

Testrapport:

SATS skoleadministrativt system 1999 – 2016 - arkivuttrekk fra fagsystem for grunnskolen i Stange kommune

Generelt om systemet

SATS (systemleverandør IST - International Software Technology AS) omfatter fagsystemer for skole- og oppvekstsektoren (barnehage, grunnskole og videregående skole). SATS grunnskole består av flere moduler: 'Modul elev-skole', 'Modul fagdokumentasjon', 'Modul eksamen og prøver', 'Modul personal', 'Modul økonomi', 'Modul rom og utstyr' og 'Modul grunndata'. Det finnes også andre moduler i fagsystemet som ikke er innkjøpt av Stange kommune. Hvilke moduler den enkelte kommune velger å bruke (dvs. kjøpe), vil variere.

SATS ble tatt i bruk ved skolestart skoleåret 1999/2000 og avsluttet etter skoleåret 2015/2016 – så sent på høsten 2016 at all klagebehandling mm fra skoleåret var avsluttet, jf. nedenfor om SIP og AIP.

SATS inneholder bl.a. demografisk informasjon om alle elever, adresser og evt. flytting, foresatte og kontaktinformasjon, skoleår og skole- og klassetilhørighet. Videre inneholder systemet informasjon om fag, faggrupper, elevprosjekter, faglærere, karakterer og annen evaluering, fravær, eksamen, skoler, rom og utstyr, skoleskyss mm. Systemet inneholder data på individnivå – om elever, foresatte, ansatte og eksterne kontaktpersoner. Individdata kalles med en fellesbetegnelse for 'mikrodata'. SATS har rapportfunksjoner for et stort antall standard rapporter, for statistikk til SSB, og en rekke standardtekster – flettetekst – som legges inn i utskrifter og rapporter.

OBS! Systemet er utviklet for et internasjonalt marked og inneholder derfor mange tabeller og felt som ikke er relevante for norsk skole. Mange tabeller og felt er derfor tomme, og mange tabell- og feltnavn er på engelsk.

Hjemmelsgrunnlag for funksjonsområdet er *Lov om grunnskolen og den videregående opplæringa (opplæringslova)*, LOV-1998-07-17-61. Det vises spesielt til lovens kapittel 15: *Bruk av forvaltningslova* for bestemmelser om opplysningsplikt, taushetsplikt m.m. Dessuten kommer *Forskrift om sikring av skyssberettigede skoleelever i buss*, FOR-2012-12-19-1342, til anvendelse. Se også 'Klausulering og brukertjenester' nedenfor bl.a. om behandling av personopplysninger.

Teknisk sett er SATS en relasjonsdatabase der alt innhold består av strukturerte data i tabeller og felt. Denne versjonen av systemet, og dermed dette arkivuttrekket, omfatter 164 tabeller – herav 71 tomme tabeller, dvs. tabeller som bare inneholder feltnavn i første rad. SATS inneholder f.eks. ikke ferdig genererte vitnemål som 'fulltekstdokumenter' (eks. som pdf). Vitnemål og annen dokumentasjon må på vanlig måte genereres i spørringer basert på tabeller og felt, jf. nedenfor. Databaseplattformen i dette tilfellet er Oracle. SATS ble driftet av IT-avdelingen i Stange kommune inntil Hedmark IKT (HIKT) ble opprettet i 2006. HIKT er et IKT-samarbeid mellom Hamar, Stange, Løten og andre kommuner i Hedmark (som ikke er tilknyttet IKA Øst). HIKT overtok fysisk serveren fra Stange kommune.

SATS er benyttet sammen med fysiske elevmapper. Mappene inneholder analog informasjon. Innholdet i mappene må også bevares, men overføres ikke til arkivdepot før materialet er 25 – 30 år gammelt i hht. bestemmelsene i [Riksarkivarens forskrift § 6-2](#).

Bevaringsvurdering

Bevaringsvurderingen skal avklare om SATS skole inneholder informasjon som er bevaringsverdig etter Arkivlovens § 1: *‘(...) arkiv som har monaleg kulturelt eller forskingsmessig verdi eller som inneholdt rettsleg eller viktig forvaltningsmessig dokumentasjon.’* Paragrafen sidestiller m.a.o. ‘informasjonsverdien’ (forskning) og ‘dokumentasjonsverdien’ (rettighet/arkivskapers virksomhet). Bevaringsvurderingen følger [Riksarkivarens hovedprinsipper](#), [Bevaringsutvalgets](#) formål, kriterier og metodikk (formålene F1 – F4 med tilhørende kriterier, jf. vedlegg 1 nedenfor), fagspesifikke bestemmelser i [Riksarkivarens forskrift](#) og Riksarkivarens anbefalinger om [‘Innhaldsvurdering av arkiva og systema’](#), pkt. 3.2.

En generell erfaring mht. bevaringsvurdering av fagsystemer etter Bevaringsutvalgets metodikk, er at konklusjonen ofte blir ‘nei’ etter vurdering av F1-kriteriene (jf. pkt. 4 i den skjematiske fremstillingen side 61 i Bevaringsutvalgets rapport – vedlegg 1 her). Da skal vurderingen uansett fortsette med F2-kriteriene (informasjonsverdien). Det er dekning for å starte vurderingen direkte med F2-kriteriene *‘dersom man på forhånd vet at den enheten man skal vurdere har et innhold med stor informasjonsverdi for ettertiden’* (Bevaringsutvalget, side 62, andre avsnitt). Teknisk sett er informasjonen i SATS tilrettelagt for maskinell bearbeiding og analyse. Generelt er dette forhold som tilsier stort potensial for forskning og dermed også stor etterspørsel. Bevaringsvurderingen av SATS starter derfor i punkt 5 i den skjematiske fremstillingen side 61 i Bevaringsutvalgets rapport. Dersom svaret blir ‘ja’ på spørsmålet om *unikhet* (punkt 5a i skjemaet), leder det til konklusjonen ‘Bevares for ettertiden’ (pkt. 7a), og i så fall er det unødvendig å vurdere F3 og F4. Konklusjon ‘7a’ innebærer at både dokumentasjonsverdien og informasjonsverdien bevares.

Fagsystemet SATS er en digital videreføring av manuelle prosesser og rutiner i et funksjonsområde med lang historie som også vil videreføres i nye generasjoner fagsystemer etter hvert som systemer byttes ut. Fagsystemet representerer derfor *kontinuitet* innenfor funksjonsområdet. All historikken i SATS er bevart. Systemet inneholder komplett registrering av alle elever med tilhørende data i perioden systemet var i bruk, og oppfyller alt i alt [Bevaringsutvalgets](#) kriterium om *tidsspenn/kontinuitet* og *omfang*. Dette tilsier bevaring.

Videre inneholder systemet nasjonale koblingsnøkler: fødselsnummer/D-nummer, organisasjonsnummer (for skolene og kommunen), postnummer og kommunenummer og oppfyller dermed også kriteriet om *lenkbarhet* (kobling). Det tilsier også bevaring. Elevene kan følges videre gjennom skoleløpet i fagsystemer for videregående opplæring, og i høyere utdanning vha. fødselsnummeret. Populasjonen i SATS kan f.eks. sammenstilles med tilsvarende populasjoner fra fagsystemer i samtlige kommuner til en landsomfattende populasjon. Hele eller deler av populasjonen kan selvsagt også kobles med alle andre totalpopulasjoner som inneholder fødselsnummer, til tverrsnitts- og paneldata.

Fagsystemer er generelt utviklet for å tjene klart definerte og spesialiserte formål, tilrettelagt for høy grad av maskinell saksbehandling og med svært ensartet og strukturert innhold. Dette gjelder også SATS som dermed har liten *tematisk variasjon* – en egenskap som trekker i retning kassasjon. Bevaringsutvalget anbefaler likevel (s. 56) at dette kriteriet ikke tillegges vekt i retning kassasjon når

det er snakk om unik informasjon. Dette må også vurderes opp mot dekningsgrad, kontrollerbarhet og representativitet (Bevaringsutvalget s. 56, andre avsnitt). SATS 'scorer' høyt på disse kriteriene. Kriteriene *kvalitative egenskaper* og *alder* har liten relevans for denne typen materiale: Digitalt skapt arkivmateriale er alltid ungt materiale, men en levetid på 16 år er likevel lenge i digital sammenheng.

Databaseplattformen Oracle har generell import- og eksportfunksjonalitet og SATS har i tillegg eksport- og importfunksjonalitet for en rekke spesifikke formål. Demografiske data importeres fra Det sentrale folkeregister fire ganger årlig. Data utveksles med systemer for videregående opplæring på fylkesnivå (eks. VIGO ved Inntakskontoret), til PAS (Utdanningsdirektoratet for nasjonale prøver), med kommunens økonomisystem (Agresso) om foreldrebetaling, lønn for ansatte etc., personalsystemer (kommunalt HRM og sentrale tjenestemannsregistre), med andre skolars fagsystemer (Extens bl.a.), flyttmeldinger mm. SATS utveksler også informasjon med læringsplattformer (Itslearning i Stange kommune), med PP-tjenestens fagsystem (PPT-HK i dette tilfellet). Det er derfor mulig å finne noe av informasjonen fra SATS andre steder (*redundans*), men da vil informasjonen være spredd utover forskjellige arkivskapere og systemer og inngår i en annen kontekst og en annen proveniens. Sammenstillingen av data i SATS representerer totalt sett *unik* informasjon som bare finnes samlet dette ene stedet – med egen kontekst og proveniens. Bevaringsutvalgets kriterium om *unikhet* er dermed oppfylt, og dette tilsier bevaring.

I fagspesifikke bestemmelser i Riksarkivarens '[Forskrift om utfyllende tekniske og arkivfaglige bestemmelser om behandling av offentlige arkiver](#)': IV. *Bevarings- og kassasjonsbestemmelser for fylkeskommunale og kommunale arkiv skapt etter 1950*, § 7-28, pkt. 3 bokstav e: [Opplæring og oppvekst](#) - spesifiseres følgende:

Om alle elever bevares følgende dokumentasjon:

- *Hvilke elever som har gått på hvilken skole, inkludert fravær*
- *Underveisvurdering, inkl. halvårsvurdering med karakter*
- *Sluttvurdering, inkl. standpunktkarakter og eksamenskarakter*
- *Ved fritak for vurdering med karakter skal det bevares annen dokumentasjon av opplæringen.*

Alt dette er dokumentasjon som finnes i SATS som strukturert informasjon i tabeller og felt. All informasjon kan enkelt gjenskapes via spørringer basert på tabelluttrekkene fra SATS - som spørring om enkeltelever, om den enkelte klasse, om alle elever, grupper av elever, om den enkelte skole eller for alle skoler samlet osv. Spørringene kan evt. redigeres som rapporter.

Et tilleggskriterium som også skal vurderes (Bevaringsutvalget, s. 66), er fremtidig *etterspørsel*. Det kan selvsagt være vanskelig å forutsi, men det er solid empiri for at digitale mikrodata generelt er sterkt etterspurt i forskning - i særlig stor grad gjelder dette digitale mikrodata som kan kobles og analyseres maskinelt. Dette gjelder også SATS. Det er all grunn til å tro at denne typen etterspørsel bare vil øke i fremtiden. SATS inneholder en rekke variabler som kan gjenbrukes i historisk og samfunnsvitenskapelig forskning. 'Utdanning' er en mye brukt analysevariabel. Populasjonen i SATS lokalt kan f.eks. sammenstilles med tilsvarende populasjoner fra fagsystemer i samtlige kommuner til en landsomfattende populasjon. Elevenes utdanningsløp er dokumentert gjennom hele grunnskolen og er i praksis paneldata.

I hht. Bevaringsutvalget skal også følgende tilleggskriterier vurderes:

- Symbol/identitetsverdi

- Estetisk verdi
- Pedagogisk verdi
- Gjenstandsverdi
- Brukbarhet
- Fysisk tilstand

De første kriteriene har størst relevans for papirmateriale, men 'Brukbarhet' og 'Fysisk tilstand' er relevant på den måten at datagrunnlaget faktisk må være teknisk intakt og lesbart. Det kan selvsagt også hevdes med en viss rett, at arkivuttrekket fra SATS har pedagogisk verdi, f.eks. som illustrasjon i opplæringsøyemed.

Alt i alt tilsier bevaringsvurderingen at informasjonen i systemet skal langtidsbevares ved godkjent depotinstitusjon. Den forvaltningsmessige dokumentasjonen i systemet vil ha tidsbegrenset bevaringsverdi, men den kulturelle og forskningsmessige verdien er evigvarende og tilsier at informasjonen 'bevares for ettertiden', dvs. konklusjon '7a' etter Bevaringsutvalgets metodikk, s. 61.

For at disse konkrete bevaringskravene skal oppfylles, må alle tabeller og alle felt fra SATS bevares. Når f.eks. kriteriet om *lenkbarhet* er oppfylt og inngår som en del av begrunnelsen for bevaring, må selvsagt også lenkbarheten være intakt i arkivuttrekket, dvs. at nasjonale koblingsnøkler (fødselsnummer/D-nummer, organisasjonsnummer osv.) må være integritetssikret. Også tomme tabeller må være med i SIP for å opprettholde proveniensens og som dokumentasjon på at ingen informasjon er underslått. At alle tabeller og alle felt inkluderes, er en absolutt betingelse for at informasjonen skal kunne gjenskapes, jf. nedenfor om kardinalitet. Det vises også til ER-diagram. Det er ikke noe alternativ å plukke ut enkelte tabeller eller felt og bevare disse, mens resten av fagsystemet kasseres ('plukk-kassasjon'). Konsekvensen vil da være at vitale data går tapt slik at informasjon ikke kan gjenskapes. All kassasjon skal ha en hjemmel - det vil være vanskelig å finne en hjemmel for plukk-kassasjon. En slik tilnærming vil dessuten være brudd med proveniensprinsippet. Det ville også bryte med prinsippet om at digitale data skal bevares autentisk og integritetssikret i arkivdepot. Endelig ville en arkivpakke (SIP) som ikke er identisk med arkivskapers versjon, være en sekundærkilde med svakere beviskraft enn primærkilden. (SATS vil eksistere hos arkivskaper som et avsluttet system for oppslag og søking så lenge det er lesbart, og vil derfor inntil videre være primærkilden). Historikere er faglig forpliktet til å bruke primærkilder som grunnlag i sin forskning. Så lenge arkivdepotets versjon er identisk med primærkilden, kan publikumstjenester besvares av arkivdepotet. Dersom SIP skulle inneholde manipulererte data, data som ikke er identisk med primærversjonen, vil dette være mindre interessant for historikernes bruk.

I et hvilket som helst databasesystem vil informasjonsinnholdet være lagret i felt i tabeller som igjen knyttes sammen med nøkler i relasjoner. Eventuelle 'fulltekstdokumenter', dvs. fritekst/merknader i ustrukturert form som 'ser ut'/'kan se ut' som tradisjonelle A4-dokumenter (f.eks. i RTF-felt eller BLOB-felt), er også lagret som felt i tabeller.

Ulike typer informasjon genereres ved spørringer og funksjoner der aktuelle tabeller, felt og rader knyttes sammen etter behov. Kombinasjonsmulighetene av tabeller og felt er tilnærmet uendelige. Informasjon som genereres gjennom kombinasjon av tabeller og felt representerer en 'merverdi' som ikke finnes i de enkelte tabeller og felt hver for seg. Denne 'merverdien' lagres ikke i systemet fordi den er dynamisk og kan gjenskapes når som helst via spørringer basert på underliggende tabeller og felt. De enkelte tabeller og felt er statiske og forseglet med sjekksummer, dvs. et forsøk på å endre selv en BIT, vil føre til en annen sjekksum. For å kunne gjenskape merverdien og ikke minst generere ny merverdi, må alle felt og

tabeller være med på et arkivuttrekk. Den virkelige store gevinsten for forskning er at data fra ulike fagsystemer og forskjellige arkivskapere kan kobles sammen (lenkes) for et gitt forskningsprosjekt.

For å genere en karakterutskrift/et vitnemål, må arkivdepotet vite elevens fødselsnummer og hvilket skoleår eleven ble uteksaminert - hvis det er sluttresultatet det gjelder. Det ligger jo mye annen informasjon om elevene i systemet, som det også vil bli etterspørsel etter. Det vil alltid være elevens fødselsnummer som er søkekriterium og identifikasjon. Fødselsnummer er entydig i motsetning til navn. Spørringen resulterer i en tabell, som så kan redigeres slik at den ser ut som en karakterutskrift/vitnemål. Deretter kan dette skrives ut, evt. på papir med bakgrunn som er vanlig for vitnemål.

Som aktivt system har SATS funksjonalitet for å genere vitnemål. Det er flere arkivskapere som har ønsket å få den underliggende kildekoden for dette fra IST, men dette vil leverandøren ikke utlevere. Det er forståelig ettersom dette er noe av det leverandøren tjener penger på. Dette vil være forretningshemmeligheter. Likevel er det fullt mulig å rekonstruere denne funksjonaliteten på bakgrunn av teknisk dokumentasjon i arkivuttrekket – hvilke tabeller og hvilke felt som må inngå i en slik spørring kan utledes fra teknisk dokumentasjon.

SIP¹ og AIP²

IKA Øst følger bestemmelsene for arkivuttrekk som er nedfelt i Riksarkivarens forskrift gjeldende fra 01.01.2018 - [Kapittel 5. Elektronisk arkivmateriale. Lagringsmedier, format og avlevering \(§§ 5-1 til 5-32\)](#). IKA Øst velger å følge bestemmelsene for statlig sektor, selv om dette ikke er et krav for kommunale systemer i forskriften – jf. '[Virkeområde](#)' pkt. 1) og 2).

Tabellene er lagret som kommaseparerte filer med variabel felt- og radlengde, jf. [Riksarkivarens forskrift § 5-12 b](#). Semikolon er benyttet som feltskille og CR/LF som radskille. Doble anførselstegn er benyttet som tekstkvalifikator. Som file-extension er det benyttet CSV og ikke DAT (som er standard). Dette er valgt fordi CSV blir 'forstått' automatisk av databasesystemer ved re-import av datatabellene. Tabellene inneholder også feltnavn i første rad, fordi dette forenkler re-import til nytt databaseverktøy. Tegnsettet er ISO 8859-1:1998, Latin 1, jf. [Riksarkivarens forskrift § 5-11](#). Øvrige relevante bestemmelser for innleveringen finnes i Riksarkivarens forskrift [§ 5-10](#) til [§ 5-13](#), [§ 5-24](#) til [§ 5-26](#), [§ 5-27](#) til [§ 5-28](#), [§ 5-30](#) til [§ 5-32](#). Alt dette er spesifikasjoner som defineres spesifikt i uttrekksprosessen i Oracle SQL Developer. SIP inkluderer også faste SQL-oppsett og views. ER-diagram fra systemleverandør er også inkludert som svg-filer. Dette er imidlertid et proprietært format som ikke vil være lesbart over tid.

SATS ble avløst av OppAd og Vokal fra skolestart høsten 2016. SIP ble opprinnelig innlevert til IKA Øst som et prøveuttrekk fra SATS i oktober 2016. Tabelluttrekket ble utført av HIKT. I løpet av juni 2018 ble det imidlertid avklart at det ikke hadde vært noen oppdatering i systemet etter oktober 2016. Arkivskaper har derfor avgjort at prøveuttrekket skal endre status til endelig innlevering.

Virussjekk av SIP 16.10.2016: Ingen virus eller trusler funnet

Virussjekk av AIP 19.09.2018: Ingen virus eller trusler funnet

¹ Submission Information Package – overføringspakke fra arkivskaper til arkivdepot - 'originalen'

² Archival Information Package – depotpakke, som inkluderer SIP + teknisk og annen dokumentasjon og testrapport bl.a.

Tekniske metadata (*metadata.xml* - også kalt *addml-fil*) er laget av IKA Øst vha. [Arkadukt](#) som er Riksarkivarens verktøy for standardisering av tekniske metadata. Versjonen som er benyttet, er Arkadukt 2 - ADDML³ 8.2.

Stange kommunes versjon av SATS består av 164 tabeller, dvs. at SIP består av uttrekk fra 164 tabeller. Av disse er 71 tabeller tomme, dvs. disse tabellene inneholder bare feltnavn. Slike tabeller har ikke vært i bruk og har ingen praktisk funksjon i forhold til koblinger og relasjoner mellom de øvrige tabellene, men de er en del av proveniensens.

Det vises til INNHOLD.TXT for struktur på avleveringspakken, til *metadata.xml* og til ER-diagram for teknisk dokumentasjon. ER-diagrammene er laget av IKA Øst. ER-diagrammene dokumenterer nøkkelfelt, relasjoner og kardinalitet for hovedmodulene. Kodetabeller og tabeller med systemteknisk innhold er som hovedregel ikke inkludert i ER-diagrammene. Tabeller som danner tverrforbindelse mellom moduler, er inkludert i aktuelle modul-diagram. Brukerhåndbøker og annen dokumentasjon er levert fra HIKT og lagret under mappen SYSDOK.

Format, nøkler, relasjoner, og kardinalitet

SIP inneholder primærversjonen av tabelluttrekkene. Tabellene i SIP reflekterer derfor bl.a. også sluttbrukeres autentiske tastetrykk. I enkelte tilfelle innebærer dette at sluttbrukere har lagt inn tegn i merknadsfelt/fritekstfelt som kan skape problemer når tabellene skal re-importeres i et nytt databasesystem. Det finnes f.eks. semikolon inne i enkelte fritekstfelt som sluttbrukere har lagt inn i noen tabeller. Semikolon er som hovedregel brukt som feltskilletegn i kommaseparerte filer. Ved reimport av tabellene til nytt databasesystem vil semikolon tolkes som feltskilte, men dette unngås med doble anførselstegn som tekstkvalifikator.

Noen tabeller inneholder også CR/LF (linjeskift) inne i merknadsfelt. Dette skyldes at sluttbrukere har trykket 'Enter' inne i feltet for å skape en slags formatering. Dette vil også skape problemer ved re-import av tabellene til ny databaseplattform, ettersom CR/LF da vil tolkes som et radskille – 'falske linjeskift'. Tabellene er beholdt uendret i SIP, dvs. med de ekstra linjeskiftene, men det er opprettet nye versjoner av disse tabellene i AIP, der ekstra linjeskift inne i merknadsfelt er fjernet maskinelt (i mappen 'SYSDOK\Tabeller med fjernet CRLF'). Linjeskift er erstattet med mellomrom. Tabellnavn for disse tabellene inneholder <navn>2.csv. En forutsetning for maskinell fjerning av CR/LF, er at feltene ligger innenfor doble anførselstegn.

Samtlige SIP-tabeller med innhold (nummerert 1 – 93 i AIP) er importert til Microsoft Access uten feilmeldinger mht. feltlengder, formater og tegnsatt med unntak av tabeller med falske linjeskift, jf. over. Antall rader ved import er identisk med antall rader i alle tabeller i SIP.

Det er altså opprettet kopier av enkelte tabeller i AIP for å kunne gjennomføre nødvendige kontroller, import til MS Access og konvertering av BLOB-felt. SIP-tabellene, dvs. primærdata, er alltid uendret. Det må være opp til fremtidige brukere om de vil velge å bruke SIP-tabellene eller kopiene i AIP i slike tilfelle. Dersom SIP-tabellene brukes, må altså 'problemene' med CR/LF og semikolon som er dokumentert her, løses på nytt.

³ ADDML (Archival Data Description Markup Language) – Arkivverkets egenutviklede standard for teknisk beskrivelse av datasett.

Nøkkelkontroll er bare mulig å gjennomføre på tabeller med innhold. Samtlige primærnøkler validerer. Samtlige fremmednøkler validerer i alle tabeller i følgende moduler:

- ‘Modul eksamen og prøver’
- ‘Modul personal’
- ‘Modul rom og utstyr’
- ‘Modul elev-skole’

I ‘Modul økonomi’ har det ikke vært mulig å etablere kobling mellom tabellene *salarytable.csv*, *salaryclass.csv*, *salarygrade.csv* og hovedtabellen i modulen: *salaryitem.csv*.

‘Modul grunndata’ inneholder kodetabeller/hjelpetabeller som kan knyttes til de enkelte hovedmodulene. En rekke koblinger er testet, og samtlige testede knytninger validerer.

Tabeller med lange binærdata (BLOB)

SATS inneholder seks tabeller hvor det finnes BLOB-felt. Det vises spesielt til dokumentasjon og merknader i *metadata.xml*.

Dette gjelder følgende tabeller:

- *report.csv* – tabellen inneholder 719 rader, feltet ‘template’ er et blobfelt – konvertering viser at dette feltet inneholder formateringskoder for visning/utskrift av ulike typer rapporter og lister.
- *diploma_memo.csv* – tabellen inneholder 45 rader, feltet ‘memoblob’ er et blobfelt – konvertering viser at feltet inneholder fritekstmerknader som ‘ren tekst’.
- *diplomareport.csv* - tabellen inneholder 316 rader, feltet ‘blobnote’ er et blobfelt - konvertering viser at feltet inneholder fritekstmerknader som ‘ren tekst’.
- *objectblob.csv* – tabellen inneholder 341 rader, feltet ‘blobvalue’ er et blobfelt – konvertering viser at feltet inneholder grafikk (en logo).
- *subjectblob.csv* – tabellen inneholder 949 rader, feltet ‘subjectblobcontents’ er et blobfelt – konvertering viser at innholdet er fritekst som ‘ren tekst’.
- *studentreport.csv* – tabellen inneholder 3555 rader, feltet ‘blobnote’ er et blobfelt – konvertering viser at feltinnholdet er fritekstmerknader som ‘ren tekst.’

Det er opprettet kopier av disse tabellene under mappen SYSDOK\BLOB_felt\ . Konverteringen er gjort i et SAS-script.

Validering av fødselsnummer, D-nummer og dublettkontroll

SATS inneholder flere tabeller med fødselsnummer:

- *elevfravær.csv*
- *studentgrade.csv*
- *diploma.csv*
- *employeeaccount.csv*

- *invoices.csv*
- *person.csv*

Hovedtabellen er *person.csv*. Dette er en felles persontabell for både elever og ansatte. Tabellen inneholder 21 901 rader. Feltet *socialsecno* inneholder fødselsnummer/D-nummer (bare 30 D-nummer). Det er 9 686 rader som er blanke.

Fødselsnummerkontrollen er gjennomført på 12 215 rader fra denne tabellen med utfyllt innhold i feltet. Fødselsnummer og D-nummer er validert etter Modulus 11-algoritmen. Gyldige D-nummer passerer gjennom denne kontrollen. Validering av fødselsnummer er kjørt på en kortversjon av tabellen *person.csv* som inneholder følgende felt: *socialsecno*, *birthday*, *firstname*, *lastname*. Valideringen er kjørt i en SAS-macro. SAS-logen fra validering av fødselsnummer viser:

NOTE: There were 12215 observations read from the data set DATA.PERSON.

NOTE: The data set DATA.INVALID has 256 observations and 5 variables.

NOTE: The data set DATA.VALID has 11959 observations and 5 variables.

Det er 256 fødselsnummer som ikke validerer. Dette skyldes fiktive/ugyldige datoer, at personnummerdelen er blank, inneholder 00000 eller 99999 eller en kombinasjon av dette. Alt dette er autentiske feil bekreftet av arkivskaper. Tabellene er skrevet ut som hhv. *gyldigefnr.csv* og *ugyldigefnr.csv* og lagret under mappen SYSDOK. Det er tatt stikkprøver mot Folkeregisteret for kontroll av navn og (gyldig) fødselsnummer. Ingen feil er avdekket.

Deretter er det kjørt dublettkontroll på tabellen med gyldige fødselsnummer (*data.valid*), som viser at det finnes 77 dubletter blant observasjonene med gyldig fødselsnummer, jf. SAS-logen:

NOTE: There were 11959 observations read from the data set DATA.VALID.

NOTE: The data set DATA.DUBFRI has 11882 observations and 5 variables. NOTE: The data set DATA.DUBLETT has 77 observations and 5 variables.

Primærnøkkelen (*personoid*) i *person.csv* validerer. Fødselsnummer er ikke nøkkelfelt i noen av tabellene med fødselsnummer. Fødselsnummer vil imidlertid være det viktigste søkekriteriet for enkeltpersoner i en DIP⁴ fra systemet og i tillegg en nasjonal koblingsnøkkel som gjør det mulig å koble sammen mikrodata fra SATS med mikrodata i andre systemer fra andre arkivskapere for forskningsformål, f.eks.

Klausulering og brukertjenester

I hht. databehandleravtale mellom Stange kommune og IKA Øst av 12. november 2013 er IKA Øst databehandler og Stange kommune behandlingsansvarlig. Databehandleravtalens hensikt er å regulere rettigheter og plikter etter [Lov om behandling av personopplysninger \(personopplysningsloven\)](#), LOV2018-06-1538, ikrafttredelse 20.07.2018 og [Forskrift om behandling av personopplysninger](#), FOR2018-06-15-876, med ikrafttredelse 20.07.2018. Avtalen skal sikre at personopplysninger om de registrerte ikke brukes urettmessig eller kommer uberettigede i hende. Bestemmelsene er ikke til hinder for 'Behandling av personopplysninger for arkivformål i allmennhetens interesse, formål knyttet til vitenskapelig eller historisk forskning eller statistiske formål'- jf. [personopplysningslovens § 8](#), og en rekke andre unntak for samme formål i loven og i [Europaparlaments- og rådsforordning \(EU\) 2016/679](#)

⁴ Dissemination Information Package – bruksversjon basert på en eller flere SIP.

av 27. april 2016 om vern av fysiske personer i forbindelse med behandling av personopplysninger og om fri utveksling av slike opplysninger samt om oppheving av direktiv 95/46/EF (generell personvernforordning - GDPR).

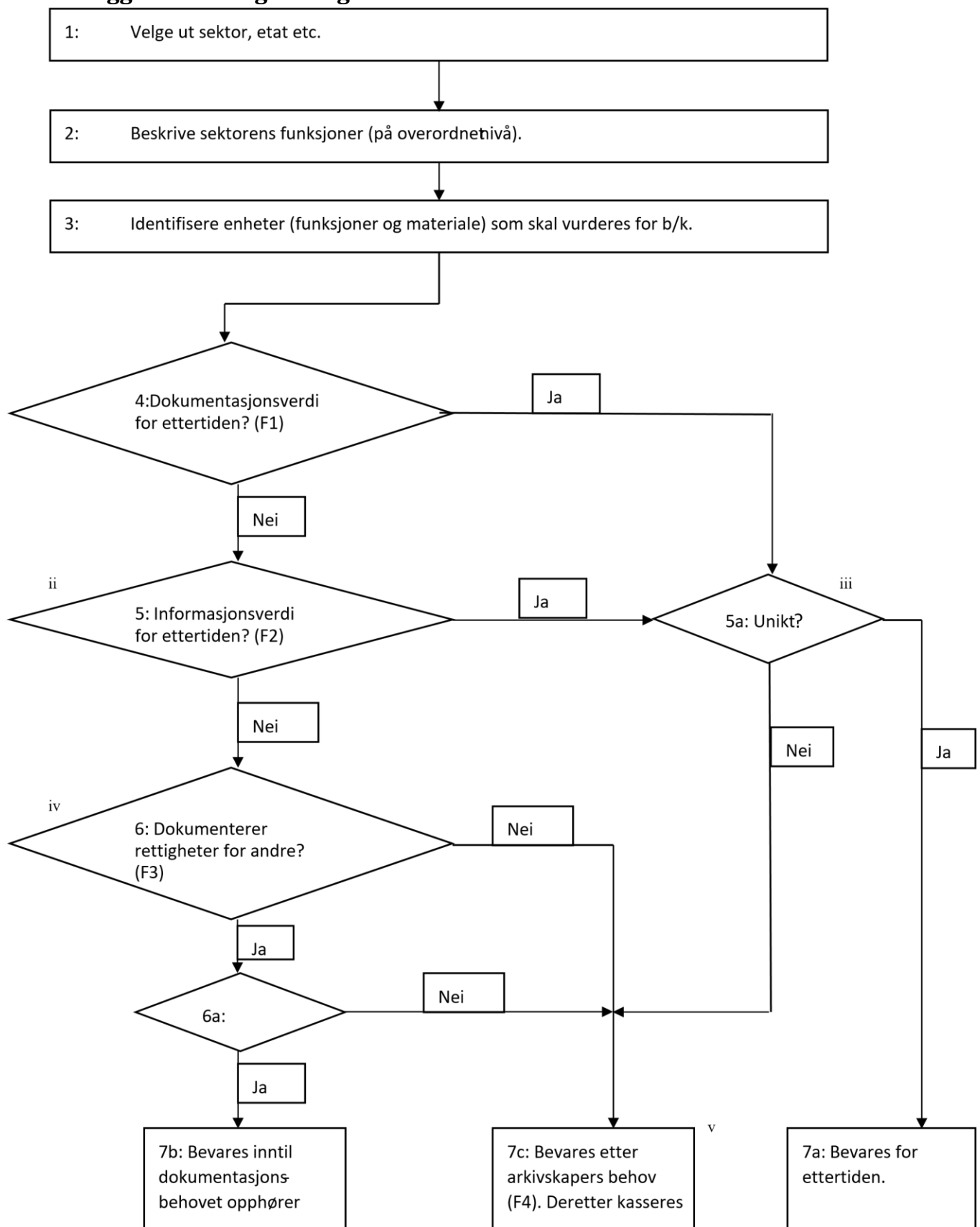
Så lenge arkivskaper beholder papirbasert dokumentasjon (som senere skal til arkivdepot, jf. over), må uansett arkivskaper ha muligheten til å betjene forespørsler fra en rekvirent (elev/tidligere elev). Brukertjenester vil i mange tiår fremover i praksis blir gjennomført ved at forespørsler fra arkivskaper (på egne vegne) besvares av IKA Øst. Dette er uansett klausulerte data, som selvsagt ikke kan publiseres. Publikumsforespørsler (partsinnsyn) må alltid gå via arkivskaper til IKA Øst, og samme vei tilbake, evt. direkte mellom IKA Øst og rekvirent dersom arkivskaper (behandlingsansvarlige) tillater. Dersom det viser seg å være stor etterspørsel etter visse typer informasjon, vil IKA Øst levere en DIP til arkivskaper, slik at arkivskaper kan betjene forespørslene på egenhånd. Elever/tidligere elever har krav på informasjon om seg selv, men ikke om andre. Det har ingen hensikt å lage generelle 'innsynsløsninger' som er åpne for publikum så lenge datagrunnlaget er klausulert. Den generelle klausuleringen vil i dette tilfellet være gjeldende til 2076. En DIP kan gjøres mer målrettet når det foreligger empiri for etterspørselen. Uansett må personvern- og klausuleringsbestemmelsene varetas.

For forskningsformål vil etterspørselen være helt annerledes enn hvis formålet er å dokumentere historikk for enkeltelever. Forespørsler fra forskere vil typisk besvares i form av 'skreddersydd', anonymiserte data (DIP), forutsatt at alle vilkår om tilgang til data for forskningsformålet er oppfylt. Tekniske metadata (addml-fil, ER-diagram mm.) er derimot katalogdata som fritt kan distribueres til forskere som grunnlag for å formulere en bestilling. Dette er imidlertid svært teknisk informasjon som krever betydelig datakompetanse for å bli forstått. Arkivdepotet må derfor være forberedt på å bistå forskere mht. å lese og forstå addml-filene.

Dokumentasjon av arkivskapers egne, vanlige (historiske) spørringer finnes i en del 'Views' som er tatt med i SIP. Disse vil likevel ha begrenset verdi med tanke på forskningens etterspørsel. Forskere (historikere og andre) vil i utgangspunktet stille helt andre og nye spørsmål til datagrunnlaget – spørsmål som *ikke* har vært stilt før. Det typiske vil derfor være at data vil brukes til helt andre formål enn de er innsamlet for. Forskere (som kan få tilgang etter unntaksbestemmelsene i Forvaltningsloven og Personopplysningsloven bl.a.) vil neppe søke opp en og en opplysning, men bestille et definert datasett (DIP) med en større eller mindre populasjon og utvalgte variable, normalt koblet sammen med variable fra andre fagsystemer og arkivskapere (skreddersydd datagrunnlag). Utgangspunktet kan f.eks. være at et forskningsprosjekt definerer en populasjon (en gruppe individer som oppfyller bestemte kriterier) med utvalgte variable fra et fagsystem (eks. SATS) og kobler dette sammen med de samme individene (dvs. de som får 'match') i ett eller flere andre fagsystemer/registre for å legge på analyse- og forklaringsvariabler som finnes spesifikt i andre systemer – eks. økonomiske variable, yrkeskoder, husholdningsdata etc.

Slike koblinger må utføres innenfor sikre soner i et arkivdepot, og det ferdige analysegrunnlaget (datasettet) må anonymiseres før utlevering til forskeren. Forskeren kan deretter analysere datasettet i sitt eget 'laboratorium'. Det vanlige er at slike datasett slettes når analysen er ferdig. Anonymiserte datasett kan uansett ikke kobles videre og brukes i annen forskning.

Vedlegg 1. Bevaringsutvalgets metodikk



Tilleggsriterier.

- Etterspørsel
- Symbol/identitetsverdi
- Estetisk verdi
- Pedagogisk verdi
- Gjenstandsverdi
- Brukbarhet
- Fysisk tilstand

ⁱ Administrativt nivå, saksbehandlingstype, saksbehandlingsledd, ekstraordinære/ordinære aktiviteter, pionervirksomhet, primær-/internfunksjon i virksomheten

ⁱⁱ Tidsspenn/kontinuitet, omfang, informasjonstetthet, tematisk variasjon, lenkbarhet med annet arkivmateriale, kvalitative egenskaper og alder ⁱⁱⁱ Dvs. spørsmålet om redundans. ^{iv} Saksbehandlingens konsekvenser, hjemmelsgrunnlag ^v Administrative og driftsmessige behov