

Notat

Vedrørende:	Danskuddannelse til voksne udlændinge m.fl. - beskrivelse af fastlagte krav til indberetning af statistikoplysninger fra udbydere	02.11.2012
Skrevet af:	UNI-C: Jesper Lund/Lars Strange/Henrik Rosendahl-Kaa	
Version:	1.9: Rev. af Ankestyrelsen, januar 2014	

Dette dokument beskriver i detaljer strukturen for de danskundervisningsdata (aktivitetsdata), som skal modtages i Ankestyrelsen **dækkende kvartalerne 2010-4** og frem.

I forhold til seneste version 1.8, rev. 2 (notat af 30-08-2010), er der til version 1.9 implementeret følgende ændringer:

- **Henvisningsoplysninger:** Fra og med indberetningen 4. kvartal 2010 skal der indberettes i henhold til de reviderede indberetningskoder, som er meldt ud i det tidligere Integrationsministeriets informationsbrev nr. 5/2010.
- **Henvisningstabellen:** Der er ikke ændret på strukturen heri, og henvisningskategorien indmeldes stadig i feltet HHenvisKat. Men kodelisten for dette felt er revideret i overensstemmelse med de nye kategorier.
- **XML skema definition:** Er ændret, således at feltet for henvisningskategorier accepterer de nye talkoder.
- **Fejlkontrol:** Denne ændres således, at der ikke længere accepteres henvisningsdatoer, som ligger før det aktuelle kvartal for indberetning.
- **Versioner:** Version 1.8, rev. 2 af XML skemaet skal fortsat bruges til indberetninger for alle kvartaler op til og med 3. kvartal 2010. Nærværende version 1.9 skal (kun) bruges for indberetninger fra og med 4. kvartal 2010.

Der kan fra og med 1. januar 2010 kun indberettes med to versioner af XML skemaet.

- Version 1.8 bruges til at indberette data fra 2008-1 til 2010-3. Bemærk, at feltet Lektionslængde (UnLekMin) sættes til værdien 45, hvis det er ukendt for data fra før 1. kv. 2009, hvor det blev en obligatorisk del af indberetningen.
- Version 1.9 bruges til at indberette data fra 2010-4 og frem.
- **Der kan ikke længere indberettes data fra 2004-2007 via indberetningssystemet. Data fra disse år betragtes som lukkede.**

NB. Små udbydere, som benytter SM og UNI•Cs registreringsformular udarbejdet i MS InfoPath vil blive kontaktet af SM og UNI•C med henblik på, hvordan de indberetter for data fra og med 4. kvartal 2010. For kvartaler til og med 3. kv. 2010, kan de benytte den hidtidige udgave, som er i version 1.6, hvis de overholder følgende betingelser:

- De er ikke prøveafholdende, og indberetter derfor ikke prøve karakterer
- De har ikke lektionslængder, som afviger fra 45 minutter.

I nedenstående afsnit gives først en verbal beskrivelse af rammerne for og indholdet af indberetningerne under version 1.9. Herefter følger en præcis skematisk fortegnelse af data og hvilke værdier disse skal og kan antage. Endelig beskrives i sidste afsnit hvordan data, i forbindelse med indberetningen, skal organiseres i et XML dokument (evt. i bilagsdokument).

I forbindelse med spørgsmål vedr. indberetning af data og datastruktur kan der rettes henvendelse til følgende kontaktpersoner:

Christian Stoustrup
Ankestyrelsen, Data og analyse
E-mail: chs@ast.dk
Telefon: 61 89 74 94

Mette Agerskov Smith
Ankestyrelsen, Data og analyse
E-mail: mags@ast.dk
Telefon: 61 89 74 85

Charlotte Qvistgaard
Ankestyrelsen, Data og analyse
E-mail: cqv@ast.dk
Telefon: 61 89 74 91

Beskrivelse af datastrukturen

I dette afsnit gives en verbal gennemgang af hvilke data som skal indberettes set fra decentral (udbydernes) side.

Hvem skal indberette

Alle udbydere, som varetager danskundervisning i opdrag fra kommunerne, indberetter direkte til Ankestyrelsen.

Hvilke kursister skal indberetningen omfatte

Indberetningen skal omfatte alle kursister, som i den pågældende periode (kvartal) har en eller flere bevægelser på deres **forløbsoplysninger** (jf. næste afsnit), dvs. er blevet henvist til udbydere og/eller har modtaget undervisning, har afsluttet undervisning (på normal vis), er ophørt et uddannelsesmodul i "utide" og/eller har aflagt prøve hos udbyderen. En kursist kan naturligvis have haft flere af disse hændelser i den samme periode.

Hvilke oplysninger skal indberettes

De oplysninger, der skal indberettes, består af ét sæt **stamoplysninger** for hver CPR nr. ved udbyderen (se tabel: Kursist) og nogle **forløbsoplysninger** (et variabelt antal pr. kursist). De **forløbsoplysninger**, som skal indberettes er følgende:

- Kursistens henvisninger i perioden (se tabel: Henvisning).
- Kursistens indplaceringer eller afslutninger på et undervisningsmodul på et af de tre uddannelsesniveauer (se tabel: Modul).
- Kursistens deltagelse i undervisning under et sådant modul (se tabel: Undervisning).
- Kursistens prøveresultater (se tabel: Prøver)

Hvad skal data i hver tabel indeholde

Stamoplysninger:

- **Kursisttabel:** Oplysningerne i denne tabel gives én gang per CPR nr. inkluderet i indberetningen. Oplysningerne omfatter faste baggrundsoplysninger på kursisten, dvs. køn og nationalitet.

Indberettes kursisten over flere kvartaler, skal stamoplysningerne indberettes hver gang. I Ankestyrelsens statistikdatabase bliver det de sidst indberettede stamoplysninger, som bliver registreret for et CPR nr. hos en udbyder. Nye stamoplysninger for et CPR nr. hos en udbyder vil altså overskrive tidligere indberettede oplysninger.

Forløbsoplysninger:

- **Henvisninger:** For hver af de henvisninger, kursisten har for perioden, angives henvisningsdato, henvisningskategorien samt kursistens bopælskommune (henvisningskommune). Henvisningsdatoen skal ligge inden for den pågældende indberetningsperiode (kvartal). Der bør være en henvisning med dato før startdato på hver modul som indberettes for kursisten, men det er pt. ikke et krav for at indberetningen godkendes.

- **Uddannelse (Modultabel):** Her indberettes alle de *uddannelsesforløb*, som kursisten enten har modtaget undervisning på i indberetningsperioden, har afsluttet "normalt" (har bestået modultest) i perioden eller er ophørt fra i perioden. Det enkelte uddannelsesforløb identificeres ved **modul kategori** (1-6) og **uddannelses nr.** (1-3) samt **forsøg nr.** (fortløbende nr. for kursistens modulsætninger på denne kombination). Der registreres endvidere startdato for uddannelsen.

Når kursisten afslutter den pågældende danskuddannelse og modul herpåtilføjes en slutdato til den samme registrering og en ophørskode, som indikerer enten at kursisten er stoppet og hvad ophørsårsagen er (enten ophørskode eller dato for bestået modul test).

Kursister som er henvist og modulsat, men aldrig mødt op eller aldrig startet på undervisningen, registreres med startdato og slutdato samme dag samt med den relevante ophørsårsag.

- **Fravær i hele måneder indberettes med de tilbudte timer og samme antal timer som fravær. Fraværsindberetningen er dog frivillig.**

For at indberette modulet som et nyt forsøg skal 'pausen' i den pågældende DU/Modul være på mindst 3 måneder.

Bemærk at moduloplysninger **kun** skal indberettes for kursister som **enten** har haft undervisning i kvartalet **eller** er ophørt på denne modulkombination.

Bemærk at den samme *record* (dvs. kombination af modul nr, uddannelses nr. og forsøgs nr. for en kursist) i denne tabel godt kan gå igen i flere indberetningsperioder. Er kursisten startet med at modtage undervisning på et uddannelsesforløb i kvartal 1, indberettes dette med start dato. Denne record gentages så i alle fremtidige indberetninger så længe kursisten modtager undervisning på modulet. Den sidste gang recorden optræder i en indberetning er når kursisten afslutter den pågældende modulplacering. Her tilføjes så en slutdato samt **enten** en ophørskode hvis modulet afsluttes uden modultest, **eller** en dato for bestået modultest.

Bemærk at hvis kursisten er afsluttet "unormalt" på et modul (fx er gået på barsel eller aldrig er mødt op eller på anden måde har afbrudt modulet), **skal** der startes på en ny registrering af uddannelse i denne tabel, også selvom kursisten på et senere tidspunkt starter på **samme** uddannelsesforløb. Dvs. så snart der sættes en slutdato for en record i Modultabellen, lukkes der for yderligere opdateringer. Feltet **forsøg nr.** bruges her til at markere, når kursisten starter på samme modul igen på et senere tidspunkt.

- **Undervisning:** Denne tabel er en undertabel til Modultabellen, og heri registreres oplysninger om den undervisning kursisten har modtaget i indberetningsperioden på et givet uddannelsesforløb. Registreringerne organiseres på den måde, at der for hver måned i kvartalet opsummeres, hvor mange timer kursisten er blevet undervist fordelt efter en række kategorier både for undervisningsformen og undervisningens indhold. De præcise oplysninger er specificeret i næste afsnit. Der kan altså optræde op til tre records i denne tabel for hver kursist for hvert uddannelsesforløb. Der skal naturligvis være konsistens mellem modul- og undervisningstabellerne forstået på den måde, at der fx ikke kan være registeret undervisning i en måned, som ligger efter måneden for en slutdato angivet i Modultabellen.

- **Prøver:** Alle kursister som **har aflagt eller er tilmeldt** prøve hos udbyderen i perioden registreres her. En registrering i prøvetabellen er ikke betinget af, at kursisten optræder med en uddannelses/undervisningsregistrering, idet kursister kan aflægge prøve ved udbyderen uden at have modtaget undervisning samme sted eller i samme kvartal.

Der oplyses i prøvetabellen om datoen for prøvens aflæggelse, prøvens type, status for prøven, den evt. opnåede karakter o.a. oplysninger specifikt for den enkelte prøveaflæggelse. Har kursisten tilmeldt sig/aflagt flere forskellige prøver i perioden, optræder der flere registreringer for det samme CPR nr. Omvendt kan den samme **prøvedato** og **prøvetype** kun optræde én gang for et CPR. Der skal altså kun indberettes den seneste prøvestatus for kursisten for en given prøvedato og -type.

Overgangen til den nye 7 trins karakterskala (12 skalaen) er i indberetningen håndteret således, at for alle prøver, som indberettes i indberetningskvartaler før 4. kvartal 2007, skal den gamle 13- skala benyttes. Omvendt for prøveresultater for fra og med 4. kvartal 2007 – her skal alene den nye 12-skala benyttes.

Hvilke krav stilles der til validiteten af de indberettede data

Følgende er et krav til de indberettede data:

- Alle datafelter skal overholde nogle grundlæggende egenskaber specificeret senere i dette notat. Er et datafelt angivet som talfelt vil der fx ikke blive accepteret tekst. Ligeledes skal felter som datoer og CPR nr. m.v. overholde en række konventioner.
- For datafelter, som benytter **lukkede kodesæt** (se feltspecifikationerne senere), **skal** der benyttes en af de udmeldte koder. For Nationskoder vil der dog blive accepteret en vis afvigelse.
- Nogle datafelter er **obligatoriske** for enhver registrering i en given tabel og må derfor ikke mangle. For andre felter kan manglende accepteres, måske dog kun i et nærmere bestemt omfang.
- Der vil blive foretaget kontrol af den tidsmæssige konsistens i data, således at fx slutdatoen i et uddannelsesforløb ikke ligger før en måned, hvori der er angivet undervisning og lignende sammenhænge.

Hvordan skal data indberettes

Data organiseres i en XML fil og indberettes kvartalsvist til Ankestyrelsen. Skemadefinitionen på denne XML fil er beskrevet i sidste afsnit i dette notat. XML filen skal være opbygget hierarkisk, således at alle data, som vedrører en given CPR kursist står samlet.

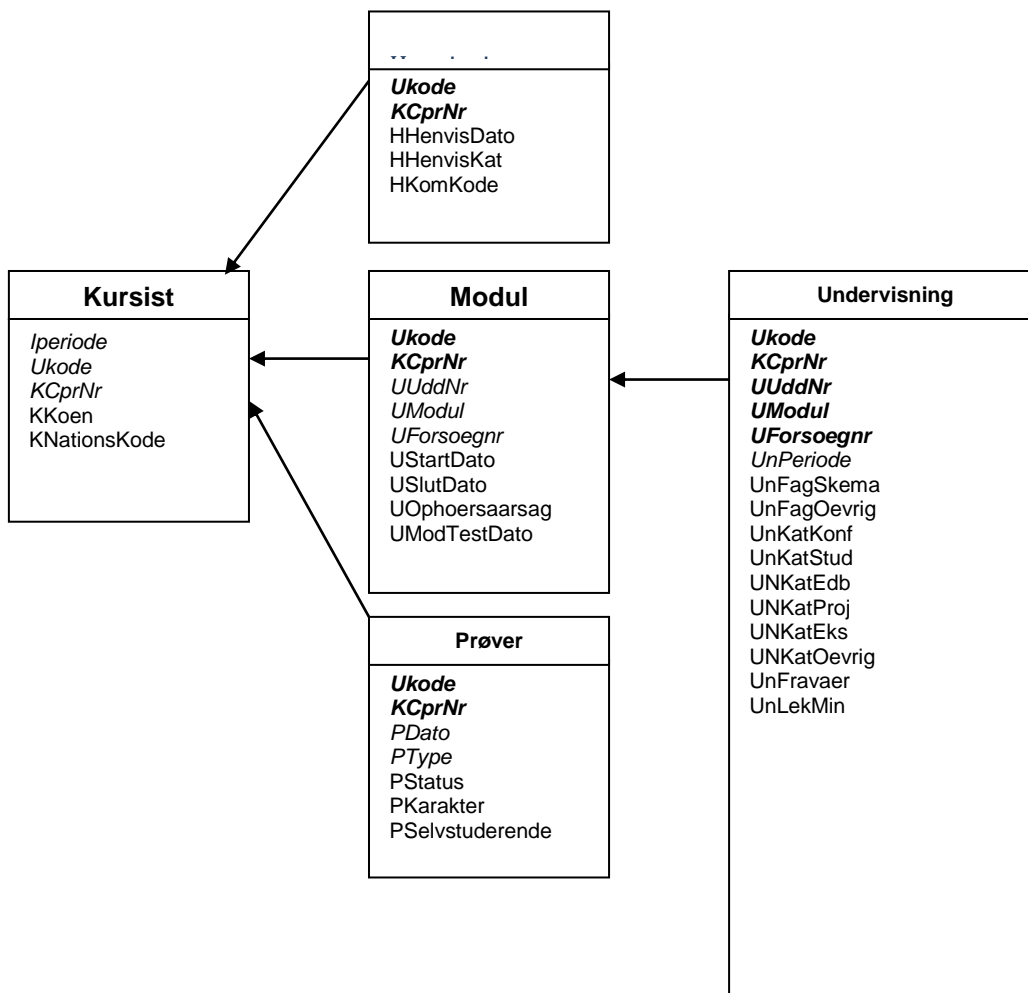
Opdateringer af allerede indberettede data

Så længe der ikke er indberettet for et nyt kvartal, vil der blive mulighed for udbyderen at gå ind og erstatte de indberettede data med nye data for kvartalet. Det er også muligt at Ankestyrelsen på baggrund af den fejlsøgning der laves i data, vil bede om en ny indberetning for samme kvartal.

Datastrukturen i skematisk form

Nedenstående figur viser datastrukturen i skematisk form. Datastrukturen er **hierarkisk**, dvs. alle tabeller på nær stamtabellen har en relation til en overliggende tabel, og alle records i alle tabeller kan føres tilbage til stamtabellen (via CPR nr. og udbyderkode). Felter som er **eksterne nøglefelter**, dvs. fastlægger forbindelsen til en overliggende tabel, står med **fed** skrift. Felter som er *interne nøglefelter*, dvs. medvirker til unikt at identificere den enkelte record, står med *kursiv* skrift. De fleste af nøglefelterne er både eksterne og interne.

I det dataformat som indberetningen skal foregå i (XML, se et senere afsnit), er det ikke nødvendigt at angive nøglefelterne mere end én gang. De er blot medtaget i alle tabeller for at illustrere den hierarkiske struktur i data samt for at vise, hvordan den enkelte record skal kunne identificeres unikt.



Felt specifikationer

I det følgende angives de nøjagtige specifikationer for de felter, som er indeholdt i hver tabel. For en række felter er de lovlige værdier af overskuelighedsgrunde dog angivet i en tabel efterfølgende, eller de er kun tilgængelige på en web side.

Bemærk: Er et felt markeret med X i kolonnen obligatorisk, afvises indberetningen, hvis feltet er tomt eller helt mangler. For ikke obligatoriske felter kan Ankestyrelsen evt. stille krav om en maksimal andel manglende oplysninger.

Kursist tabellen (stam data)

Felt Navn	Info	Lovlige værdier	Obligatorisk (X)
Iperiode	Indberetningsperiode (kvartal)	Tekst, kvartal nr YYYYQ, fx 20041	X
Ukode	Udbyder kode (=sprogcenter kode i offentlig udbyder regi)	Tal, seks cifre	X
KCprNr	Kursistens CPR nummer	Tekst, 10 tegn der accepteres kun alm. CPR, som overholder modulus 11 test.	X
KKoen	Køn på kursist	Tekst, 1 tegn M = Mand K = Kvinde	X
KNationsKode	Kursistens nations kode	Tekst, 3 tegn Følger FNs "ISO Alpha 3-code", se web siden: http://unstats.un.org/unsd/methods/m49/m49alpha.htm	X

Henvisning tabellen

Felt Navn	Info	Lovlige værdier	Obligatorisk (X)
HHenvisDato	Henvisningsdato	Dato, format: ÅÅÅÅ-MM-DD	X
HHenvisKat	Henvisningskategori (nye kategorier per 1. august 2010).	Talkode, 1 ciffer (1-9) Se listen nedenfor	X
HKomKode	Kommunekode for bopæls-kommunen	Talkode, 3 cifre. officiel kommune kode eller kode 999 for grænse- gængere	X

Henvisningskategori (**HHenvisKat**) kan være en af følgende talkoder 1-9, jf. det tidligere Integrationsministeriums informationsbrev nr. 5/2010:

- 1= I1** - Henvist efter integrationsloven, omfattet af integrationsprogrammet og modtager introduktionsydelse.
- 2= I2** - Henvist efter integrationsloven, omfattet af integrationsprogrammet og modtager ikke introduktionsydelse.
- 3= I3** - Henvist efter integrationsloven og omfattet af introduktionsforløbet.
- 4= D** - Henvist efter danskuddannelsesloven.
- 5= K1** - Henvist efter lov om en aktiv beskæftigelsesindsats og modtager kontanthjælp eller starthjælp.
- 6= K2** - Henvist efter lov om en aktiv beskæftigelsesindsats og modtager ikke kontanthjælp, starthjælp eller dagpenge.
- 7= A** - Henvist efter lov om aktiv beskæftigelsesindsats og modtager dagpenge
- 8= Ø** - Kursister, der ikke er omfattet af danskuddannelseslovens målgruppe, men som deltager som selvbetalere.
- 9= R** - Rekvireret undervisning

Modul tabellen

Felt Navn	Info	Lovlige værdier	Obligatorisk (X)
UUddNr	Danskuddannelsens nummer	Talkode, 1 ciffer 1 = Danskuddannelse 1 2 = Danskuddannelse 2 3 = Danskuddannelse 3	X
Umodul	Angiver uddannelses modul	Talkode, 1 ciffer 1 = Modul 1 2 = Modul 2 3 = Modul 3 4 = Modul 4 5 = Modul 5 6 = Modul 6	X
UForsoegNr	Fortløbende nr. for kursistens modsætning på dette modul på dette udd. niveau.	Tal 1, 2, 3 .. osv	X
UStartDato	Kursistens startdato på dette uddannelsesmodul	Dato format: ÅÅÅÅ-MM-DD	X

USlutDato	Slutdato for dette uddannelsesmodul	Dato format: ÅÅÅÅ-MM-DD	
UOphoersaarsag	Årsag til ophør	Talkode, 1 ciffer (1-8) Se listen nedenfor	X, ved angivet slutdato
UModTestDato	Dato for bestået modultest	Dato format: ÅÅÅÅ-MM-DD	

Ophørsårsag (**UOphoersaarsag**) kan være en af følgende talkoder:

- 1= Danskuddannelse afsluttet med prøve
- 2= Danskuddannelse afsluttet uden prøve
- 3= Ophør på grund af beskæftigelse
- 4= Ophør på grund af uddannelse
- 5= Ophør som følge af 3-års grænsens udløb
- 6= Ophør på grund af flytning til anden kommune eller anden udbyder
- 7= Kursisten ophørt på grund af meddelt sygdom, barsel, orlov eller børnepasningsproblemer (periode over tre måneder)
- 8= Andet/ukendt

Undervisning tabellen

Felt Navn	Info	Lovlige værdier	Obligatorisk (X)
UnPeriode	Det måned undervisningen forløber i.	Talkode, 2 cifre 1 = januar 2 = februar 12 = december	X
UnFagSkema	Skemalagt lærerstøttet undervis. - antal skemalagte lærerstøttede undervisningslektioner	Heltal, antal lektioner 0 – 400	X
<i>UnKat.. felter er valgfri underopdelinger af antal skemalagte lærerstøttede undervisningslektioner angivet i UnFagSkema. Lovlige værdier er som for UnFagSkema</i>			
UnKatKonf	Konfrontationslektioner	Heltal, antal lektioner	
UnKatStud	Studieværksted	Heltal, antal lektioner	
UnKatEdb	Edb-lokale/Sproglaboratorium	Heltal, antal lektioner	
UnKatProj	Projektundervisning	Heltal, antal lektioner	
UnKatEks	Ekskursion	Heltal, antal lektioner	
UnKatOevrig	Andet	Heltal, antal lektioner	
UnFravaer	Antal fraværslektioner af totalt antal skemalagte lærerstøttede undervisningslektioner	Heltal, antal lektioner 0 – 400	
UnFagOevrig	Øvrige timer, antal tilbudte lektioner	0 – 400	X
UnLekMin	Lektionslængden i minutter for lektioner angivet under UnFagSkema.	Heltal, antal minutter 15 – 120	X

Note. Er lektionslængden ikke den samme for alle timer angivet for et CPR nr. i et måned på en DU/Modul/forsøg nr. indberettes den gennemsnitlige lektionslængde og angives i minutter afrundet til heltal.

Prøve tabellen

Felt Navn	Info	Lovlige værdier	Obligatorisk (X)
PDate	Dato for denne prøve	Dato, format: ÅÅÅÅ-MM-DD	X
PType	Prøvetype	Talkode, 2 cifre 1 = Prøve i Dansk I – Lytteforståelse/ mundtlig kommunikation 2 = Prøve i Dansk I - Læseforståelse/ skriftlig fremstilling 3 = Prøve i Dansk II - Lytteforståelse/ mundtlig kommunikation 4 = Prøve i Dansk II – Læseforståelse 5 = Prøve i Dansk II - Skriftlig fremstilling 6 = Prøve i Dansk III - Lytteforståelse/ mundtlig kommunikation 7 = Prøve i Dansk III - Læseforståelse 8 = Prøve i Dansk III - Skriftlig fremstilling 9 = Studieprøve - Lytteforståelse 10 = Studieprøve - Mundtlig kommunika- tion 11 = Studieprøve - Læseforståelse 12= Studieprøve - Skriftlig fremstilling	X
PStatus	Status for denne prøve	Talkode, 1 ciffer 1 = Tilmeldt 2 = Ikke mødt 3 = Syg 4 = Gennemført 9 = Andet	X
PKarakter	Karakter givet ved prø- ven (12 eller 13 skala)	Tekst, 2 tegn Karakter fra 12 skala , hvis lperiode >= 20074. Karakterer fra 13 skala , hvis lperiode < 20074. NB karaktererne 0, 2 og 3 indberettes som hhv: "00", "02" og "03"	X, hvis PStatus = 4
PSelvstuderende	Har kursisten været selvstuderende	Tekst, 1 tegn J = Ja N = Nej	



XML skemadefinition

Nedenstående XML skemadefinition findes også på Ankestyrelsens hjemmeside som separat dokument.

Vær opmærksom på følgende:

- "Namespace" referencen i headeren er nu <http://statistik.uni-c.dk/INMindb/v1.9>.
- Feltet **version** angiver hvilken version af skemaet. Kun XML dokumenter af samme version, for denne 1.9, validerer mod dette skema.
- For nogle af tabellerne kræves det at felterne kommer i en bestemt rækkefølge (samme som i definitionerne ovenfor). Dette er angivet med `<xs:sequence>` tag'en.
- For nogle felter kræves det, at der **altid** står en tag for feltet uanset om der findes data eller ej. Dette krav kan ses ved specifikationen `nillable="true"` i `<xs:element>` tag'en tilhørende feltet.
- Feltet **produktkode** er systemleverandørens mulighed for at angive fx systemnavn og version på det system, der har produceret XML filen. Maksimalt 256 tegn fri tekst.

NB. Selve XML skemaet for version 1.9 findes i en separat fil:

INMindb2009_v1-9.xsd