Library of Sweden—Making a Swedish BERT. *arXiv preprint arXiv:2007.01658*.

¹⁷ <https://data.jobtechdev.se/annonser/historiska/index.html>

¹⁸ <https://esco.ec.europa.eu/en/about-esco/data-science-and-esco/machine-learning-assisted-mapping-multilingual-occupational-data-esco-part-1>

¹⁹ <https://esco.ec.europa.eu/en/about-esco/data-science-and-esco/machine-learning-assisted-mapping-multilingual-occupational-data-esco-part-2>

annat angreppssätt som med fördel etableras i dialog inom ett större samarbete med fokus på frågor kring standardisering, gärna på nordisk, europeisk eller internationell nivå.

För att förenkla både i redigering och presentation av terminologier gjordes ett val att begränsa de begrepp som publiceras på Sveriges dataportal till begrepp som har högst ett bredare begrepp trots att SKOS-standarden tillåter flera bredare begrepp. Vår förhoppning är att de få fall när denna begränsning är problematisk kan man lösa situationen genom att dela upp terminologin i flera. Om uppdelningen görs rätt kan man då ändå åstadkomma samma struktur genom att ha flera skos:broadMatch mellan begrepp i olika terminologier (istället för flera skos:broader inom samma).

Arbetet med matrisparet påvisade att tillvägagångssätt samt själva bedömningen görs på helt olika sätt då personer analyserar kodkombinationer, om man inte först tar fram ett detaljerat gemensamt tillvägagångssätt. Dessutom kan tilläggas att hur kodkombinationerna representeras i populationen inte nödvändigtvis var det som deltagarna primärt hade med i sin bedömning. Det man analyserade när identifikationen gjordes, var egentligen om textbeskrivningarna för koderna var ihopkopplade. Till skillnad från en bedömning av hur stor andel i populationsregistren hade de olika kodkombinationerna.

När vi tittade på de detaljerade nivåerna så var det slående att samtliga utpekade i denna övning, ca 200 kombinationer faktiskt varit korrekt bedömda. Även om det selektiva urvalet kan ha gjorts med tanke på den kunskapsnivå som analyspersonen har om just de valda koderna, så kan man nog påstå att ju mer detaljerad nivå desto bättre träffsäkerhet i bedömningarna.

2.2.1 Huvudsakliga lärdomar

- Vid manuella bedömningar märks stora skillnader i tillvägagångssätt och hur bedömningar görs. Det behövs mer arbete för att prata sig samman för att göra mer likartade bedömningar
- Samtliga manuella kombinationer som analyserades på detaljerade nivåerna var korrekt bedömda. Även om det finns metodologiska betänkligheter finns det fog att tro att bedömningarna har bättre träffsäkerhet desto mer detaljerad nivå de görs på.
- Textbeskrivningarna för koder på mer aggregerade nivåer är inte alltid heltäckande gentemot innehållet i underliggande mer detaljerade koder, och nedkortade textbeskrivningar bör kanske ses över
- SSYK och SUN är närmare ihopkopplade än någon bilateral kombination med någon av dessa och SNI
- Att tillgängliggöra de olika terminologierna enligt SKOS-standarden underlättar informationsutbyte mellan myndigheter. Det blir lättare att konsumera ett begrepp från en annan myndighet när alla följer samma standard för att beskriva strukturen i sina terminologier.
- Det finns begränsningar i SKOS som kan vara problematiska för vissa ändamål men det är tillräckligt uttrycksfullt för att representera de väsentliga delarna av de terminologier som kan vara aktuella att publicera på Sveriges dataportal.

- En fördel med att använda SKOS är att det kan ge kostnadsbesparingar då verktyg som utvecklas för terminologier på SKOS-standarden kan återanvändas.
- Det finns många anledningar att fortsätta följa hur algoritmer och metoder inom så kallad "ontology alignment" kan användas för att undersöka och skapa relationer mellan relevanta begreppsstrukturer inom KLL-domänen.

2.3 Finess krävs för översättning mellan aspekter på kunnande

På samma sätt som när information går från en engelsk till en svensk kontext menar vi med översättning i det här avsnittet att så mycket som möjligt av ursprungliga och avsedda konnotationer överförs till det nya sammanhanget. I *Semantikprojektet – Lägesrapport, 23 augusti 2022*²⁰ konstaterade projektgruppen att det behövs mer än en översättning för att förmedla relevant information från ett utbildningssammanhang till ett arbetssammanhang.

Eftersom översättningen i projektet framförallt sker mellan utbildning och arbetsliv är läranderesultat eller resultat av lärande en bärande princip. Läranderesultat kan uttryckas som det en individ vet, kan och förstår, och/eller kan göra efter en avslutad lärandeprocess, men det fångar bara det som individen har uppnått (jfr engelskans *achieved learning outcomes*).

Dessa läranderesultat behöver sättas i relation till vad individen förväntas veta, förstå, och/eller kunna göra efter en avslutad läroprocess. Läranderesultat kan därför också ses som en kravställning eller kompetenskrav (jfr engelskans *intended learning outcomes*), vilka ofta definieras för ett arbetsområde eller yrke, eller genom betygskriterier eller examensmål för en utbildning.

Med den förståelsen i ryggen går det att utarbeta mer preciserade beskrivningar av utmaningarna med att hitta översättningsnycklar och/eller gemensamma begreppsstrukturer. Förståelsen gör det även möjligt att begränsa sökandet efter översättningsnycklar till läranderesultat i form av kompetenskrav (*intended learning outcomes*) eftersom bedömningen om enskilda individer har uppnått dessa ligger utanför projektet. Men projektet har även börjat ringa in några sökvägar för att hitta lösningar på de utmaningarna:

- Med fokus på hur informationen används i respektive sammanhang har projektet identifierat existerande experter för två led i översättning mellan information om utbildningar och yrken: (1) lärare och mer specifikt yrkeslärare då det gäller läranderesultat med tydlig koppling till arbetslivet, samt (2) branschprocesser för att sammanställa yrkesprofiler.
- Oavsett om översättningsnycklar tas fram med hjälp av manuella redaktioner eller algoritmer för maskininlärning, behövs sådana domänexperter för att bedöma kvalitet på framtagna översättningsnycklar.

²⁰ Dnr: Vinnova-2021-04628

2.3.1 Läranderesultat har ingen direkt motsvarighet i den svenska läroplansstrukturen

Skolväsendets kurs- och ämnesplaner (även benämnt som styrdokument i texten) ska innehålla ett *syfte* som ger svar på frågan varför ämnet behövs och ringar in ämnets centrala kunskapsdomäner, ett *centralt innehåll* som anger vad som ska tas upp i undervisningen och *betygskriterier*. I ämnenas syfte formuleras mål för undervisningen, vilket har vissa likheter med avsedda lärandemål (eng. *intended learning outcomes*). Men skillnaden är att målen siktar in sig på lärarens planering av undervisningen och inte på bedömningen av elevernas kunskaper.

I ämnenas centrala innehåll (eng. *content standards*) formuleras såväl innehållsstoff som handlingar och beteenden som eleverna ska få möta, pröva och öva på. Även här är det läraren som är adressat, dvs. att det som beskrivs är krav på input och inte krav på output. I ämnenas betygskriterier formuleras de mest centrala kännetecknen på relevanta ämneskunskaper (eng. *criteria for grading*). De är identiska mellan stegen E/C/A med undantag av enstaka värdeord. Tolkningen av betygskriterierna är i hög grad beroende av inte minst det centrala innehållet, som ger kontext och konkretion åt de relativt allmänt hållna skrivningarna i kriterierna.

Av ovanstående följer att det inte finns någon given kategori i kurs- och ämnesplanerna som motsvarar de externa ramverkens och kvalifikationernas läranderesultat (*intended learning outcomes*).

Visserligen har betygskriterierna det gemensamt med läranderesultat att de beskriver kvaliteter som mäts efter att undervisningen avslutats, men betygskriterierna refererar i sin tur till det centrala innehållet och är dessutom mer övergripande skrivna än vad läranderesultat behöver vara för att fungera optimalt. Den ämnesbetygsmodell som ska börja gälla från och med 2025 skapar ytterligare behov att formulera betygskriterier mindre detaljerat eftersom samma betygskriterier behöver fungera för att sätta betyg på samtliga nivåer inom ett ämne, ett ämne som består av flera nivåer som bygger på varandra.

Ett införande av läranderesultat som uttrycklig kategori i ämnesplanerna skulle möjliggöra en direkt jämförelse men också innebära en systemförändring som Skolverket bedömer inte är möjlig utan föregående utredning och nödvändiga politiska beslut.

2.3.2 Aspekter av kunnande uttrycks på olika sätt

Skolväsendets styrdokument och kompetenskraven för arbetslivets yrkesroller uttrycker därför olika aspekter av kunnande. Utöver skillnader i aspekter har utbildningen i gymnasieskolan också ett bredare syfte. Skolväsendets uppdrag, så som det regleras i skollagen, läro-, kurs- och ämnesplaner är tredelat och omfattar elevernas utveckling till ansvarskännande människor och samhällsmedborgare, deras förberedelse för aktivt deltagande i det demokratiska samhällslivet och deras förberedelse för vidare studier och arbetsliv. För att uppnå det breda syftet ingår olika kurser i utbildningen som dels ger en bredd, dels ger en grund för det valda studie-, yrkes- eller kompetensområdet.

För att underlätta ett arbete med att ta fram möjliga synonymer och översättningar mellan utbildning och kompetenskrav behöver man –eller maskin– härleda vilka resultat av lärande som förväntas vara uppnådda efter en utbildning för att kunna jämföra den med de kompetenskrav som finns inom arbetslivet.

För övergången mellan gymnasieskola och högre studier är översättningar mellan utbildningsbeskrivning och kompetenskrav mindre problematiska eftersom kompetenskraven definieras som den behörighet i termer av utbildningar som krävs för tillträde till högre studier. Däremot behöver processen härleda till om individens kunnande motsvarar kraven på läranderesultat som den sökande vill åberopa genom reell kompetens för behörighet eller tillgodoräknande. Men dessa fall sker huvudsakligen från andra erfarenheter än studier, som till exempel från arbetslivserfarenhet.

För att identifiera synonymer för att översätta individens kunnande efter en utbildning till kompetenskrav i arbetslivet behöver således ett steg tas som omvandlar utbildningens innehåll och omtolkar det till läranderesultat (intended learning outcomes). Detta steg bedöms kunna göras av en individ som har kunskap om utbildningen och kompetensområdet för yrket, till exempel en yrkeslärare. En yrkeslärare bedöms ha förmågan att såväl tolka fackuttryck som att kunna läsa mellan raderna utifrån sin dubbla yrkeskompetens som lärare och yrkesutövare och detta kan ske såväl intuitivt som medvetet.

Bedömningen är att den individ som kommer från enbart ena sidan – från utbildning eller arbetsliv – inte har samma förutsättningar att läsa mellan rader och göra den härledning som krävs för att en översättning ska bli meningsfull. Det första steget, att förtydliga och konkretisera utbildningens mål, genom att härleda dessa ur utbildningens innehåll, är nyckeln till att formulera beskrivningar av relationer mellan begrepp i styrdokument och kompetenskrav i arbetslivet.

För att fungera som översättningsnycklar mellan utbildningar och yrken behöver även relationer mellan arbetslivets kompetenskrav och de kritiska arbetsuppgifter som ger upphov till dessa kompetenskrav hållas aktuella och kvalitetssäkras. För att översättningen ska fungera hela vägen mellan arbetsuppgifter och styrdokument behövs därför nära kontakt med en god representation av yrkeslivet. Detta kan med fördel genomföras i ett nära samarbete mellan företrädare för utbildning och arbetsliv.

2.3.3 Möjlighet att använda Europass tillägget som översättningsnyckel

Skolväsendets läroplaner avspeglar att skolans uppdrag är bredare än kvalifikationer för specifika yrkesroller och kurs- och ämnesplanernas struktur medger inte en direkt koppling till läranderesultat. Frågan är då på vilket sätt, i vilken omfattning och var i skolväsendet som det går att hitta lämpliga kopplingar mellan läroplanen och externa kvalifikationsramverk?

Skolverket ser en möjlighet till en vidareutveckling av de tillägg till yrkesexamen på gymnasial nivå (Certificate Supplement – Europass) som finns för de olika yrkesprogrammets inriktningar, att de tas fram på en ökad detaljnivå för de olika yrkesutgångarna men också för yrkespaket inom kommunal vuxenutbildning och

gymnasieskolans introduktionsprogram. Skolverket avser att undersöka detta vidare under 2023. Genom sådana aggregerade mål (Europass) skulle översättningar mellan utbildningsinformation och kvalifikationer kunna göras utan direkta kopplingar till skolväsendets styrdokument.

2.3.4 Huvudsakliga lärdomar

- Slutsatsen av casemetodiken är att det är svårt att jämföra läranderesultat som kan härledas ur gymnasieskolans styrdokument med de kompetenskrav som ställs upp av arbetslivet. Det är inte meningsfullt att jämföra dem på så aggregerad nivå som examensmål, men det är också svårt att jämföra dem på ämnes- och kursnivå eftersom kompetenskraven inte är formulerade med samma skärning.
- Det är viktigt att förstå skillnaderna i de skärningarna, exempelvis för att kunna ta fram översättningsnycklar mellan ämnesplaner, dokumentation av lärande, läranderesultat och arbetslivets kompetenskrav.
- I ämnenas syfte formuleras mål för undervisningen, vilket har vissa likheter med avsedda lärandemål (eng. *intended learning outcomes*). Men skillnaden är att målen siktar in sig på lärarens planering av undervisningen och inte på bedömningen av elevernas kunskaper.
- Oavsett om översättningsnycklar tas fram med hjälp av manuella redaktioner eller algoritmer för maskininlärning, behövs domänexperter för att bedöma kvalitet på framtagna översättningsnycklar.
- Skolverket ser en möjlighet till vidareutveckling av en typ av översättning mellan utbildningsinformation och kvalifikationer genom det som i Europass kallas Certificate Supplement.

2.4 På väg mot en värdering av gemensamma begrepp och/eller översättningsnycklar

Det är en utgångspunkt för projektet att semantisk interoperabilitet²¹ är en grundförutsättning för att förmedla information från en del av KLL-domänen till en tjänst som behövs i en annan del. Som exempel behövs information om vilka utbildningar som är lämpliga och tillgängliga för att lyckas få till en önskad omställning mitt i arbetslivet.

Projektet har på olika sätt sökt svar på vilka nyttor och risker det finns med gemensamma begrepp och/eller översättningsnycklar. Det hade varit önskvärt att nå hela vägen fram till en regelrätt konsekvensanalys, inklusive explicita beräkningar av mervärden och kostnader. I synnerhet eftersom sådana beräkningar skulle underlätta bedömning av lämplig finansiering, segmentering och tidsplanering inför fortsatt förvaltning. Även om det inte når hela vägen finns i detta avsnitt en sammanställning av de underlag till värderingar och bedömningar som projektet arbetat fram.

Digitaliseringsrådet beskriver bristen på semantisk interoperabilitet som ”..data från olika källor är strukturerade på olika sätt med olika namn, klassificeringar och definitioner vilket gör att uppgifter på ett enkelt sätt inte går att dela eller bearbeta maskinellt och rutinmässigt. Denna avsaknad av standardiserad informationsklassificering och hantering skapar stora inläsningseffekter.”

²¹ Digitaliseringsrådet (2019) [En lägesbild av digital infrastruktur](#), Dnr: 19-2608

2.4.1 Tillgänglighet är ett värde i sig

En grundförutsättning för att kunna skapa tjänster utifrån den här typen av data är att de är tillgängliga. Genom att publicera de begreppsstrukturer som nämnts i uppdraget (SSYK, SUN, SNI, SeQF och ESCO) på Sveriges dataportal ökar möjligheten till vidareutnyttjande. Projektet har, som beskrevs i avsnitt 2.1, skapat möjlighet för myndigheterna att på ett strukturerat sätt koppla ihop taxonomierna med varandra. Gemensam ingång till data underlättar fortsatt samarbete för att beskriva kopplingarna mellan data. Samtidigt som det ger alla med ambition att utveckla digitala tjänster tillgång till dessa begreppsstrukturer och kopplingar mellan dem i ett format som följer SKOS²² standard för beskrivning av kunskap.

2.4.2 Kopplingar mellan begrepp kan betyda flera saker

- Förväntningar på vad en lösning kan hjälpa till med.

En lösning skulle innebära ett tillgängliggörande av samlad kunskap. En kunskap, om hur begrepp i olika etablerade system eller strukturer relaterar till varandra, som idag finns utspridd på de myndigheter som deltagit i projektet. I lösningen får begreppen, exempelvis i SUN och SSYK, kopplingar till varandra, antingen direkt eller genom en separat begreppsstruktur som förbinder dem.

För en användare som tar del av och vidareutnyttjar kopplingarna måste det vara tydligt vad en koppling består i, det vill säga *hur* begreppen relaterar till varandra. Kan inget tolkningsunderlag för varje koppling tillhandahållas måste det framgå i en generell instruktion under vilka förutsättningar de togs fram och av vilka roller. Risken att en användare (i det här fallet ett system som använder sig av lösningen) ska tolka en koppling som ett lika-med-tecken mellan två begrepp bör så långt det går minimeras för att säkerställa att den tjänst eller de slutsatser som medborgare i ”andra änden” tar del av blir så bra som möjligt.

Diskussioner om vad digitalisering kan hjälpa till med tenderar att hamna i binära resonemang om att fullständigt automatisera processer eller beslutstagande. När vi pratar om en infrastrukturlösning är det sant att vi inte helt kan styra hur konsumerande system använder sig av den. En lösning betyder dock inte att mänsklig finess, och känsla för sammanhang, eller beslutsfattande ska uteslutas eller bli sekundärt. Snarare att den information som lösningen erbjuder fungerar som ett komplement till det underlag som en slutanvändare baserar sitt beslut på, avseende exempelvis matchning eller karriärvägledning. För att uppnå god effekt med en sådan lösning är det viktigt att såväl dokumentation som interaktion med de som använder data upprätthålls över tid.

Att erbjuda bra information om hur delar av samhällslivet, med starka bilaterala kopplingar, i detta fall utbildning och arbetsmarknad, förhåller sig till varandra är i sig angeläget i och med att digitaliseringen springer framåt. Huruvida detta görs bäst med hjälp av etablerade strukturer eller om en mer detaljerad begreppsstruktur behöver skapas för att få god kvalitet i översättningarna är än så länge osagt. Med de

²²Potentiella konsekvenser från gemensamma begrepp och/eller översättningsnycklar på en finare detaljnivå fanns i intervju svaren även för andra typer av tjänster än de som namngivits i uppdraget.

[SKOS Simple Knowledge Organization System Reference \(w3.org\)](http://w3.org)

etablerade strukturerna finns fördelen att de redan är överenskomna, men det är möjligt att de kommer med en detaljering som inte är tillräckligt precis för ändamålet.

Både i dataportal- och matchningsindikator-spåret fanns aktiviteter som innehöll att manuellt para ihop och jämföra begrepp och koder mellan begreppsstrukturerna SSYK, SUN och SNI.²³ En viktig lärdom kopplat till värdet med sådant arbete är att anledningen till att person a gör en koppling inte behöver vara samma som tolkningen person b gör då de ser den koppling som person a gjort.

Som exempel är det lätt hänt att person b för en viss koppling mellan SUN och SSYK drar slutsatsen att utbildningen med SUN-kod x är en väg för att komma till kopplat yrke med SSYK-kod y. Eller från andra hållet: att det finns anledning att tro att det i yrket används kunskap som man kan nå genom utpekad utbildning. Att den tolkningen kan leda fel blir tydligt då den jämförs med statistik och beskrivningar av hur det är med yrkesspridning och utbytbarhet i samhället. Se skrivningar om det i exempelvis SCB:s *Trender och Prognoser*.²⁴

För den som gjort kopplingen kan anledningen vara att samma ord förekommer i båda de rubriker som koderna avser. Är det ett centrala och viktiga ord där rubriken beskriver stor del av underliggande innehåll kan det vara en rimlig tolkning att utbildningen då leder till yrket. Men den som läst utbildningen, förvärvat kunskaper på annat sätt eller är verksam i yrket kan troligen peka ut mycket mer än synonyma ord i kodrubrikerna för att beskriva på vilket sätt utbildningen är en lämplig eller otillräcklig väg för att komma till utpekad yrke.

2.4.3 Kartläggning av processer och användning är ett stort steg på vägen

Case-spåret har bidragit med att systematiskt identifiera vilka processer och vilka personer i dessa processer som gör översättningar mellan information om utbildningar och information om arbetslivet.

Legitima lösningar är svåra eftersom organisatoriska uppdelningar av utbildning och arbetsmarknad följer olika principer. Parternas organisationer med intresse att beskriva behovet delar sällan upp sig längs samma linjer som utbudet från utbildningsformer och Regioner. Därutöver kan en tydlig avgränsning inom arbetsmarknaden få en otydlig avgränsning inom utbildningssystemet och tvärtom, som en följd av existerande yrkesspridning och utbytbarhet.

Detaljer i de case som används kan påverka vilka processer som berörs av en livshändelse. Exempelvis vilka processer som är relevanta för att validera tidigare kunskaper för en yrkeserfaren undersköterska. Case-strukturen, med exempel på persona och deras livshändelser, är resurskrävande men gör det möjligt att synliggöra vilka frågor som behöver svar. Svar som behövs för att ge den semantiska infrastrukturen sammanhängande väldefinierade avgränsningar inom relevanta policyområden.

²³ Standard för svensk yrkesklassificering (SSYK); Svensk utbildnings nomenklatur (SUN); respektive Standard för svensk näringsgrensindelning (SNI)

²⁴ [Trender och Prognoser 2020 \(scb.se\)](https://www.scb.se/2020/09/24/trender-och-prognoser-2020)

2.4.4 Resultat från intervjuer om möjliga konsekvenser

Relativt samstämmigt uttryckte alla att det handlar om behovet att den som tar emot informationen förstår det som avsändaren avser förmedla, att detaljnivån på informationen spelar stor roll och att målet handlar om good-enough snarare än perfekt. Det bekräftar att gemensamma begrepp och/eller översättningsnycklar kan vara ett lämpligt sätt att ringa in utmaningen.

Det var talande hur mycket mer tid som intervjupersonerna la på att prata om konsekvenser av att få till kopplingar på mer detaljerad nivå jämfört med konsekvenser av att få till kopplingar mellan SCB:s taxonomier. Det stämmer också väl med arbetsgruppens erfarenheter att det är viktigt att inte bara se till kopplingar mellan existerande begreppsstrukturer utan även inkludera översättning av information om detaljer. Arbetsgruppen återkom ständigt till parametrarna detaljnivå och progression i samtal om projektets målsättning.

För konsekvenser av att ta fram standardiserade **kopplingar mellan existerande klassifikationer** (SSYK, SUN och SNI) beskriver de här fyra citaten väl bredden i svaren:

- *behöver inte vidareutvecklas då den struktur som finns i dagsläget är fullt tillräcklig.*
- *idag finns ingen konsensus och varje aktör skapar sin mappningsstruktur.*
- *man kan koppla vad analytiker säger till våra prognoser, ex. antagning.se eller Ladok.*
- *aldrig funnit de användningsbara. Alltid så mycket tolkningar bakom kulisserna. Har aldrig litat på det där. Vad gör din drifttekniker vs. driftingenjör? Olika koder gör samma sak, eller tvärtom, samma kod men gör olika saker.*

Det framgår av citaten att det blir missvisande att dra genomsnittliga slutsatser. Konsekvenserna blir uppenbart olika för olika processer och sammanhang. Det är en anledning till att det är viktigt att bena ut, vilka processer och vilken koppling mellan informationsmängder olika användare är intresserade av, innan det går att göra uppskattningar av värdet med olika lösningar.

Det kom betydligt fler svar angående konsekvenser av att ta fram gemensamma begrepp och/eller översättningsnycklar mellan information på finare detaljnivå. Det gör det meningsfullt att dela upp svaren och koppla dem till det i uppdraget formulerade behovet att underlätta utvecklingen av tjänster för validering, matchning och vägledning:

Vägledning är troligen det tydligaste och mest direkta användningsområdet. Det framgår av dessa tre citat:

- *En mer professionell vägledning. Vägledaren har fortfarande kompetens att hålla i mötet, men innehållet blir bättre. Vi kan idag inte få fram snabb information om vilka kompetenser som krävs för en specifik yrkesroll. Denna datamängd är omöjlig att ha i huvudet.*

- *Större vikt, branschöverskridande, att man pratar samma terminologi. Inom en yrkeskategori har man det inarbetat.*
- *Bättre överföringsbarhet, 20 år på en återvinningscentral, om man spjälkar upp vad man faktiskt gjort. Färdigheter man kan ta med sig till annan yrkesroll.*

De här fyra citaten återspeglar svar om konsekvenser inom **validering**. Skillnader i svaren beror delvis på om det gäller fördjupad kartläggning, validering mot utbildning eller validering mot branschkrav:

- *Ungefär samma process. Man måste ändå gå igenom det. Nu har vi checklistor och dokumentationer. Helt annan systematik.*
- *Nyttan man vill åt - att folk kan värdera den kunskap man byggt. Jag har gått kurs och har kunskap från den här nivån.*
- *Med yrkesprofiler²⁵ går det att identifiera när olika moduler innehåller samma kunskaper. Ex. dricksvatten och avloppsvatten. "Jag har redan svarat på det här i förra modulen."*
- *Det är svårt att få med sig den "spårbarheten" när man flyttar till annat bolag. Idag är mycket vad man skriver på CV och "word by mouth".*

I **matchningen** kan gemensamma översättningsnycklar avhjälpa brister, det framgår av de följande fyra citaten. Fast det är inte självklart hur och på vilket sätt:

- *Finns en poäng med att ha samma terminologi. Det finns olika benämningar på relativt samma sak. Du kan säga så, men det är inte så man säger inom branschen.*
- *Bristituationer löser man genom en dialog med personer i en intervju. Det är inte att man inte förstår. Det går inte att läsa sig till, det krävs en dialog i alla fall.*
- *Gemensamt språk kan hjälpa men det stora arbetet handlar om att hjälpa sökande och arbetsgivare att beskriva vad de kan och vad de vill ha.*
- *Det är inte självklart att utomstående förstår. Exempelvis: Ah, du har erfarenhet från tunga fordon. Men, hur många gånger tror du att bussförare backar?*

Potentiella konsekvenser från gemensamma begrepp och/eller översättningsnycklar på en finare detaljnivå fanns i intervju svaren även för andra typer av tjänster än de som namngivits i uppdraget.

En av de största vinsterna man erfarit i branscher som jobbat med yrkesprofiler är att människor **motiveras** av att få tydliggjort vad de kan och hur de kan utveckla sin yrkesskicklighet. Det kan även ge yrket status och medarbetare bättre självkänsla.

²⁵ Branschgemensamt beskrivna behov på specificerad detaljnivå. [Yrkesprofilering - Nordiskt Valideringsforum](#)

Vid **utformning av utbildning** kan standardisering bidra till tydligare beskrivningar av behov och det tillkommer bättre kvalitetssäkring och spårbarhet.

Introduktion och praktik uppges kunna kortas betydligt genom att ge stöd till handledarrollen, underlätta validering och att det mer systematiskt går att bocka av ett kriterium i taget. En standardiserad måttstock för on-boarding kan även ge nya möjligheter att bredda rekryteringsbasen.

För **kompetens- och verksamhetsutveckling** kan det skapa tydliga utvecklingsplaner som gör det lättare att få till den fortbildning som behövs för att inte bli övertalig vid verksamhetsförändringar. Det finns redan indikatorer om att kompetensutveckling matchar verksamhetsutvecklingen bättre när det finns en strukturerad beskrivning av kompetens behoven.

Vid **upphandling** kan det finnas nya möjligheter att ställa krav, så att anbud kan sorteras utifrån kompetens hos medarbetare, och att beskrivningen av kompetensen knyts till en standardiserad yrkesprofil snarare än ett specifikt certifikat.

Om något av detta förbättrar matchningen generellt, eller konkurrenskraften i många företag, så skulle det kunna ge utslag även på **makronivå**. Men det är långsökt att ta fasta på det. Sådana konsekvenser ligger långt bort i kedjan och det finns många om och men på vägen dit.

2.4.5 Risker och kostnader

Av intervjuerna framgår också att det finns en balansakt att göra i arbetet med översättningsnycklar. Det är inte bara potentiella vinster utan också kostnaderna för utveckling och förvaltning som är större när kopplingar görs på finare detaljnivå.

Det är viktigt att ha med sig att kvaliteten på sammankopplandet är beroende av kvaliteten på beskrivningarna av de begrepp som kopplas samman. Begrepp som inte definieras mer än med ett namn kan vara svåra att förstå i förhållande till avgränsning mot andra begrepp och vilka delmängder de innehåller.

En rikare beskrivning, mer information på mer detaljerad nivå, leder till mer underbyggda beslut om kopplingar och färre gissningar, samtidigt som den också ställer högre krav på förvaltning i och med att den totala informationsmängden blir större. Ett exempel från Arbetsförmedlingen är den nivå av begrepp, ”yrkesbenämningar”, som kopplas till begrepp i SSYK-strukturens lägsta nivå (4-ställig). Till skillnad från SSYK-strukturen uppdateras yrkesbenämningarna mer löpande. Däremot är de inte definierade med hjälp av beskrivningar, såsom begreppen i SSYK-strukturen är, utan betydelsen av yrkesbenämningarna får istället härledas utifrån deras kopplingar mot andra begrepp, exempelvis de som finns i SSYK-strukturen. Den generella förståelsen för yrkesbenämningarna riskerar därför att bli låg om inte ingående kännedom om begreppsmängdens karaktär och begreppens olika sammanhang finns, vilket påverkar förmågan att relatera dem mot begrepp i *andra* strukturer.

Vad exemplet ovan försöker att illustrera är att sammankopplandet av etablerade strukturer kan vara mindre problematiskt än att utgå ifrån existerande information

på mer detaljerad nivå, inte bara för att ingående begrepp är färre till antalet eller för att strukturerna redan är överenskomna, utan också för att informationen på mer detaljerad nivå av olika skäl, oftast kostnadsmässiga, brister i hur väl den är kvalitetssäkrad.

Samtidigt skulle det kunna vara kommunikation och införande som är den allt överskuggande kostnaden och i så fall spelar valet mellan grövre och finare detaljnivå mindre roll för totalkostnaden.

Med det sagt är det inte självklart att det är just Arbetsförmedlingens yrkesbenämningar som svarar mot behovet av mer detaljerad information och därmed skulle vara aktuella att koppla ihop med andra strukturer, men det är ett faktum att begreppsmängder som tagits fram utanför de stora processerna av att etablera strukturer som SSYK, SUN och SNI inte alltid har haft samma dokumentationskrav och att mer ingående kännedom om begreppens sammanhang därmed inte följer med.

Den mänskliga förmågan att klassificera och koppla samman begrepp har i sin tur påverkan på maskinella metoder (exempelvis inom Machine Learning), som producerar bättre resultatet om det mänskliga facit har god kvalitet och textmängden är rik. Att den kvaliteten säkras av domänkunniga blir i sammanhanget viktigt och har direkt bäring på hur bra lösningar, de användare som vidareutnyttjar kopplingarna, kan bygga.

Det förekommer också en oro att fler standardiserade beskrivningar av kunnande kommer att leda till mindre flexibilitet inom KLL-domänen. Som vid all standardisering är det inte kostnaden utan den relativa kostnaden för speciellösningar som ökar. Mot den minskade flexibiliteten står också möjligheter till ökad rörlighet och nya kombinationsmöjligheter. Liksom för legobitar går det att öka kombinationsmöjligheterna genom att minska storleken (i.e. ökad detaljnivå) på de bitar som standardiseras.

För det livslånga lärandet finns också ett bredare syfte, mål som ligger bredvid eller utanför det arbetslivslånga lärandet. Stärkta kopplingar mellan information om utbildningar och information om arbetsmarknader bidrar inte på något självklart sätt till mål om bildning, vidare studier eller elevers utveckling till ansvars-kännande människor och medborgare. Å andra sidan tas inte heller något bort från dessa mål. Möjligen kan det bidra till att tydliggöra distinktioner mellan dessa olika mål, men det är inget som projektet har tittat närmare på.

Slutligen: För att få till lämpliga underlag och värderingar av att skapa gemensamma begrepp och/eller översättningsnycklar är det centralt att se det som den infrastruktursatsning det är. På samma sätt som för annan infrastruktur är en av poängerna att användningen är spridd över flera olika sammanhang, och liksom i andra sammanhang finns många steg, ofta med långa ledtider, från planering av produktion av infrastruktur till dess att en satsning kan nå utlovade värden för arbetsgivare, arbetstagare och individer vid vägskälen i sitt livslånga lärande.

2.4.6 Huvudsakliga lärdomar

- För att kunna använda gjorda kopplingar på ett värdefullt sätt är det viktigt att beskriva och vara transparent med varför och hur en koppling gjorts. Det gäller oavsett hur väl bedömningen stämmer med verkliga jobbmöjligheter.
- Utmaningen med att hitta gemensamma begrepp och/eller översättningsnycklar är konkret och hanterbar när den formulerats på tillräcklig detaljnivå.
- En översättningsnyckel behöver dock inte vara möjlig att ta fram på detaljnivå även om beskrivningar på detaljnivå behövs för att göra kopplingen på övergripande nivå. Det är exempelvis en viktig skillnad mellan översättningar mellan, å ena sidan arbetsuppgifter och läranderesultat och å andra sidan läranderesultat och olika formuleringar av utbildningsmål.
- En rikare beskrivning, mer information på mer detaljerad nivå, leder till mer underbyggda beslut om kopplingar och färre gissningar, samtidigt som den också ställer högre krav på förvaltning i och med att den totala informationsmängden blir större.
- För att få kvalitet och legitimitet krävs ofta involvering av intressenter från de specifika delar av arbetsmarknaden och utbildningssystemet som berörs. Involveringen är inte sällan kopplat till krav på kontroll över att kontexten följer med och att informationen inte ändras på vägen.
- Potentiella konsekvenser av att ta fram gemensamma begrepp eller översättningsnycklar varierar mellan olika verksamheter och processer men från intervjuerna framgår att flera ser betydligt större värde av kopplingar på finare detaljnivå än mellan existerande begreppsstrukturer.
- Potentiella konsekvenser från gemensamma begrepp och/eller översättningsnycklar på en finare detaljnivå fanns i intervjusvaren även för andra typer av tjänster och verksamheter än de som namngivits i uppdraget.
- Nyttor och kostnader med olika varianter av gemensamma begrepp eller översättningsnycklar behöver klargöras ytterligare. Det är en bidragande orsak till förslaget om en fullföljd samhällsekonomisk analys (se avsnitt 3.2.1).

3 Vad får det för följder i fortsatt förvaltning?

Då gemensamma begrepp och översättningsnycklar är infrastruktur och inte löser de specifika utmaningarna i respektive verksamhet/projekt riskerar investeringar att bli underdimensionerade. Samma utmaningar tenderar att lyftas i alla verksamheter där översättningsnycklar efterfrågas, men ingen enskild aktör kan bära kostnaderna för att ta fram det som behövs.

Hur ser samhället till att data som tas fram eller bör tas fram, i ett av KLL-domänens politikområden, förstås på ett sätt så att det möter behov där dessa data används, inom ett annat av KLL-domänens politikområden?

Det är vad det här kapitlet handlar om, fast på konkret nivå. Avsnitten är formulerade i termer av projekt och verksamheter som bedömts som lämpliga nästa steg för att utveckla gemensamma begrepp eller översättningsnycklar.

3.1 Motivering till projekt som kommer drivas under 2023

3.1.1 WikiPOC – en gemensam källa för översättningsnycklar

Under 2023 kommer Arbetsförmedlingen driva det Vinnovafinansierade projektet *Wiki-POC – en gemensam källa för översättningsnycklar mellan arbetsuppgifter och läranderesultat*. Projektet utvecklar ett koncept för ett nytt IT-verktyg: ett stöd för gemensam insamling av, och en gemensam källa för, tidigare svårtillgänglig och/eller ej digitaliserad information.

Syftet är att därigenom utveckla förutsättningarna att skapa gemensamma begrepp eller översättningsnycklar på en mer detaljerad nivå än existerande officiella klassificeringar. En detaljnivå som behövs vid såväl validering och matchning som vägledning och utbildningsutformning. Den ökade detaljnivån öppnar även för möjligheter att utveckla mer sammanhängande beskrivningar av resultat från skilda verksamheter inom KLL-domänen.

Semantikprojektet har bidragit till förutsättningarna för detta projekt genom att förtydliga en möjlighet och en avgränsning. Möjligheten kommer från lärdomen att utmaningar med översättningsnycklar är konkreta och hanterbara så fort vi kommer ner på tillräcklig detaljnivå. Avgränsningen kommer från lärdomen att det krävs flera översättningar, än denna mellan arbetsuppgifter och läranderesultat, för att förmedla information från arbetslivet till utbildningsvärlden.

Avgränsningen som finns i projektnamnet, *översättningsnycklar mellan arbetsuppgifter och läranderesultat*, ökar förutsättningarna för att projektet ska nå sitt syfte. Dels genom att arbetsuppgifter och läranderesultat är begrepp som befinner sig på en detaljnivå där kopplingar mellan informationsmängderna är hanterbara. Dels genom att avgränsa bort de ytterligare utmaningar som behöver hanteras för att få till översättningsnycklar mellan läranderesultat och olika formuleringar av utbildningsmål.

IT-verktyget siktar på att erbjuda en lösning som samtidigt behåller ett distribuerat ägande, för olika delar av arbetsmarknaden, och skapar en gemensam källa för de användare som i andra änden har behov av de kvalitetssäkrade översättningsnycklar som behövs i olika processer.

Verktyget kan på så sätt bidra till att skala upp en utveckling av processer för att sammanställa yrkesprofiler. Ett arbete som redan påbörjats i flera partsgemensamma organisationer – att standardisera i återanvändbara moduler vad man behöver behärska för att utföra verksamhetskritiska arbetsuppgifter.

3.1.2 SCB:s utvecklingsprojekt om klassifikationerna SNI, SUN och SSYK

Under 2023 kommer SCB driva de två Vinnova-finansierade projekten ”Förbättra tillförlitligheten och införa modellbaserade arbetssätt i SSYK- och SUN-klassifikationerna på SCB” och ”Undersöka samband mellan SUN, SSYK och SNI maskinellt”.

Gemensamt för de båda projekten är att man vill förbättra och utveckla klassifikationer inom arbetsmarknads- och utbildningsområdena genom att i större utsträckning använda maskinella metoder och AI/ML vid klassifikationsutförandet.

Projekten kan med fördel nyttja insikter baserade på processdata från arbetet med matchningsindikatorer som input i projekten. Bland annat kan jämförelser göras mellan de utpekade bedömda kodkombinationerna på mer aggregerade nivåer med maskinella korrelationsmetoder baserade på en objektiv vetenskaplig grund. Även de faktiska populationsandelarna kan jämföras med de maskinella korrelationsmått.

Projekten syftar till att kartlägga klassifikationerna ur ett användarperspektiv samt att synliggöra och förbättra kvaliteten i dessa klassifikationer. Tanken är dessutom att arbetsprocessen med framställande av årsuppsättningar av Utbildningsregistret (UREG) och Yrkesregistret (YREG) ska kunna gå snabbare och smidigare med nya mer automatiserade och säkerställda lösningar, med positiv påverkan på kvaliteten i klassifikationerna.

Något annat som är centralt i projekten är att skapa ytterligare insikter om klassifikationerna och hur dessa är kopplade till varandra. I och med att användandet av de tre klassifikationerna ökar, med en allt större målgrupp för klassifikationerna, bör kunskapsnivån om hur klassifikationerna är relaterade till varandra bli allt viktigare, och en höjd kunskapsnivå kommer höja nyttan med klassifikationerna.

Specifikt maskinellt framtagna, uppföljningsbara mått på jämförbarhet/samband mellan klassifikationerna samt analyser av dessa kommer möjliggöra denna kunskapshöjning.

3.1.3 Begreppslistor kommer i skarpt läge under nästa år

Dataportalspåret har skapat förutsättningar för fortsatt arbete att dels publicera färdigt de datakällor inom begreppsområdet som identifierats som prioriterade och dels möjlighet att publicera relationer/översättningsnycklar mellan terminologierna baserat på SKOS-standard. Det arbetet kommer att bedrivas dels i några projekt inom semantikområdet, i Sänkta trösklar-projektet inom uppdragets delområde Ökad datadelning för stärkt innovationskraft och dels i sekretariatets arbete med att ta fram en framtida förvaltning.

För en framtida förvaltning behöver vi se vilka begrepp och begreppstrukturer som ger mest mervärde att relatera och vem/vilka som ansvarar för dessa, inklusive att formulera förslag på lämpliga juridiska och organisatoriska former för förvaltningen av de relationer där kvalitetskraven på relationerna bäst uppfylls av ett gemensamt eller distribuerat ansvar.

3.2 Förslag på utveckling som bör få förankring under 2023

3.2.1 Fullfölj den samhällsekonomiska analys som påbörjats

Avsnitt 2.4.4 beskrev ett antal typiska svar från intervjuer om möjliga konsekvenser av att ta fram gemensamma begrepp eller översättningsnycklar. Genomförda intervjuer är början till en grund för att beräkna konsekvenser av den infrastrukturutveckling som semantikprojektet utrett.

Urvalet av intervjupersoner gjordes för att få en bred representation. Det finns dock en tonvikt på kompetensförsörjning, så det behöver kompletteras med intervjuer med mer tonvikt på livslångt lärande för att fånga potentiella konsekvenser för hela domänen. Efter en utvidgad mängd konsekvensintervjuer vore det lämpligt att sedan göra en samhällsekonomisk nyttoanalys av värdet med att producera och/eller tillgängliggöra gemensamma begrepp eller översättningsnycklar inom KLL-domänen.

Inspiration till metod, process och rapport för att ta fram en sådan analys kan med fördel hämtas från Lantmäteriets delrapport *Tillgängliggörande av särskilt värdefulla datamängder* inklusive dess bilaga 2, *Värdet av öppna data – Samhällsekonomisk nyttoanalys av värdefulla datamängder*.^{26, 27}

3.2.2 Synk med EU initiativ för Data Space for Skills

En viktig inriktning för kommande arbete inom uppdraget eller inom kommande samverkansorganisation är stöd för skills data space. Det vill säga att lösningar för att hantera den nationella semantiska interoperabiliteten inom KLL-domänen tar ansvar också för att datainfrastrukturen synkar med hantering på EU-nivå av inte bara offentligt producerad utan även privat producerad information inom utbildning och arbetsmarknad.

3.2.3 Mäta kvaliteten på framtagna översättningsnycklar

En viktig del av det fortsatta arbetet med att ta fram översättningsnycklar, oavsett om det görs för hand eller med hjälp av algoritmer, är att kunna mäta kvaliteten på dessa nycklar.

Som ett komplement till, eller möjligen som en del av SCB:s utvecklingsprojekt om klassifikationer, skulle den datamängd som Matchningsindikatorspåret (avsnitt 2.4.3) resulterade i kunna användas för att jämföra kopplingar mellan begrepp framtagna för hand och kopplingar mellan begrepp framtagna av en algoritm. Förutsatt att man kan enas om kriterier för vad god kvalitet på kopplingarna innebär.

Därigenom kan man beräkna ett mått på hur bra en algoritm kan ersätta manuellt arbete. Det skulle innebära en ny typ av underlag då användare och producenter i olika delar av KLL-domänen väljer metod eller algoritm för att ta fram nya översättningsnycklar. Det skulle också vara intressant att undersöka om befintliga system som BERTMap eller Europeiska kommissionens BERT-modell kan användas för att ta fram nya översättningsnycklar inom KLL.

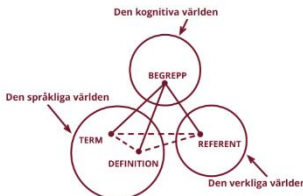
²⁶ [lm2019_007157-delrapport---tillgangliggorande-av-sarskilt-vardefulla-datamangder.pdf \(lantmateriet.se\)](#)

²⁷ [Värdet av öppna data \(lantmateriet.se\)](#)

Bilaga 1 – Begreppslista och bilder

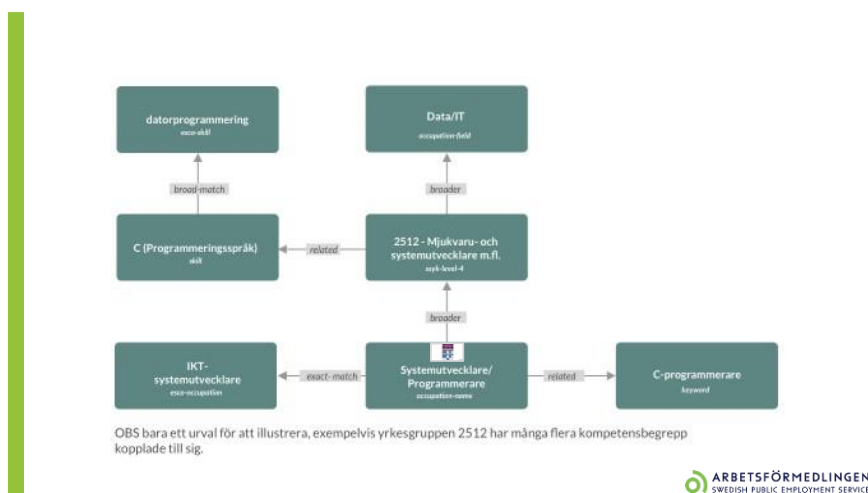
I det här dokumentet har följande centrala begrepp använts med dessa definitioner.

Begrepp om berörda verksamheter		
<i>KLL-domänen</i>	Ett sätt att referera till alla verksamheter som är del av samhällets insatser och stöd till Kompetensförsörjning och Livslångt Lärande.	Egen
<i>Interoperabilitet</i>	Interoperabilitet är förmågan att fungera tillsammans och kunna kommunicera med varandra. Vanligtvis lyfts teknisk, juridisk, organisatorisk och semantisk interoperabilitet.	Begrepp - Digitaliseringsrådet
<i>Persona</i>	En fiktiv karaktär som beskriver en målgrupp.	Egen
<i>Läranderesultat / Resultat av lärande</i>	Uppgifter som rör vad en individ vet, förstår och kan göra efter att ha avslutat en inlärningsprocess, vilket uttrycks i kunskaper, färdigheter, ansvar och självständighet. Resultat av lärande har två sidor som lätt förväxlas då båda kan benämnas som en dokumentation: Å ena sidan dokumentation av en individs lärande (eng. achieved learning outcome). Å andra sidan dokumentation av kravställningen för ett lärande (eng. intended learning outcome).	EQF-rekommendationen Egen
<i>Dokumentation av lärande</i>	Dokumentation av bedömningen av en individs lärande i relation till en kravställning för ett lärande.	Egen
<i>Utbildningsmål</i>	Har en bred betydelse inom skolväsendet, dels som den enskilde elevens mål med utbildningen, men också formulerat som utbildningens mål, t.ex. examensmål.	Egen
<i>Yrkesprofil</i>	Beskriver vad man behöver kunna för att arbeta inom en preciserad del av arbetsmarknaden. En yrkesprofil kan med fördel, men behöver inte, formuleras i det format som preciserats som läranderesultat enligt EQF-rekommendationen.	Nordiskt Valideringsforum

Begrepp om begreppsstrukturer		
<i>Begrepp</i>	<p>Det abstrakta innehållet i en språklig term, i kontrast till termen själv och de konkreta eller abstrakta objekt den syftar på (refererar till).</p> 	<p>Wikipedia</p> <p>(bild) Terminologifrämjandet</p>
<i>Begreppsstruktur</i>	<p>En förteckning över det system av termer och beteckningar som används inom ett fackområde.</p> <p>I rapporten använt som ett paraplybegrepp för att kunna referera till alla olika typer av begreppsstrukturer (i.e. från vokabulär till taxonomi).</p>	<p>Wikipedia</p> <p>Egen</p>
<i>Taxonomi</i>	<p>En taxonomi är en hierarkiskt uppbyggd förteckning eller ett system för kategorier eller grupperingar av objekt, där varje objekt eller observation endast kan förekomma i en av dessa kategorier.</p>	<p>SCB</p>
<i>Klassifikation</i>	<p>Ett bredare begrepp än taxonomi då det innefattar begreppsstrukturer som inte är hierarkiskt uppbyggda.</p>	
<i>Terminologi</i>	<p>En namngiven samling begrepp som har en gemensam tematik och förvaltas gemensamt.</p> <p>Ett bredare begrepp än klassifikation då det innefattar begreppsstrukturer där kategorier eller grupperingar av objekt inte är ömsesidigt uteslutande. I texten används 'terminologi' ibland ännu bredare, som synonym till nomenklatur.</p>	<p>Dokumentation – Sveriges dataportal</p> <p>Egen</p>
<i>Nomenklatur</i>	<p>Ett bredare begrepp än terminologi, då det innefattar begreppsstrukturer där det saknas beskrivningar av ingående begrepp. Samtidigt ett snävare begrepp än vokabulär, då vokabulär inte behöver ha en gemensam förvaltning eller etablerade namngivningsregler.</p>	<p>Egen utveckling från terminologifrämjandets ordlista</p>

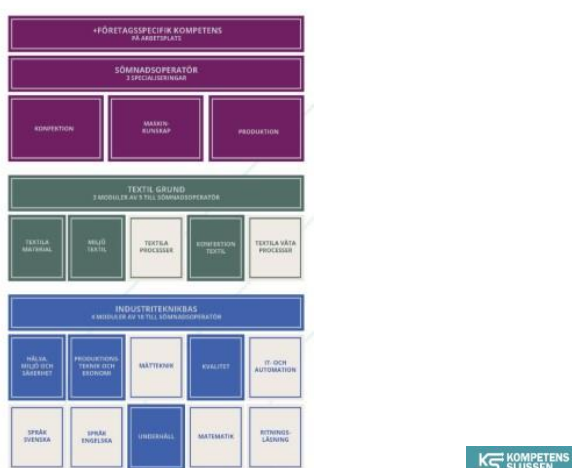
Begrepp om relationer mellan data		
<i>Länkade data</i>	<p>Länkade data handlar om att komplettera den existerande webben av dokument med en webb av data. Konkret uppfyller länkade data två principer:</p> <ul style="list-style-type: none"> - adresserbarhet: ge varje ting en URI (en webbadress) som gör att påståenden kan hämtas via HTTP. - enhetlig informationsmodell: används språket RDF för att uttrycka påståenden om ting <p>Den fjärde nivån i W3C:s standard för öppen data.</p> <p>I rapporten används länkade data ibland med syfte på både länkade och sammanlänkade data.</p>	<p>Egen utveckling från Vitbok länkade öppna data</p> <p>5-star Open Data (5stardata.info)</p>
<i>Sammanlänkade data</i>	<p>I rapporten används det om länkade data som även uppfyller en tredje princip.</p> <ul style="list-style-type: none"> - länka ihop: förbind ting med varandra genom relationer av olika slag, gärna mellan olika datakällor. <p>Den femte nivån i W3C:s standard för öppen data.</p>	<p>Egen utveckling från Vitbok länkade öppna data</p> <p>5-star Open Data (5stardata.info)</p>
<i>Översättningsnyckel</i>	<p>Data eller information som <u>beskriver</u> en koppling/relation mellan ett begrepp i ett sammanhang och ett annat begrepp i ett annat sammanhang.</p>	Egen
<i>Semantik</i>	<p>Studiet av språklig betydelse eller studiet av teckensystems innebörd och tolkning.</p> <p>Syftar här på de former som text använder för att förmedla innebörden av information från ett sammanhang till ett annat.</p>	<p>Wikipedia</p> <p>Egen</p>
<i>Översättning</i>	<p>överföring av ett budskap från ett språk till ett annat,</p> <p>Till skillnad från en <i>översättningsnyckel</i> mellan enskilda <i>begrepp</i> överför en översättning så mycket som möjligt av den ursprungliga förståelsen av innehållet. Med andra ord inkluderar den, givet förutsättningarna, även överföring av form och kontext.</p>	<p>NE</p> <p>Egen</p>

Övergripande begrepp - olika perspektiv, samma referenter Utveckling från tabell i <i>Validering – för kompetensförsörjning och livslångt lärande</i> (SOU 2019:69)						
Kunnande (det samlade lärandet) Det en individ..	Fyra former av kunskap	Förväntade utbildningsresultat	Läranderesultat	Kompetenser	"Competencies"	Kunskap
..vet	fakta	kunskap och förståelse	kunskaper	kunskaper	kunskaper	
..förstår	förståelse					episteme (förståelse)
..kan göra	färdighet	färdighet och förmåga	färdigheter	färdigheter	kognitiva färdigheter, praktiska färdigheter	techne (färdigheter)
	förtrogenhet	värderingsförmåga och förhållningssätt	ansvar och självständighet	attityder	attityder, känslor, värden och etik, motivation	fronesis (omdöme)
SOU 2019:69	Skolväsendet / Läroplanens fyra F	Högskolan Bologna-processen	Referensramar för kvalifikationer EQF/SeQF	EU:s nyckelkompetenser	OECD	Aristoteles



Figur 4: Exempel på kopplingar mellan begrepp i Arbetsförmedlingens Taxonomy.

Yrkesprofil Sömnadsoperatör



Figur 5: Exempel från Proteko på återanvändning av moduler till en yrkesprofil.

MÅLERI- BRANSCHENS YRKESNÄMND

Kvalifikation Byggnadsmälare UTKAST

Validering Väst

2. Tak&Vägar (invärdigt)

Syfte: Byggnadsmälaren ska efter godkänd dekkvalifikation, självständigt och i samverkan med andra kunna på ett fackmannamässigt sätt kunna utföra underbehandling samt repetering, värgmållning och målning av tak och väggar, på både nytt och tidigare behandlad underlag.

Delområde	Kunskaper	Färdigheter	Ansvar och självständighet
Underbehandling	Byggnadsmälaren ska kunna: <ul style="list-style-type: none"> redigera förodda ytor till tekniska tillståndet för vidare målning som avses vid underbehandling av tak&väggar inomhus redigera förodda ytor till tekniska tillståndet för vidare målning av tak&väggar utomhus redigera förodda ytor till tekniska tillståndet för vidare målning av tak&väggar i svåråtkomliga ytor redigera förodda ytor till tekniska tillståndet för vidare målning av tak&väggar i svåråtkomliga ytor 	Byggnadsmälaren ska kunna: <ul style="list-style-type: none"> välja rätt verktyg, metod och arbetsförfarande vid målning och utföra underbehandling av tak&väggar inomhus i enlighet med gällande tekniska och kvalitetskrav välja rätt verktyg, metod och arbetsförfarande vid målning och utföra underbehandling av tak&väggar utomhus i enlighet med gällande tekniska och kvalitetskrav 	Byggnadsmälaren ska kunna: <ul style="list-style-type: none"> skötas och samverka med andra i ett team för att utföra tekniska och kvalitetskrav på ett fackmannamässigt sätt skötas och samverka med andra i ett team för att utföra tekniska och kvalitetskrav på ett fackmannamässigt sätt
Färdigbehandling	Byggnadsmälaren ska kunna: <ul style="list-style-type: none"> redigera förodda ytor till tekniska tillståndet för vidare målning som avses vid färdigbehandling av tak&väggar inomhus redigera förodda ytor till tekniska tillståndet för vidare målning som avses vid färdigbehandling av tak&väggar utomhus redigera förodda ytor till tekniska tillståndet för vidare målning som avses vid färdigbehandling av tak&väggar i svåråtkomliga ytor redigera förodda ytor till tekniska tillståndet för vidare målning som avses vid färdigbehandling av tak&väggar i svåråtkomliga ytor 	Byggnadsmälaren ska kunna: <ul style="list-style-type: none"> välja rätt verktyg, metod och arbetsförfarande vid målning och utföra färdigbehandling av tak&väggar inomhus i enlighet med gällande tekniska och kvalitetskrav välja rätt verktyg, metod och arbetsförfarande vid målning och utföra färdigbehandling av tak&väggar utomhus i enlighet med gällande tekniska och kvalitetskrav 	Byggnadsmälaren ska kunna: <ul style="list-style-type: none"> skötas och samverka med andra i ett team för att utföra tekniska och kvalitetskrav på ett fackmannamässigt sätt skötas och samverka med andra i ett team för att utföra tekniska och kvalitetskrav på ett fackmannamässigt sätt

Figur 6: Exempel från Måleribranschen på den mest detaljerade nivån (strecksatser) i beskrivningen av kraven i en yrkesprofil.

Poängplan Fordons- och transportprogrammet

	Gymnasie gemensamma kurser	Program gemensamma kurser	Inriktning	Individuellt va gymnasiearbete	Program fördjupning
Karosseri och lackering	600 p	400 p	400 p	300 p	800 p
Lastbil och mobila maskiner	600 p	400 p	400 p	300 p	800 p
Personbil	600 p	400 p	400 p	300 p	800 p
Transport	600 p	400 p	500 p	300 p	700 p

Figur 7: Exempel på de olika kurser som ingår i Fordons- och transportprogrammet

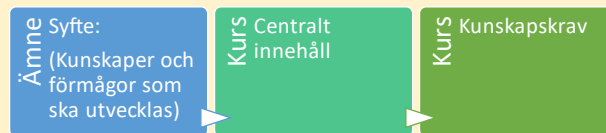
Yrkesutgång/Yrkespaket Lastbilmekaniker

Gymnasie gemensamma kurser 600 p	Program gemensamma kurser 400 p	Inriktningskurser Lastbil och mobila maskiner 400 p	Individuellt val, gymnasiearbete 300 p	Programfördjupningskurser 800 p
Svenska 1 Engelska 5 Matematik 1 Naturkunskap 1a1 Historia 1a1 Religionskunskap 1 Samhällskunskap 1a1 Idrott och hälsa 1	Fordons- och transportbranschens villkor och arbetsområden Fordonsteknik – introduktion	Lastbilar och mobila maskiner – Basteknik Lastbilar och mobila maskiner – service och underhåll 1 & 2 Lastbilar och mobila maskiner – verkstad och elteknik	Individuellt val Gymnasiearbete	Lastbilar och mobila maskiner – förbränningsmotorer Lastbilar – kraftöverföring Lastbilar – chassi och bromsar Lastbilar – system- och diagnosteknik 1 & 2 El- och hybridfordon 1

Figur 8: Exempel på en möjlig kombination av kurser i ett yrkespaket för komvux.

Gymnasieskolans ämnen/kurser

- De gymnasiala (del-) kvalifikationerna är i första hand beskrivna genom utbildningen som ska leda till att eleverna utvecklar kunskaper, färdigheter och förmågor, och i andra hand genom uppnådda läranderesultat som går att härleda ur enskilda kursers centrala innehåll kombinerat med kunskapskraven.



Figur 9: Kombinationen av delar i skolväsendets styrdokument ur vilka en erfaren lärare skulle kunna härleda något som motsvarar läranderesultat.

Bilaga 2 – Kartläggning med hjälp av case

Arbetsgruppen kunde genomlysna ett case hela vägen och påbörjade arbete med ytterligare ett. Det första bestod av en person som hoppat av el- och energiprogrammet och som nu vill ta sig via en utbildning till en anställning som bussmekaniker. Där både bussmekanikerutbildningens inriktning och arbetsgivarens kompetensförsörjning är i behov justeringar med anledning av övergång från fossildrivna bussar till en el-driven fordonsflotta. Exemplet var bra på att visa utmaningen med att jämföra brottstycken av kunnande som finns, och behöver fortsätta finnas, i skilda format.

Läranderesultat i kursen Bussteknik

Undervisningen i kursen ska behandla följande centrala innehåll	Kunskapskrav	Transportföretagens beskrivning delkvalifikationer*
<ul style="list-style-type: none"> Risker i samband med service och reparationsarbete på eller i närheten av högvoltsystem och gassystem avsedda för framdrivning. 	Eleven arbetar på ett säkert, miljö- och kvalitetsmässigt sätt utifrån lagar och andra bestämmelser samt fordons- eller utrustningstillverkarens krav.	Bussmekanikern ska kunna: <ul style="list-style-type: none"> Redogöra för ellärans grunder samt grundprincipen för bussens elektriska system. Demontera och montera elkomponenter och kablage samt efter reparation och montering utföra kontroll av systemets funktion.

* ur delkvalifikationen *Elsystem och instrument*

Figur 10: Ett exempel som användes i case-spåret för att åskådliggöra utmaningen med att hitta översättningsnycklar mellan läranderesultat och skolväsendets styrdokument.

Ett andra case bestod av en nyanländ arbetssökande med erfarenhet som sjuksköterska i annat land och en erfaren undersköterska som möter en arbetsgivares behov av kompetensförsörjning. Arbetsgivaren har i exemplet behov av både en undersköterska och en specialistundersköterska. Exemplet var bra på att fånga hur mycket hemvisten för de olika stödprocesserna skiftade beroende på detaljer i de arbetssökandes bakgrund.

Resultaten växte fram som en succesiv kartläggning av svar på: Var och hur möts begreppsstrukturer i samband med tänkta livshändelser?

I arbetet blev det uppenbart att frågan behöver delas upp för att kartläggningen skulle bli bra. Frågan om *var* begreppsstrukturer möts bryts med fördel i två delar:

- Vilka institutioner och processer ger stöd i den livshändelse som identifierats?
- Vilka begreppsstrukturer och informationsmängder används i de processerna?

Likaså frågan om *hur* begreppsstrukturer möts bör delas i två delar:

- Vem eller vilka, i de processerna, gör översättningen eller tolkar vad som är en lämplig koppling mellan olika begreppsstrukturer/informationsmängder?
- Vilket IT-stöd, har den som gör översättningen, tillgång till?

Figur 11 tjänar mest som ett kikhål på den typ av tabell med beskrivningar som dokumentationen från arbetet med case resulterat i.²⁸

Var och hur möts begreppsstrukturer i samband med livshändelsen?						
Livshändelse	+ Utmaningar att undersöka närmare längs datavärdekedjan: (data om utbildning/arbete) till önskemål - via datamellanhänder - från dataproducenter.					
	Föräldraledig nyanländ sjuksköterska registrerad som hemma.					
		Migrationsverket	Komvux	Komvuxläraren mot ett yrkespaket utifrån Socialstyrelsens allmänna råd. (Socialstyrelsen) Utfärdande av legitimation för legyrken.	Socialstyrelsen	Socialstyrelsen
Hur 2) Vem gör översättning mellan begreppsstrukturer/informationsmängder?			SYV kommer att ha ett ansvar för att - samla in dokument och få en beskrivning av vad individen har för kompetenser. - göra studieplan bl.a. för att komma till sfi tidigt i processen. - se till att ansökan om svensk SSKleg. skickas till UHR - bedömning om det stämmer att det går snabbare att komma till arbete som USK än som SSK	Komvuxläraren mot ett yrkespaket utifrån Socialstyrelsens allmänna råd. (Socialstyrelsen) Utfärdande av legitimation för legyrken.		
Var 3) Vilka begreppsstrukturer/informationsmängder har hen tillgång till?			utbud: vad individen kommer med. efterfrågan: centrala innehållet i komvux yrkespaket.	Gymnasiekurser (kunskapskraven) inom yrkespaketet för USK. "Dataunderlag:" samtal och vad individen visar. - tester(?), samtal o praktiskt visa moment - utifrån yrkeslärarens	krav kring sjuksköterskedelen. Vad är kraven för att få SSK-leg? Är de samma som krav för att få SSK-examen? Finns de i högskoleförordningen?	

Figur 11: Utdrag ur tabell från kartläggning

För många celler i tabellen blev det uppenbart att det inte gick att komma fram till svar på sittande möte. De frågor som pekade mot större undersökningsbehov bör också delas i två, med olika följd för hur lämpliga svar tas fram:

- Vilka verksamheter, personer eller informationskällor kan vi fråga för att få de svar som behövs?
- Konceptuella utmaningar. Hur skapar vi det forum som behövs för att studera utmaningen närmare?

²⁸ <https://forum.sunet.se/file/file/download?guid=171cca15-5301-4601-991e-d659a73f6129&download=1>

Bilaga 3 – Överblick av SKOS

SKOS (Simple Knowledge Organization System) är en standard för att uttrycka begreppsstrukturer. Hjärtat i SKOS utgörs av begrepp som hänger ihop med varandra genom relationer. Ett begrepp i SKOS kodar någonting med en bestämd betydelse. En begreppsstruktur för arbetsmarknaden kan exempelvis innehålla ett begrepp för yrket konditor.

Ett begrepp i SKOS har en beständig identifierare, en så kallad URI (Uniform Resource Identifier), som kan användas för att referera till begreppet när det kopplas samman med andra begrepp eller när det används för att koda information. Ett begrepp har en huvudsaklig benämning (*preferred label* i SKOS), men kan även ha alternativa benämningar (*alternative labels* i SKOS) och en beskrivning (*description* i SKOS) av begreppet. Benämningar för begreppet och eventuell beskrivning utgörs av textsträngar.

Begrepp i SKOS kan kopplas ihop med relationerna *bredare* (benämnd *broader* i SKOS) och *smalare* (benämnd *narrower* i SKOS). Kopplingen "A bredare B" betyder att begreppet A kopplas mot ett begrepp B som har en bredare betydelse än A. En sådan koppling medför den omvända kopplingen "B smalare A" men även det omvända sambandet gäller. Ett begrepp i SKOS tillhör exakt en terminologi (benämnd *scheme* i SKOS). Alla begrepp i en terminologi ska vara av samma sort, t.ex. begrepp för yrken och grupper av yrken. Relationerna *broader/narrower* kan endast koppla ihop begrepp inom samma terminologi.

Utöver att koppla ihop begrepp genom relationer kan även en mängd av begrepp som hör ihop på något vis inordnas i en samling (benämnd *collection* i SKOS). Om man vill koppla ihop begrepp som tillhör olika terminologier kan relationerna `skos:closeMatch`, `skos:exactMatch`, `skos:broadMatch` och `skos:narrowMatch` användas. Dessa relationer skulle kunna användas för att uttrycka vissa av de översättningsnycklar som Semantikprojektet utreder.

Bilaga 4 – Underlag om Matchningsindikatorspåret

Övergripande syfte med detta huvudspår var att utveckla semantiken för att möjliggöra bättre förutsättningarna att tillhandahålla digitala tjänster inom domänen kompetensförsörjning och livslångt lärande. Olika språkbruk, klassificeringar och sätt att benämna kompetenser, utbildningar, beskrivande texter och en lång rad andra viktiga termer gör att dataöverföring, jämförelser, statistik, analys och planering försvåras och förhindrar att konsistens och samstämmighet kan uppnås.

Aktivitetsspår: Matchningsindikatorer - SUN/SNI/SSYK - aktiviteter

	status	april-22			maj-22			
		15	16	17	18	19	20	21
1. Etablerande av matriser och matchningsindikatorer								
1.1 Leverans av avidentifierade individdata från UREG, YREG och FDB för årgångarna 2015 - 2020	100							
1.2 Framställning av årgångsspecifika årspopulationstabeller – 6 tabeller	100							
1.3 Framtagning av definitiv innehållsspecifikation	100							
1.4 Framställning av matriser med matchningsindikatorer	100							
1.5 Hantering av sekretessfrågor samt framtagande av matriser per årgång nedladdade i Excelfiler	100							
	status	september-22			oktober-22			
		36	37	38	39	40	41	42
2. Logisk identifikation av samhörighet mellan klassifikationer samt utvärdering								
2.1 Utformning av reviderat förslag till plan för arbete med vidareanalyser av matchningsindikatorer	100							
2.2 Logisk identifikation av samhörighet i klassifikationer (nivå 1,2) - manuell	100							
2.3 Logisk identifikation av samhörighet i klassifikationer (detalj) – manuell (urval)	100							
2.4 Utvärdering av identifikation av samhörighet i klassifikationer (nivå 1,2,3,detaljnivå) - maskinell	100							

Figur 12: Aktivitetsbeskrivning

Beskrivning av klassifikationer - informationsmatriser

Som en slags förberedelser på arbetet i detta huvudspår genomfördes en detaljerad informationshämtning angående de i detta sammanhang centrala klassifikationerna SUN, SNI och SSYK. De tre klassifikationerna ägs och förvaltas av SCB, och ingår dessutom i SCB:s centrala klassifikationsnätverk (KLASNET), där samtliga ur användarsynpunkt större och viktigare kodförteckningar (klassifikationer och kodlistor) ingår. För var och en av dessa finns en utpekad klassifikationsansvarig (KLASA). Informationshämtningen kan bland annat sägas bestå av följande delar; två separata intervjuer (grundläggande aspekter samt kvalitetsaspekter) med KLASA för de tre klassifikationerna, kompletterande informationsinhämtning via rapporter och publicerade delar på nätet och framtagning av informationsmatriser. Denna informationshämtning genomfördes feb 2022 - april 2022. De tre KLASA bidrog dessutom med presentationer av sina respektive klassifikationer i samband med den workshop som genomfördes i Semantikprojektet i februari 2022.

Nationell klassifikation	Internationell motsvarighet
<p>SNI - Svensk näringsgrensindelning</p> <p>Standarden för svensk näringsgrensindelning, SNI, är främst en statistisk standard som används för att klassificera enheter som företag och arbetsställen efter deras ekonomiska aktiviteter. SNI gör det möjligt att jämföra och analysera data både nationellt och internationellt samt över tid. Den kan också användas för administrativa ändamål. SNI är byggd på ett hierarkiskt sätt. SNI börjar med en bokstavs nivå (Avdelning), följd av tvåsiffriga nivå (Huvudgrupp), tresiffriga nivå (Grupp), fyrsiffriga nivå (Undergrupp) och avslutningsvis femsiffriga nivå (Detaljgrupp).</p> <p>Register: Företagsdatabasen (FDB)</p>	<p>NACE - Nomenclature of Economic Activities</p> <p>Den gällande versionen är NACE Rev. 2. och motsvarar SNI 2007. Statistik som produceras utifrån NACE är jämförbar på europeisk nivå och delvis även på global nivå med FN:s nomenklatur ISIC (Rev. 4). Användandet av NACE är obligatoriskt inom det europeisk-statistiska systemet. Den nationella versionen fastställs av SCB:s generaldirektör och godkänns av EU enligt den förordning som finns för NACE. SNI och NACE är helt identiska på de fyra första positionerna. Därutöver har SNI en femte nivå, som är en svensk tilläggsnivå (även övriga EU-länder har på liknande sätt en nationell tillämpning av NACE i femte koden).</p>
<p>SUN - Svensk utbildningsnomenklatur</p> <p>Används för klassificering av enskilda utbildningar. Ger förutsättningar för att framställa jämförbar statistik och analys av befolkningens utbildning och det svenska utbildningssystemet såväl nationellt som internationellt. Registret över befolkningens utbildning, Utbildningsregistret (UREG) bygger på klassifikationen. Även företag och andra organisationer använder SUN i stor omfattning för exempelvis registrering av personalens utbildning. Klassifikationen består av två indelningar: en nivå- och en inriktningsklassifikation.</p> <p>Register: Utbildningsregistret (UREG)</p>	<p>ISCED - International Standard Classification of Education</p> <p>Internationell referensklassificering utvecklad av UNESCO för beteckningar på utbildningsnivåer samt för att organisera utbildningsprogram och tillhörande kvalifikationer efter nivåer och områden. Består av två nomenklaturer; ISCED 2011 (utbildningsnivåer) har genomförts i alla EU:s datainsamlingar sedan 2014. ISCED-F 2013 (utbildningsområden) har genomförts sedan 2016.</p>
<p>SSYK - Standard för svensk yrkesklassificering</p> <p>System för att gruppera individers yrken eller arbetsuppgifter. Används inom arbetsmarknads- och individstatistiken bland annat för att organisera information om lediga platser, arbetssökande och yrken eller för klassificering av anställda. Det finns även Yrkesnomenklaturer som "ägs" av andra organisationer som bygger på SSYK. Ett exempel är Näringslivets yrkesstruktur (NYK) som finns på Svenskt Näringsliv eller de motsvarande koder som används av kommuner och regioner via SKR.</p> <p>Register: Yrkesregistret (YREG)</p>	<p>ISCO - International Standard Classification of Occupation</p> <p>Tjänar som förlaga för nationella och regionala yrkesklassifikationer, tillämpbar för länder som inte utvecklar egna nationella yrkesklassifikationer och ger en aktuell och relevant grund för internationell rapportering, jämförelser, harmonisering och utbyte av yrkesinformation.</p>

Figur 13: Övergripande informationsmatris om SUN, SNI och SSYK

Nivå	SUN2000_Nivå	SUN2000_Inr	SUN2020_Nivå	SUN2020_Inr	SNI2007	SSYK2012
1-ställig	Nivå 1-siffer (7)	Huvudinriktning 1 (9)	Nivå 1-siffer (7)	Huvudinriktning 1 (9)	Avdelning (21)	Yrkesområde (10)
2-ställig	Nivå 2-siffer (14)	Huvudinriktning 2 (25)	Nivå 2-siffer (14)	Huvudinriktning 2 (25)	Huvudgrupp (88)	Huvudgrupp (46)
3-ställig	Nivå 3-siffer (50)	Ämnesinriktning (116)	Nivå 3-siffer (50)	Ämnesinriktning (116)	Grupp (272)	Yrkesgrupp (147)
4-ställig		Specificering (350)		Specificering (376)	Undergrupp (615)	Undergrupp (429)
5-ställig					Detaljgrupp (821)	

Figur 14: Antal koder per nivåaggregat för relevanta klassifikationsversioner

Framställning av matriser med matchningsindikatorer avser populationsandelar för kombinationer av klassifikationer på olika aggregat (1-, 2-, 3- och 4-ställiga för SUN och SSYK samt, 1-, 2-, 3- och 5-ställiga för SNI). Det innebär totalt fyra matriser per årgång. Data betraktas här som en form av testdata avgränsat för känsliga uppgifter i YREG för personer inom militära- samt för personer inom det polisiära. Dessutom redovisas enbart de jämförelsekombinationer av klassifikationer på de olika nivåerna som grundas på minst 10 individer. De kombinationer som baseras på färre än 10 individer redovisas i stället i en icke-detaljerad hopslagen grupp. För individdata från UREG används klassifikationen SUN2000 för årgångarna 2015 -2018 och SUN2020 för årgångarna 2019-2020. Data aggregeras på 1-, 2- och 3-ställig nivå samt på den mest detaljerade nivån (4-ställig för SUN_Inr och SSYK samt 5-ställig för SNI). Dessutom aggregeras utbildningsnivåvariabeln i SUN2000Nivå, och SUN2020Nivå på de fyra nivåerna grundskola, gymnasieexamen, akademisk examen och total nivå.

Antalet befintliga kombinationer (där kombinationen baseras på minst 10 individer) är exempelvis 55 254 kombinationer för Matris D (SUN_Inr2020) yrke (SSYK2012) och bransch (SNI2007). Det teoretiska antalet kombinationer oavsett ordningen mellan klassifikationerna, och oavsett om faktiska data finns för kombinationen är dock betydligt större (123 miljoner kombinationer), men innehåller naturligtvis kombinationer som i verkligheten är orimliga.

Matchningsindikatorerna beräknades sedan för matristyperna A-D för var och en av de sex årgångarna. I Figur 5 härnedan visas exempelutdrag ur matrisen mellan utbildning och yrke på 1-ställig nivå för 2015 (A1_UTB_YRKE_2015) för utbildningsnivån "Akademiker", där SUN1 = "Allmän utbildning" för andel_yrke. Vi kan dessutom, se ett utdrag för A1_YRKE_UTB_2015 för "Akademiker med chefsyrken" för andel_utb.

år	utbniva	SUN1	sun1 text	SSYK1	ssyk1 text	andel_yrke
2015	Akademiker	0	Allmän utbildning	1	Chefsyrken	2,8
2015	Akademiker	0	Allmän utbildning	2	Yrken med krav på fördjupad högskolekompetens	7,5
2015	Akademiker	0	Allmän utbildning	3	Yrken med krav på högskolekompetens eller motsvarande	10,7
2015	Akademiker	0	Allmän utbildning	4	Yrken inom administration och kundtjänst	14,0
2015	Akademiker	0	Allmän utbildning	5	Service-, omsorgs- och försäljningsyrken	35,5
2015	Akademiker	0	Allmän utbildning	6	Yrken inom lantbruk, trädgård, skogsbruk och fiske	1,0
2015	Akademiker	0	Allmän utbildning	7	Yrken inom byggverksamhet och tillverkning	11,0
2015	Akademiker	0	Allmän utbildning	8	Yrken inom maskinell tillverkning och transport m.m.	8,4
2015	Akademiker	0	Allmän utbildning	9	Yrken med krav på kortare utbildning eller introduktion	9,1
						100,0

år	utbniva	SSYK1	ssyk1 text	SUN1	sun1 text	andel_utb
2015	Akademiker	1	Chefsyrken	0	Allmän utbildning	0,1
2015	Akademiker	1	Chefsyrken	1	Pedagogik och lärarutbildning	11,4
2015	Akademiker	1	Chefsyrken	2	Humaniora och konst	3,7
2015	Akademiker	1	Chefsyrken	3	Samhällsvetenskap, juridik, handel, administration	32,7
2015	Akademiker	1	Chefsyrken	4	Naturvetenskap, matematik och data	4,8
2015	Akademiker	1	Chefsyrken	5	Teknik och tillverkning	26,8
2015	Akademiker	1	Chefsyrken	6	Lant- och skogsbruk samt djursjukvård	1,1
2015	Akademiker	1	Chefsyrken	7	Hälsa- och sjukvård samt social omsorg	14,3
2015	Akademiker	1	Chefsyrken	8	Tjänster	5,1
						100,0

Figur 15: Matchningsindikatorer, exempelutdrag från A1_UTB_YRKE_2015 och A1_YRKE_UTB_2015

SUN1	SSYK1
A Jordbruk, skogsbruk och fiske	0 Militära yrken
B Utvinning av mineral	1 Chefsyrken
C Tillverkning	2 Yrken med krav på fördjupad högskolekompetens
D Försörjning av el, gas, värme och kyla	3 Yrken med krav på högskolekompetens eller motsv.
E Vattenförsörjning, avloppsrening, avfallshantering och sanering	4 Yrken inom administration och kundtjänst
F Byggverksamhet	5 Service-, omsorgs- och försäljningsyrken
G Handel; reparation av motorfordon och motorcyklar	6 Yrken inom lantbruk, trädgård, skogsbruk och fiske
H Transport och magasinering	7 Yrken inom byggverksamhet och tillverkning
I Hotell- och restaurangverksamhet	8 Yrken inom maskinell tillverkning och transport m.m.
J Informations- och kommunikationsverksamhet	9 Yrken med krav på kortare utbildning eller introdukt.
K Finans- och försäkringsverksamhet	
L Fastighetsverksamhet	SUN1
M Verksamhet inom juridik, ekonomi, vetenskap och teknik	0 Allmän utbildning
N Uthyrning, fastighetsservice, resetjänster och andra stödtjänster	1 Pedagogik och lärarutbildning
O Offentlig förvaltning och försvar, obligatorisk socialförsäkring	2 Humaniora och konst
P Utbildning	3 Samhällsvetenskap, juridik, handel, administration
Q Vård och omsorg; sociala tjänster	4 Naturvetenskap, matematik och informations- och kommunikationsteknik (IKT)
R Kultur, nöje och fritid	5 Teknik och tillverkning
S Annan serviceverksamhet	6 Lant- och skogsbruk samt djursjukvård
T Förvärvsarbete i hushåll; hushållens produktion av diverse varor och tjänster för eget bruk	7 Hälsa- och sjukvård samt social omsorg
U Verksamhet vid internationella organisationer, utländska ambassader o.d.	8 Tjänster

Figur 16: Koder och textbeskrivningar på den 1-ställiga nivån för de tre klassifikationerna.

SNI1	SUN1	Textkommentarer
A Jordbruk, skogsbruk och fiske	6 - stark koppling 4 och 0 - svag koppling	6) Lantbruk och Jordbruk förefaller synonyma från mitt distanserade perspektiv, skogsbruk är exakt samma term. Det är däremot ytterst tveksamt om djursjukvård även inkluderar fiskeaktiviteter. 4) Naturvetenskap innehåller ämnet biologi som är en viktig bas kunskap i jordbruk, skogsbruk och fiske. 0) Allmän utbildning går, om man får tro namnet, att kombinera med samtliga näringsgrenar.
B Utvinning av mineral	4 och 5 - stark koppling 0 - svag koppling	4) Läran om mineraler är en del av ämnet geologi som i sin tur är en del av ämnet naturvetenskap. 5) För att utvinna mineral används olika tekniska maskiner, verktyg och funktioner som behandlas i grenar av teknikämnet. 0) Allmän utbildning går, om man får tro namnet, att kombinera med samtliga näringsgrenar.
C Tillverkning	4 och 5 - stark koppling 0 - svag koppling	4) (före läsning av SNI-beskrivning) m.a.a. min erfarenhet att matematik och naturvetenskap länge förknippats med industriell tillverkning, samt att informationsteknik får en allt mer central roll. (efter läsning av SNI-beskrivning) avdelningen omfattar fysisk eller kemisk omvandling av material, ämnen eller komponenter, fysik och kemi ingår i ämnet naturvetenskap. 5) Identisk term, tillverkning, är en del av namnet på ämnesområdet (SUN1:5XXX) 0) Allmän utbildning går, om man får tro namnet, att kombinera med samtliga näringsgrenar.
D Försörjning av el, gas, värme och kyla	4 - stark koppling 5 och 0 - svag koppling	4) el, gas, värme och kyla ingår i ämnena fysik och kemi som är en del av ämnet naturvetenskap. 5) För att utvinna (= tillse försörjning av...) el, gas, värme och kyla används olika tekniska maskiner, verktyg och funktioner som behandlas i grenar av teknikämnet. 0) Allmän utbildning går, om man får tro namnet, att kombinera med samtliga näringsgrenar.

Figur 17: Exempel på ett utdrag av ifylld Excelfil på 1-ställig-nivå (SNI1_SUN1)

Tanken var att samtliga bilaterala kombinationer mellan de tre klassifikationerna jämfördes med utgångspunkt från den ena eller den andra klassifikationen. Dessutom fanns ett fält för textkommentarer, som kunde fyllas i med kompletterande information. Figur 7 nedan visar exempel på en analyserad kombination mellan SNI1 och SUN1 i en ifylld Excelfil.

Samma upplägg planerades även för den 2-ställiga nivån, med dessa sex kombinationer. På den mest detaljerade nivån planerades urval av manuella analyser. I detta fall för kombinationer mellan SUN4 och SSK4, SUN4 och SNI5 samt SSK4 och SNI5. Av resursskäl och beroende på stor tidsåtgång för detta arbete, kunde inte den ursprungliga planen fullföljas. I stället gjordes en mindre ambitionssänkning av de kvarvarande manuella bedömningarna på 2-ställig nivå i slutet av denna aktivitet.

Det manuella arbetssättet varierade bland personerna som utförde analyserna, och arbetssättet pekade ut de kombinationer som enligt en logisk bedömning borde höra ihop. Det kan därtill nämnas att analyspersonernas kom från Af, MYH och Skolverket. Var och en kunde ganska fritt utföra arbetet på sitt eget sätt, och inte efter något sorts standardutförande. Eftersom bedömningarna naturligtvis dessutom är personliga och subjektiva, bör man se denna övning som ett slags ovetenskapligt test av hur personer som är förtrogna med klassifikationerna bedömer samhörigheten mellan koder utifrån egna erfarenheter av de textbeskrivningar som ges för varje nivå av klassifikationerna.

kod2 ==>>>

Kod1	SNI1	SNI2	SNI3	SNI4	SNI5	SUN1	SUN2	SUN3	SUN4	SSYK1	SSYK2	SSYK3	SSYK4
SNI1						LU				LU			
SNI2							M				LU		
SNI3								M				M	
SNI4									M				M
SNI5									M				M
SUN1	LU									LU			
SUN2		LU									M		
SUN3			M									M	
SUN4				M	IU								M
SSYK1	LU					LU							
SSYK2		M					LU						
SSYK3			M					M					
SSYK4				IU	M				M				

LU = Logisk identifikation av samhörighet + utvärdering för samtliga koder (redovisade matchningsindikatorer för samtliga koder)

M = Redovisade matchningsindikatorer för samtliga koder

IU = Logisk identifikation av samhörighet + utvärdering för urval av koder

Figur 18: Illustration av redovisade matchningsindikatorer per nivå 2020.

I Figur 18 åskådliggörs vilka kombinationer som analyserats och på vilket sätt (både logisk bedömning + utvärdering för samtliga koder, enbart matchningsindikatorer för samtliga koder eller logisk bedömning av urval av koder + utvärdering för dessa koder).

Med tanken på omfattningen av koder på den detaljerade nivån gjordes selektiva urval, där ett första steg avser val av två koder med bedömd matchning på 2-ställig nivå samt två koder som bedömts att inte matcha på 2-ställig nivå ur åtminstone en av de analyserade Excelfilerna och där steg 2 avser att välja samtliga underliggande detaljkoder för de fyra utvalda koderna.

I figur 19 nedan visas resultatet av utvärderingen med hjälp av matchningsindikatorer för de nivåer som hanns med att analysera inom ramen för denna aktivitet. Om vi exempelvis tittar på kombinationen SSYK1-SUN1 märks att vi har nio jämförelsekoder att fördela ut 100 % i matchningsindikatorerna, vilket ger oss tröskelvärde₁ = 0,12 om enbart visst samband bedömts vara fallet, och "tröskelvärde₂" = 0,24 om starkt samband bedömts. Vi ser vidare att 27 kombinationer pekats ut med viss koppling, och av dessa upplevs 14 kombinationer ha stark koppling. Av de 27 utpekade kombinationerna visar det sig att 48 procent är korrekta bedömningar (dessa överstiger alltså det uppställda tröskelvärdet), och motsvarande för de i "utpekade₂" (starkt samband) är 14 procent.

- Resultat manuella jämförelser mellan kopplingar för enställiga koder och populationsdata 2020

kombination	imfkoder	tröskelvärde1	utpekade1	andel korrekta1	tröskelvärde2	utpekade2	andel korrekta2
SSYK1 - SUN1	9	0,12	27	0,48	0,24	14	0,14
SUN1 - SSYK1	9	0,12	28	0,46	0,24	n.a.	n.a.
SNI1 - SUN1	9	0,12	62	0,58	0,24	26	0,42
SUN1 - SNI1	21	0,05	81	0,95	0,10	19	0,95
SNI1 - SSYK1	9	0,12	122	0,63	0,24	n.a.	n.a.
SSYK1 - SNI1	21	0,05	129	0,35	0,10	n.a.	n.a.

- Resultat manuella jämförelser mellan kopplingar för tvåställiga koder och populationsdata 2020

kombination	imfkoder	tröskelvärde1	utpekade1	andel korrekta1	tröskelvärde2	utpekade2	andel korrekta2
SSYK2 – SUN2	25	0,04	122	0,61	0,08	n.a.	n.a.
SUN2 – SNI2	85	0,02	211	0,39	0,04	140	0,26
SNI2 – SSYK2	46	0,03	380	0,66	0,06	n.a.	n.a.

- Resultat manuella jämförelser mellan kopplingar för detaljerade koder och populationsdata 2020

kombination	utpekade	andel korrekta1
SSYK4 – SNI4	25	100 %
SUN4 – SNI5	172	100 %

Figur 19: Resultat i utvärderingen av logisk bedömning av kodkombinationer med bedömd koppling

Denna manuella övning var lite av utforskande karaktär utan några givna ramar vad gäller hur kodkombinationerna skulle bedömas.

Om man trots denna vetskap skulle göra en grov analys av Figur 19, så tycks det som om andel korrekta för nivå 1 och 2 ligger strax under 60 procent, och det verkar inte vara någon större skillnad totalt mellan dessa nivåer.

Bilaga 4 – Underlag om intervjuer

Under vår och höst gjordes totalt 11 semistrukturerade intervjuer om vardera en timme, plus en friare variant för att hinna inom 30min. Målet var att fånga intervjupersonernas syn på hur deras processer skulle förändras om det fanns en gemensam begreppsstruktur och/eller översättningsnycklar. Intervjufrågorna skickades i förväg och den centrala frågeställningen var:

Hur skulle processer i er verksamhet förändras om det fanns gemensamma begrepp och/eller översättningsnycklar i form av:

- *(typ 1) en standardiserad översättningsnyckel mellan existerande klassifikationer (SSYK, SUN, SNI)?*
- *(typ 2) information om utbildningar och arbeten som länkas på en detaljnivå där det går att beskriva vad som är kritiskt för att utföra ett jobb väl?*

De tre sista intervjuerna utgick från antagandet att de Yrkesprofiler som tagits fram inom avfallshantering kan fungera som exempel på en önskad form av översättningsnycklar. Grammatiken i frågorna ändrades därför till att handla om dåtid. Så att svaren beskrev skillnader mellan före och efter yrkesprofilerna fanns.

Här är hela listan på de intervjuer som genomförts:

Trygghetsfonden TSL, VD	Caroline Söder	220421
Svensk Industrivalidering, senior konsult	Ingegerd Green	220426
Arbetsförmedlingen, klassificeringsanalytiker	David Norman	220519
Arbetsförmedlingen, produktansvarig rusta och matcha	Alaa Al Attar	220607
Samverkansinsats mellan AF och MYH för att tillgängliggöra utbildningsdata i Hitta yrken, projektledare	Leo Douglas	220607
Arbetsförmedlingen, verksamhetssamordnare kundflöden	Porya Korshid	220610
Volvokoncernen, kompetensutvecklingsansvarig	Fredrik Eliason	220613
Ibility Institute, VD	Torild Carlsson	220614
Konjunkturinstitutet, utredare	Jula Fennö	220817
Vakin AB, kompetensutvecklare	Fredrik Sörvåg	220825
Renova AB, HR-specialist/kompetensförsörjning	Sofia Westergaard	220830 (30min)
VafabMiljö Kommunalförbund, senior strateg	Thomas Nylund	220830

Bilaga 5 – UHR och semantikdelprojektet

Delprojektet för semantik med uppgift att studera hur för uppdraget relevanta begreppsstrukturer relaterar till varandra är centralt för uppdraget och området kompetensutveckling och livslångt lärande.

UHR har sedan tidigare erfarenhet av framtagning och underhåll av gemensamma begrepp kring utbildning, framför allt då från utvecklingen av den nationella EMIL-standarden liksom för de lokala begrepp kring antagning till högre utbildning i NyA-systemet och dess användargränssnitt antagning.se.

Under våren och sommaren har förutsättningarna för engagemanget försämrats och myndigheten har inte längre resurser att delta i alla delar av regeringsuppdraget och har därför valt att fokusera på den del där myndigheten bedömt kan bidra bäst, EU-perspektivet. UHR valde därför att lämna delprojektet kring semantik efter sommaren.

Totalt har UHR fram till dess använt SEK 64 386 av de för ändamålet tilldelade SEK 142 858 för arbete i semantikdelprojektet.