

Vejledning til CSV Converter

Arkiveringsversion

Med arkiveringsversion menes et elektronisk arkivalie, der følger Rigsarkivets nugældende regler for aflevering og bevaring, jf. Bekendtgørelse nr. 342 af 11. marts om arkiveringsversioner af bevaringsværdige data fra elektroniske arkivsystemer. Med datafiler menes de filer / tabeller, der indeholder udtræk fra arkivskaberens oprindelige system.

Formål

CSV Converter er et program beregnet til at omforme datafilerne i en arkiveringsversion til tegnseparerede filer, så datafilerne er klar til at kunne indlæses i standardprogrammer, fx et regnearksprogram eller et statistikbehandlingsprogram.

Denne vejledning giver ikke nogen anvisning på hvordan tegnseparerede filer indlæses i et konkret standardprogram eller på den videre bearbejdning af data.

CSV står for comma separated values

Tekniske forudsætninger

For at anvende programmet kræver det at programmet kører på et WINDOWS styresystem.

Download af program

CSVConverter ligger i en selvudpakkende fil, som kan downloades til lokalt drev på den pc, du skal arbejde på (eller køres direkte fra det sted, hvor du åbner filen).

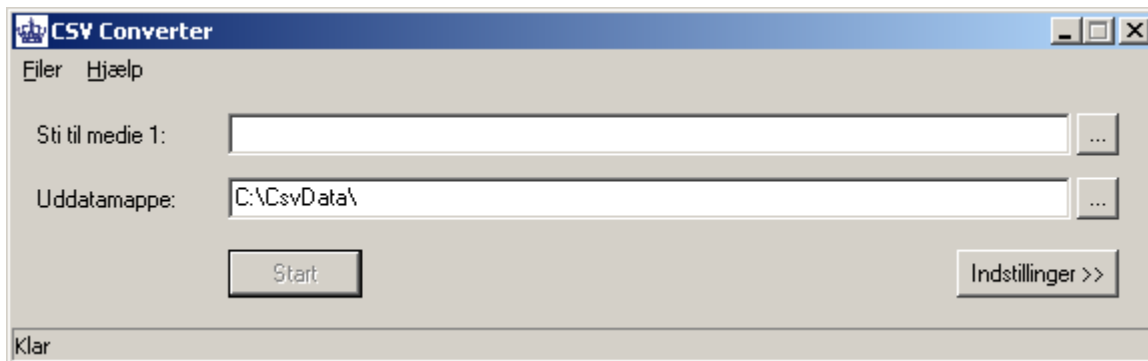
Kopiér arkiveringsversionen til lokalt drev. Det er kun arkiveringsversioner, som består af ét medie, der kan indlæses direkte fra det pågældende medie. Det er derfor lettest - uanset antallet af medier - at kopiere hele arkiveringsversionen til lokalt drev og arbejde videre derfra.

Sådan bruger du CSV

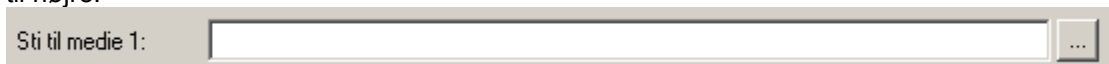
Hvis du har downloadet programmet til lokalt drev, vil det kunne åbnes via Startmenuen. Programmet starter med dette vindue:



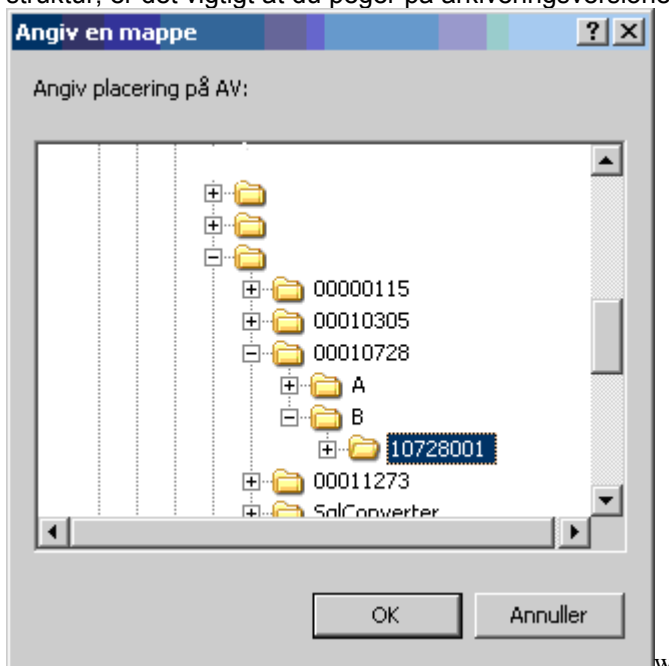
Derefter åbner CSV Converters programvindue – klar til at vælge den arkiveringsversion, du ønsker at genrejse:



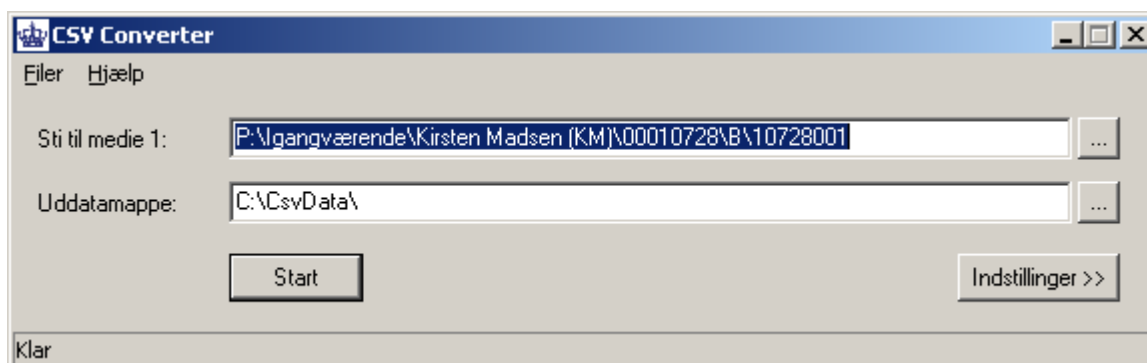
Start med at finde den arkiveringsversion du ønsker at konvertere til tegnseparerede filer via filvalgsknappen til højre:

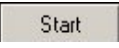


Når du skal pege på en arkiveringsversion, der ligger lagret i Rigsarkivets gældende bevaringsformat- og struktur, er det vigtigt at du peger på arkiveringsversionens 1. medie, som eksemplet her viser:



Arkiveringsversionsnr. 00010728 med 1. medie angivet som 10728**001**.
001 angiver at det er første medie ud af det samlede antal, 002 er arkiveringsversionens 2. medie og så videre.



Tryk på  for at køre programmet.

Når indlæsningen er afsluttet, dukker dette vindue op:

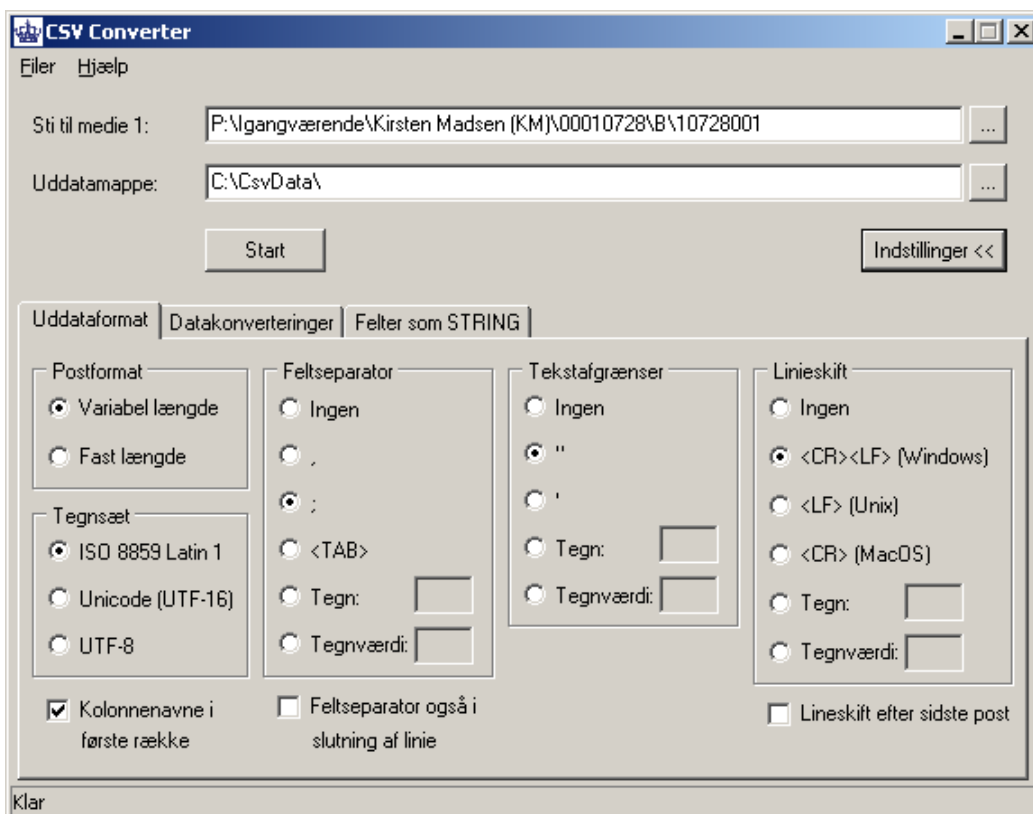


Indlæsningen er foretaget med programmets standardvalg. Hvis du har behov for at indstille efter andre kriterier, kan du læse mere om valg af feltseparator, tegnsæt mv. i næste afsnit.

Sådan bruger du Indstillinger

Hvis du har behov for at lave særlige indstillinger for konverteringen, kan du gøre det via Indstillinger. Tryk

på . Under Indstillinger>> får du følgende valgmuligheder:



Indstillinger består af de tre faneblade Uddataformat, Datakonverteringer og Felter som STRING.

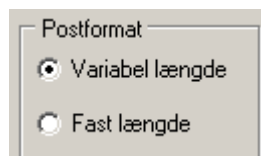
Fanebladet Uddataformat:

POSTFORMAT

Under postformat vælger du om Uddataformatet (eksportformatet) skal være med Variabel eller Fast længde.

Uddataformatet er forvalgt til Variabel Længde. Formålet med programmet er at konvertere til tegnseparerede filer, der per definition har variabel feltlængde.

Hvis den oprindelige arkiveringsversion er lagret i variabel længde¹, kan man kun konvertere til variabel feltlængde i CSV Converter.



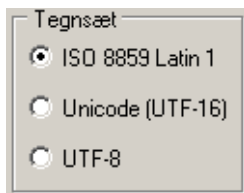
¹ Rigsarkivets krav til variabel feltlængdestruktur er baseret på at den variabel feltlængde angives i en header foran feltet. Der benyttes således ikke tegnseparator af nogen art.

Fast længde bruges yderst sjældent, men programmet understøtter principielt muligheden. Postformat som

Fast længde, indebærer at der ikke må benyttes et separatortegn af nogen art, jf. 

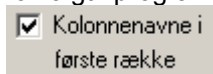
TEGNSÆT:

Her vælger du hvilket tegnsæt uddata skal have:



ISO 8859 Latin 1 er det tegnsæt, som Rigsarkivets gældende krav af bekendtgørelse nr. 342 af 11. marts 2004 foreskriver. Programmer og/eller styresystemer understøtter forskellige tegnsæt, og derfor afhænger dette valg af det konkrete system, du agter at benytte til behandling af dataene. Programmet understøtter udover ISO 8859 Latin 1 også Unicode (UTF 16) eller UTF-8, der kan vælges som uddata format i forbindelse med konverteringen.

Nederst er det angivet om uddata-filen skal tage kolonnenavnet med i første række. Denne funktion er forvalgt i programmet, jf. fluebensmarkering.



I indlæsningseksempelet med arkiveringsversionsnr. 10728, består tabellen Projekt af 14 felter. Hvis du vælger kolonnenavne i første række får du feltnavnene på de 14 felter med over i første række.

Se eksempel på valg af formater med udgangspunkt i nedenstående dataeksempel:

Post	Tal	Tekst	Dato	Tid
1	27	Tekst "linie"; Nr.1	20061201	1630
2	4.7	Tekst "linie"; Nr.2	20061231	1700

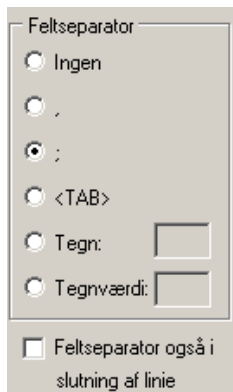
Hvis kolonnenavne er valgt, giver det følgende uddata (sammen med øvrige standardvalg), med feltnavne i første linie:

Linie	TABEL.CSV
1	"Tal";"Tekst";"Dato";"Tid"
2	27;"Tekst ""Linie""; Nr.1";01-12-2006;16:30
3	4,7;"Tekst ""Linie""; Nr.2";31-12-2006;17:00

Feltseparator

I denne funktion kan du vælge mellem de mest udbredte separator tegn eller selv vælge et andet tegn eller en decimal tegnværdi 0...255.

Den mest udbredte tegnværdi er semikolon, der er forvalgt som standard .



Ingen medfører at de enkelte felter ikke kan adskilles, og bruges kun til konvertering til fast feltlængde struktur.

Feltseparator også i slutning af linie

Denne funktion benyttes afhængigt af de programmer, du skal benytte. En feltseparator i slutningen af linien, angiver en postafslutning.

Med udgangspunkt i vort eksempel, vises her hvordan data ser ud, hvis du vælger Feltseparator i slutning af linie:

Linie	TABEL.CSV
1	"Tal"; "Tekst"; "Dato"; "Tid";
2	27; "Tekst ""Linie""; Nr.1"; 01-12-2006; 16:30;
3	4,7; "Tekst ""Linie""; Nr.2"; 31-12-2006; 17:00;

TEKSTAFGRÆNSER:

I denne funktion kan du vælge mellem de mest udbredte tekstafgrænsere eller selv vælge et andet tegn eller decimal tegnværdi 0...255.



Den mest udbredte tegnværdi er anførelsestegn ", der er valgt som standard.

Tekstafgrænser benyttes for at undgå at forveksle dataindhold i STRINGdata med separator tegn. I ovenstående eksempel indeholder linien Tekst "Linie"; Nr.1 et semikolon, som kan forveksles med feltseparatoren.

LINESKIFT:

Lineskift

Ingen

<CR><LF> (Windows)

<LF> (Unix)

<CR> (MacOS)

Tegn:

Tegnsværdi:

Lineskift efter sidste post

Lineskift markeres på forskellig vis afhængig af styresystem.

- <CR><LF> (Windows) Carriage Return Linefeed benyttes til at markere lineskift i WINDOWSs styresystemer.
- <LF> (Unix) Linefeed benyttes til at markere lineskift i Unix styresystemer.
- <CR> (MacOS) Carriage return benyttes til at markere lineskift i MacOS styresystemer.

Hvis der vælges Lineskift efter sidste post, vil uddata fra vort eksempel se sådan ud:

Linie	TABEL.CSV
1	"Tal";"Tekst";"Dato";"Tid"
2	27;"Tekst ""Linie""; Nr.1";01-12-2006;16:30
3	4,7;"Tekst ""Linie""; Nr.2";31-12-2006;17:00
4	<i>Tom linie</i>

Fanebladet Datakonverteringer

The screenshot shows a dialog box titled 'Datakonverteringer'. It is divided into two main sections. The left section, 'Tekstafgrænser i STRING data', contains radio buttons for 'Bevar', 'Fjern', and 'Dublér' (which is selected), along with input fields for 'Tegn:', 'Tegnsværd:', and 'Præfiks:'. The right section, 'Dato og tidsformater', contains three rows: 'DATE' with format 'DD-MM-CCYY' and example '30-11-2006', 'TIME' with format 'hh:mm:ss,s' and example '11:50:18,597', and 'TIMESTAMP' with format 'DD-MM-CCYY hh:mm:ss,s' and example '30-11-2006 11:50:18,597'. Below this is a section for 'Decimalseparator (REAL, EXP)' with radio buttons for 'Komma .' (selected) and 'Punktum .'.

Tekstafgrænser i STRING data

Dette valg giver mening når der vælges en tekstafgrænser. Fra det foregående eksempel har vi et tekstfelt med indholdet *Tekst "linie"; Nr.1*, som indeholder samme tegn som anvendes som tekstafgrænser. Nedenfor er vist konsekvenserne af de enkelte valg.

Bevar	"Tekst "linie"; Nr.1"
Fjern	"Tekst linie; Nr.1"
Dublér	"Tekst ""linie""; Nr.1"
Præfiks (f.eks. \)	"Tekst \"linie\"; Nr.1"

Dublér er valgt som standard. Du kan også vælge et helt andet erstatningstegn, dvs. tegn, en decimal tegnværdi 0...255 eller præfiks, hvor du bibeholder tegnet ved at foranstille et særligt tegn, som vist i eksemplet ovenfor er præfiks ofte /-tegnet (back slash).

Dato og tidsformater

De accepterede formater er afhængige af både de programmer som anvendes (samt lokale opsætninger på PC'en) Der kan vælges mellem forskellige dato og tidsformater. Følgende mest udbredte dato- og tidsformater er prædefinerede i programmet:

Format	Uddata
(uændret)	<i>Som i den oprindelige arkiveringsversion</i>
CCYYMMDD	20061207
DD-MM-CCYY	07-12-2006
M/D/CCYY	12/7/2996
hhmmss,s	123011,27
hh:mm:ss,s	12:30:11,27
hh:mm:ss	12:20:11

Herudover kan brugeren selv angive en dato eller tidsform, jf. nedenstående tabel. NB: Forskel på store og små bogstaver.

Tegn	Betydning
CC	Århundrede

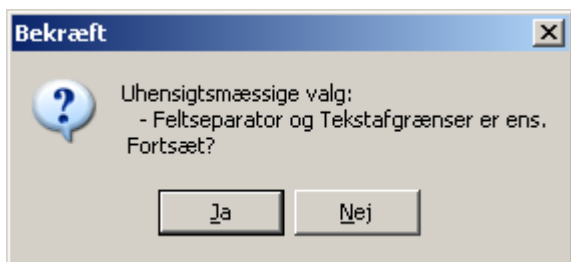
YY	Årstal (2cifre)
MM	Måned; Altid to cifre
M	Måned; Et eller to cifre
DD	Dag; altid to cifre
D	Dag; Et eller to cifre
hh	Time; Altid to cifre
mm	Minut; Altid to cifre
ss	Sekund; Altid to cifre
s	Decimal af sekund; vilkårligt antal cifre

Øvrige tegn opfattes som separator tegn. Datoeksempel vises automatisk i programmet (til højre for format), under indtastningen.

Decimalseparator (REAL, EXP)

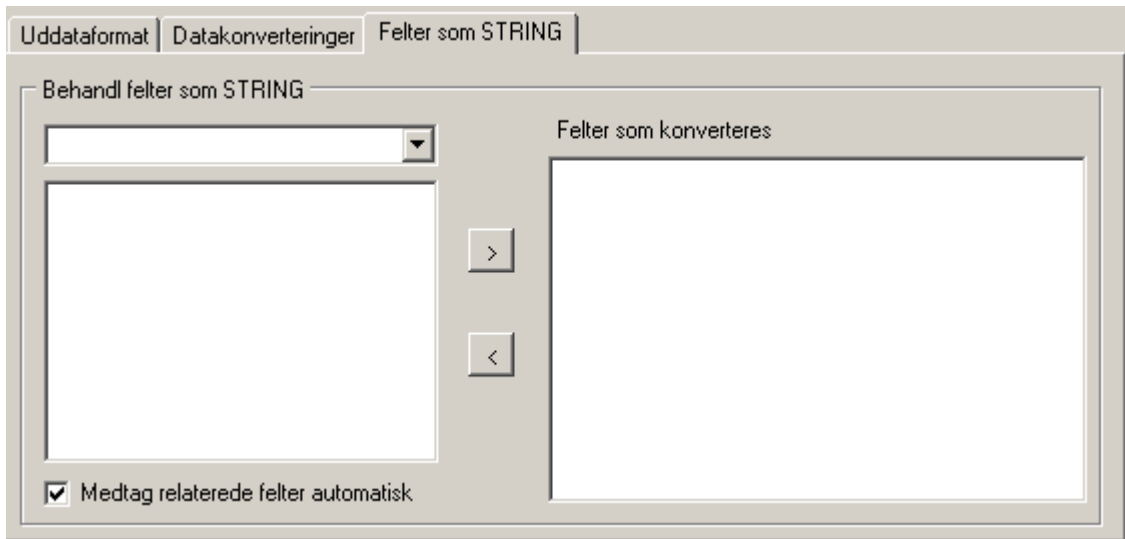
Afhængigt af både de programmer som anvendes (samt lokale opsætninger på PC'en) kan der vælges mellem *komma* (dansk) og *punktum* (amerikansk) som decimaltegn i tal.

Under fanebladet Datakonvertering, er der mulighed for mange forskellige valg af uddataformater. Da nogle af disse valg ikke giver god mening, f.eks. hvis både *Feltseparator* og *Tekstafgrænser* er valgt til semikolon giver det nogle data som sandsynligvis ikke kan indlæses. Af denne årsag advares brugeren om u hensigtsmæssige valg inden konverteringen starter, i dette eksempel fås en advarsel.



Faneblad Felter som STRING

Fanebladet Felter som STRING kan benyttes, hvis der findes nogle datatyper i tabellernes felter, som ikke overholder den pågældende datastandard.

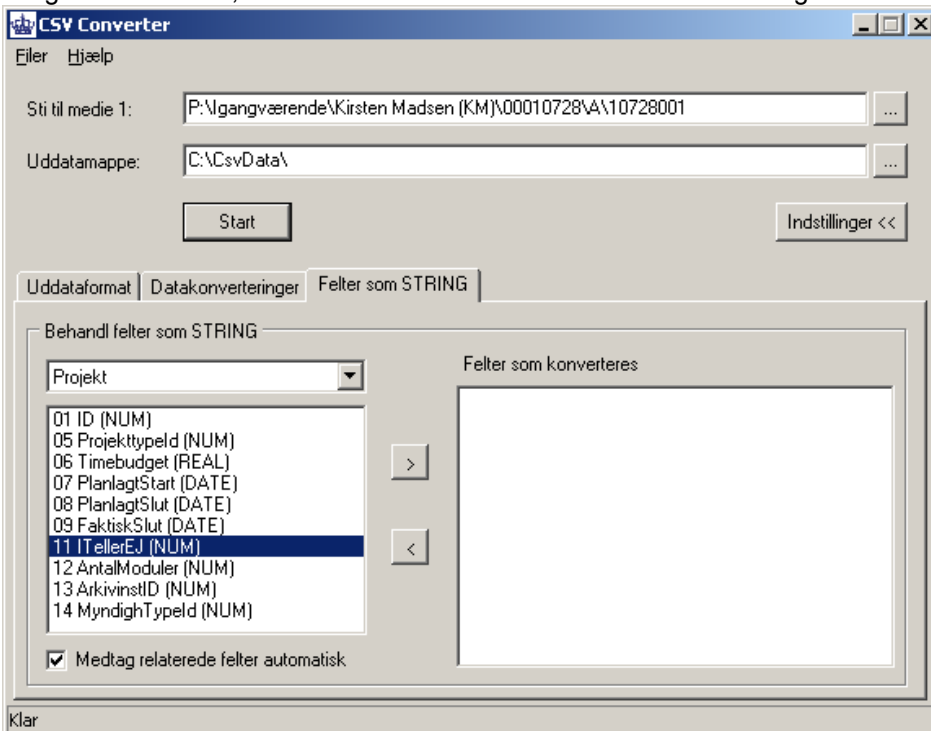


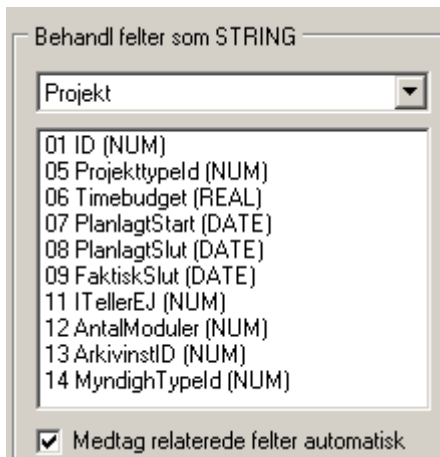
Du kan ikke på forhånd vide, at der findes sådanne data, men det vil fremgå af indlæsningsloggen, når indlæsningen er gennemført første gang.

Dropdownmenuen behandl felter som STRING vil indeholde navne på de tabeller, som indeholder felter, der er defineret som datatyperne NUM, REAL, EXP, DATA, TIME eller DATETIME. Du kan nu vælge at behandle denne type felter som STRING.

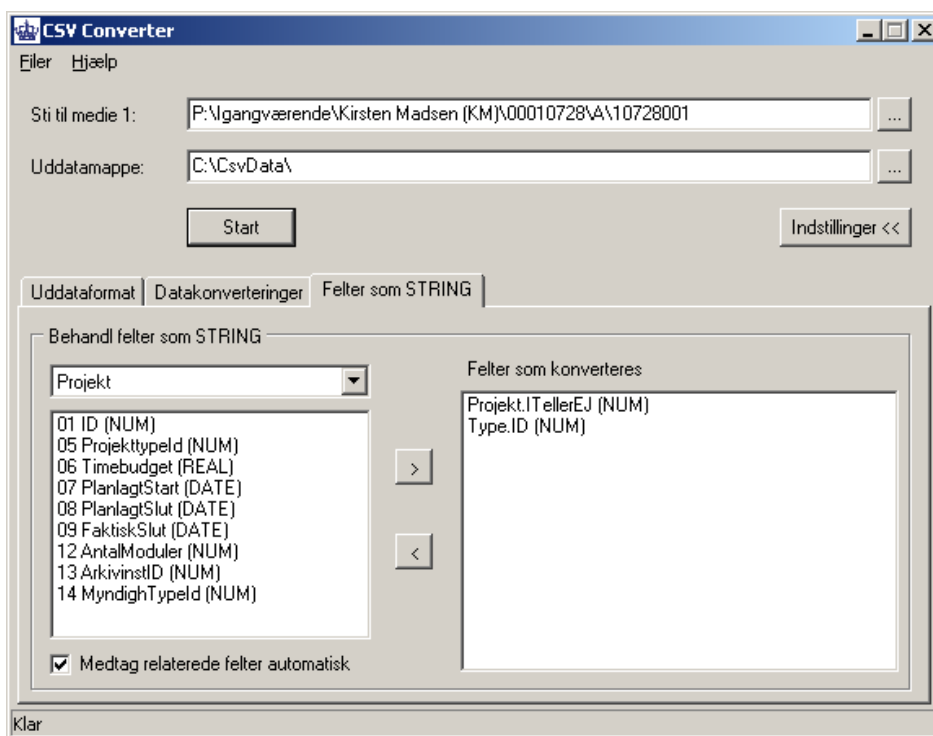
Sådan vælger du hvilke felter, der skal konverteres til STRING

Med dropdownmenuen "Behandl felter som STRING" vælger du først den pågældende tabel og dernæst vælger du det felter, der skal behandles som STRING i listeboksen lige nedenunder:





Bemærk at "medtag relaterede felter automatisk" er valgt default. Det betyder, at hvis et felt, der konverteres til STRING, er relateret til et felt i en anden tabel, konverteres dette felt også til STRING. Hvis dette ikke gøres, vil relationen ikke blive oprettet i databasen. Eksemplet her viser at flytter vi feltet 11 ITellerEJ [NUM] bliver feltet Type:ID [NUM] også konverteret til STRING.



Indlæsningslog i CSV

I indlæsningsloggen kan du bl.a. se hvilke indstillinger, der er benyttet ved konverteringen til en tegnsepareret fil. Indlæsningsloggen bliver gemt som CSVConv.log.