

Gjennomstrømming i FS-marten

Innholdsfortegnelse

Gjennomstrømming i FS-marten.....	1
Innholdsfortegnelse	1
Endringslogg	1
Innledning/bakgrunn	1
Forkortelser/akronymer	1
Dataflyt.....	2
Hva skjer på veien fra FS til DV?	3
GSP.....	3
GST	5
Semesterdimensjonen.....	5
Avstemming mellom FS og DV	5
Åpne spørsmål.....	6

Endringslogg

Dato	Versjon	Beskrivelse	Endret av
03.03.2011	0.1	Første versjon basert på workshop med Pål Erik og Royne 2.3	Stein Aske
08.03.2011	0.2	Utvidet med analyse ++	Stein Aske
15.03.2011	0.3	Utviklet alternativ 2 videre	Stein Aske
17.01.2012	0.4	Utvidet med innpassing (3 nye kilder til GSP)	Stein Aske
22.08.2012	0.5	Korrigert tidsdimensjonen i forbindelse med migrering til PowerCenter	Stein Aske
12.12.2014	0.6	STAR fase 2: Dokumentet er oppdatert med endringer som er gjort siden 2012	Stein Aske
08.04.2015	0.7	Endret kriteriet for tidsdimensjon på kvalifikasjoner (forskerutdanning flyttes ikke)	Stein Aske
18.08.2015	0.8	Rettet opp noen småfeil under åpne spørsmål	Stein Aske
21.11.2016	0.9	Fjernet GEM og tilhørende tabeller	Stein Aske

Innledning/bakgrunn

STAR-prosjektet har bestemt at gjennomstrømmingsdelen i LIST skal gjennomgå og evt. redesignes. En arbeidsgruppe bestående av Pål Erik Megaard, Royne Skaarer Kreutz og Stein Aske er satt til å starte dette arbeidet. Dette dokumentet er resultatet.

Etter innkjøp av Tableau som nytt rapporteringsverktøy for FS/DV er det behov for å oppdatere dette dokumentet, som kan være av nytte for brukere/utviklere som ønsker å forstå hvordan gjennomstrømmingstabellene i FS-marten er bygget opp.

Forkortelser/akronymer

Mart-tabeller (DV):

GSP: FAK_GJENNOMSTR_STUDPROG

GSI: FAK_GJENNOMSTR_STUDPROG_IND

GSK: FAK_GJENNOMSTR_STUDPROG_KULL

GST: FAK_GJENNOMSTR_STUDENT

STK: FAK_STARTKULL

SVP: FAK_STUDENTVURDKOMBPROTOKOLL

Kildetabeller (FS):

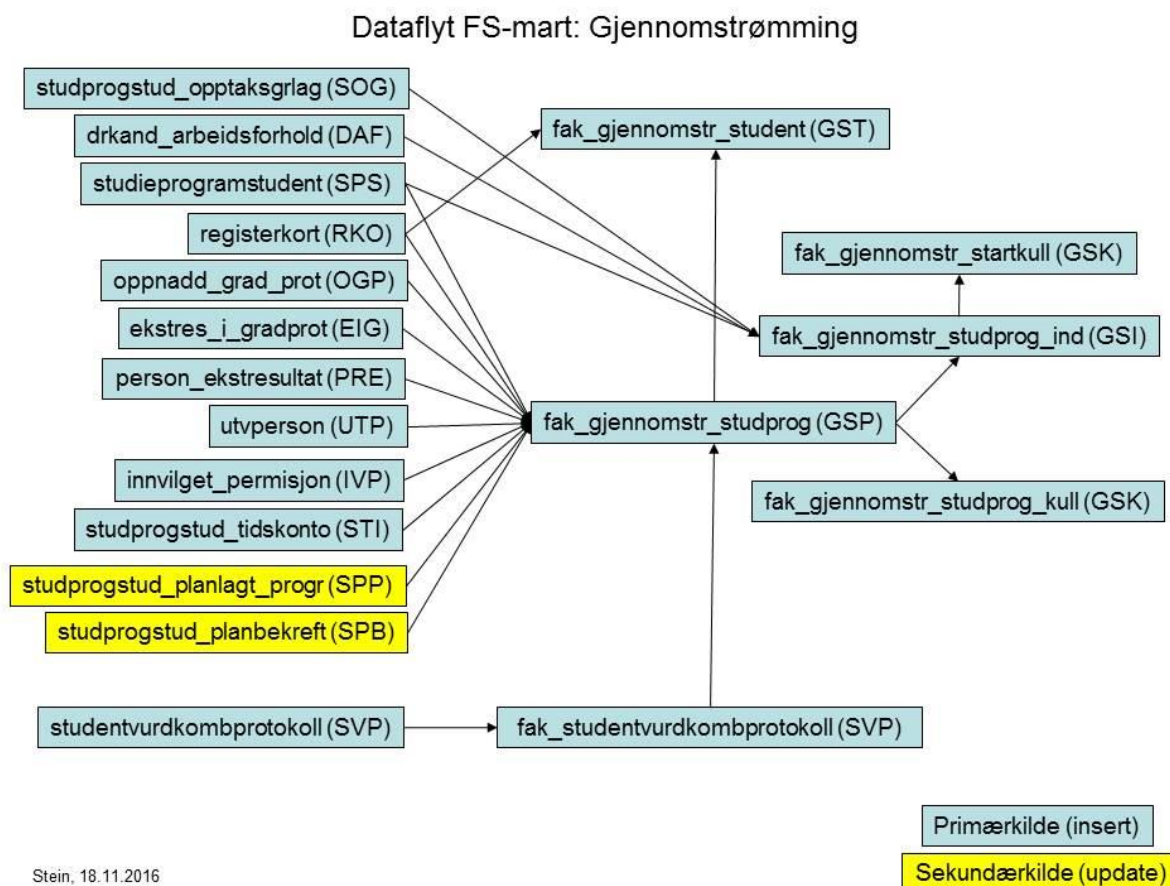
DAF: DRKAND_ARBEIDSFORHOLD
EIG: EKSTRES_I_GRADPROT
IVP: INNVILGET_PERMISJON
OGP: OPPNADD_GRAD_PROT
PRE: PERSON_EKSTRESULTAT
RKO: REGISTERKORT
SVP: STUDENTVURDKOMBPROTOKOLL
SPS: STUDIEPROGRAMSTUDENT
SEK: STUDPROGSTUD_EKSTRES
SPB: STUDPROGSTUD_PLANBEKREFT
SPP: STUDPROGSTUD_PLANLAGT_PROGR

Andre tabellforkortelser brukt i dokumentet:

EIS: EMNE_I_STUDIEPROGRAM
PER: PERSON
VKE: VURDKOMBENHET
VKO: VURDERINGSKOMBINASJON
VTI: VURDERINGSTID

Dataflyt

Figuren viser dataflyt fra FS til DV.



Hva skjer på veien fra FS til DV?

Nedenfor er de viktigste transformasjonene beskrevet, inkludert filtere og joins.

GSP

EIG/PRE:

- Filter: ARSTALL >= 1990.
- Join: EIG er drivende tabell. Denne joines med OGP og PRE. Hvis det er flere forekomster av samme grad (dvs. med samme program) på samme student så brukes den siste (med høyest dato).
- Transformasjoner: Innpassede poeng hentes fra EIG der disse er ulik NULL, ellers fra PRE (som er dominerende).
- Grader oppnådd i januar, juli og august legges til foregående semester.

SVP:

- Transformasjoner:
 - Studiepoeng aggregeres
 - Snittkarakter beregnes
 - Antall emner beregnes

IVP:

- Filtre:
 - to_number(to_char(DATO_FRA,'YYYY')) between 1990 and 2020
 - to_number(to_char(a.DATO_TIL,'YYYY')) between 1990 and 2020.
Øvre grense er satt pga noen få rader med høye/feilaktige årstall.
- Join med PER for å erstatte FODSELSDATO/PERSONNR med PERSONLOPENR.
- Transformasjoner:
 - Permisjonsfaktor = Permisjonsprosent *
Permisjonsvarighet/Terminvarighet. Permisjonsfaktor kan variere fra 0 til 1 (der 1 betyr 100% permisjon i hele semesteret).

OGP:

- Filtre:
 - to_char(DATO_OPPNAELSE,'YYYY') between 1990 and 2030. Øvre grense er satt pga noen få rader med høye/feilaktige årstall.
 - STUDIEPROGRAMKODE is not null. 10% mangler studieprogram (som er vår nøkkel). Er det mulig å tette disse hullene?
- Join med PER for å erstatte FODSELSDATO/PERSONNR med PERSONLOPENR.
- Transformasjoner:
 - Høyeste/laveste grad beregnes ved hjelp av studienivå (høy kode => høyt nivå). Hvis det er flere grader pr. student/termin/program tas den høyeste graden med.
 - Grader oppnådd i januar, juli og august legges til foregående semester, med unntak av forskerutdanning (studienivå 900-999), som beholder semester fra dato_oppnaelse.

RKO:

- Filtre:
 - TERMINKODE in ('VÅR','HØST'). Mister 12 rader (promille av promille)
 - STATUS_BET_OK = 'J'. Mister en promille.

- STATUS_REG_OK = 'J'. Mister en promille.
- ARSTALL >= 1990.
- STUDIEPROGRAMKODE_RAPPORTERING is not null. Mister 20%, men program er vår nøkkel.

SPB:

- Filter: ARSTALL_BEKREFT >= 1990
- TERMINKODE_BEKREFT in ('VÅR','HØST'). Burde denne erstattes med standardisert termin? Mister kun 0,1 promille.

SPP:

- Filtre:
 - ARSTALL >= 1990
 - TERMINKODE in ('VÅR','HØST'). Burde denne vært fjernet og erstattet med standardisert termin?
 - PROSENTTALL_AV_HELTID < 100. Prosent 100 er uinteressant/default. Hvorfor legges disse inn???
- Transformasjoner:
 - PROGRESJONSFAKTOR =
sum(PROSENTTALL_AV_HELTID)/count(*)

SPS:

- Filter:
 - STUDENTSTATKODE <> TRUKKET
 - STUDIERETTSTATKODE <> PRIVATIST
 - ARSTALL >= 2000.
- Join med PER for å erstatte FODSELSDATO/PERSONNR med PERSONLOPENR.
- Join med RKO for å filtrere vekk ikke-aktive studenter.

STI:

- Filtre:
 - to_number(to_char(a.DATO_FRA,'YYYY')) between 1990 and 2020
 - to_number(to_char(a.DATO_TIL,'YYYY')) between 1990 and 2020. Fjerner ugyldige datoer.
 - TIDSKONTODETALJTYPEKODE in ('ARBPLIKT','UTENOM'). Permisjon hentes fra IVP.
 - PROSENTTALL <= 100. Prosent > 100 anses som feilaktig.

UTP:

- Filtre:
 - to_number(to_char(DATO_FRA,'YYYY')) between 1990 and 2020
 - to_number(to_char(a.DATO_TIL,'YYYY')) between 1990 and 2020. Det er en håndfull utvekslinger med datoer > 2020. Antagelig er de feilregistrert.
 - STATUS_INNREISENDE = 'N'. Bare utveskling ut av institusjonen skal tas med.
 - STUDIEPROGRAMKODE is not null. 1% av utvekslingene mangler program, som er vår nøkkel.

- UTVPROGRAMKODE != 'INDIVID'. Individuelle avtaler tas ikke med.

GST

GSP:

- Transformasjoner:
 - Tallfelt aggregeres opp til studentnivå.

RKO:

- Filtre:
 - ARSTALL >= 1990
 - TERMINKODE in ('VÅR','HØST'). Mister noen få rader her.

Semesterdimensjonen

Knytting til tidsdimensjonen er ganske uoversiktlig. Tabellen nedenfor er derfor ment å gi en oversikt over hvilke felter i kildetabellene som brukes for å knytte seg til DIM_SEMESTER.

Tabell	Kolonne	Brukes for å knytte opp DIM_SEMESTER	Brukes for å brette ut flere terminer
SVP	ARSTALL/VURDTIDKODE joines mot VURDKOMBENHET. Fra denne brukes ARSTALL_REELL/VURDTIDKODE_REELL. som joines med ARSTALL/VURDTIDKODE i VURDERINGSTID. Fra denne brukes ARSTALL_GJELDER_I/ TERMINKODE_GJELDER_I.	x	
SPS	ARSTALL_START/TERMINKODE_START DATO_STUDIERETT_GYLDIG_TIL (hvis denne er NULL dagens dato)	x x	fra til
IVP	DATO_FRA DATO_TIL	x x	fra til
OGP	DATO_OPPNAELSE (Januar/Juli/August settes til foregående termin) ARSTALL_NSDRAPPORTERT/ TERMINKODE_NSDRAPPORTERT	x x	
RKO	ARSTALL/TERMINKODE	x	
SPP	ARSTALL/TERMINKODE	x	
EIG	DATO (Januar/Juli/August settes til foregående termin)	x	
PRE	Mangler. Må joines med EIG.	x	

Avstemming mellom FS og DV

Generelt: Avstemming av de tre gjennomføringstabellene mot de respektive kildetabellene i FS er komplisert, av flere grunner:

- DV-tabellene har ikke de samme nøklene som kildetabellene. Det er derfor nødvendig å aggregere/deduplisere kilderadene før lasting til DV. Dermed blir antall rader ulikt.
- DV-tabellene bruker i flere tilfeller andre år/termin-felter enn i kildetabellene. Dette gjøres dels for standardisering av terminer (kun VÅR/HØST), dels for samordning av terminbegrep på tvers av kiletabellene. Kapitlet om semesterdimensjonen beskriver hvordan dette foregår. Dette vanskeliggjør avstemming.

SVP:

- Antall studiepoeng fra SVP kan avstemmes, så lenge man tar hensyn til tidsdimensjonen (at det ikke er år/termin i SVP som brukes, men i VTI).

GSP:

- Skal kunne avstemmes mot SVP, siden det er en ren aggregering som skjer.
- Avstemming mot IVP/OGP skal kunne gjøres, men datofeltene i kilden kompliserer.

GST:

- Skal kunne avstemmes mot GSP, siden det er en ren aggregering som skjer.
- Skal også kunne avstemmes mot RKO, så lenge man tar hensyn til filtrene.

Åpne spørsmål

Nr	Spørsmål	Svar	Status
1	Er SVP/VPR tilstrekkelig som kilde på emnenivå?	Foreløpig ja, men dette må diskuteres nærmere. SSE og EDS er også tatt med som kilder.	OK
2	Kan vi generelt begrense oss til år > 2003?	Har brukt >= 2000	OK
3	Hvordan identifisere emnestudenter? Via undervisning?	Ja, det er de som ikke har studierett på programmet, men har undervisningmeldt seg	OK
4	Det er uavklart hvordan studiepoeng skal aggregeres opp til GST: Skal alt legges på hovedprogram, eller bare de poengene som matcher?	Alle studiepoeng aggregeres fra GSP til GST, uansett om disse tilhører hovedprogram eller ikke.	OK
5	Bør terminkoder standardiseres før kildetabellene joins? Det kan se ut som om de ikke er brukt konsistent på tvers av tabellene.	Ja	OK
6	Kan vi gjenbruke standardterminene i LIST?	Pål Erik: Ja. De stemmer med DBH-standarden.	OK
7	SPP har terminfilter. Bør erstattes med standardisert termin.		Åpen

8	10% av gradene mangler program. Disse filtreres vekk. Kan noe gjøres for å få dem med?		Åpen
---	--	--	------