Guide on Creating an Information Package with Statistical Data from Spreadsheets and CSV Files

Guide on creating an information package containing research data from spreadsheets and CSV files that comply with the provisions on Schedule 9 of the Executive Order on Information Packages.

The Danish National Archives, July 2020

Version 1.0

0. How to read this guide

Public authorities, including research institutions, are obliged to provide a copy of data and documents of preservation-worthy research data. If the research data to be delivered is from statistical files or similar (e.g. spreadsheets), the submission format is called an information package. State authorities must submit to the National Archives. Municipalities and regions can choose whether to submit data to the National Archives or set up their own archives.

The National Archives has laid down several provisions for an information package for the preservation and future use of data. All authorities must comply with them when submitting to the National Archives. These provisions are described in the Danish National Archives' Executive Order on Information Packages, Schedule 9: Information package for certain types of research data.

The National Archives has developed the ASTA (Aflevering af Statistikfiler Til Arkiv – translation: Submission of Statistical data to the Archives) program to create and test an information package. ASTA can only be used to extract data from statistical files such as SAS, Stata and SPSS to an information package. However, data used for statistical analysis can also be found in other formats (e.g. spreadsheets or CSV files) that are extracted from other programs.

This guide describes how to create an information package if your preservation worthy statistical data is found in a spreadsheet or a CSV file.

A. Guide's target audience and an application

This guide is aimed at those who need to provide statistical data from spreadsheets and CSV files in the form of an information package. It can, for instance, be the institution's data manager, IT department, the researcher himself or an external IT supplier.

B. Other guides

In addition to this guide, the Danish National Archives has prepared other guides that that are relevant to the production and submission of information packages:

- Quick guide - for the production and testing of an information package with ASTA
- Guide to Schedule 9, in the Executive Order on Information Packages
- Guide to the program Skab ArchiveIndex
- Guide to the program Skab ContextDocumentationIndex
- Guide to converting to TIFF format
- UTF-8 Guide
- Sample information package with statistical data FD.18005

All guide materials can be found on the National Archives' homepage www.sa.dk.

C. Law and legislation

Information about the related legislation can be found on the National Archives' homepage www.sa.dk.

D. Definitions

Information packages with data from statistical file(s) in general consists of context documents – which should be submitted in archival formats designated by the Archives -, the extracts of data and metadata

from the statistical files under the submission, and the two index files in xml format containing metadata about the data and the context documents submitted.

Archival formats: The National Archives (Rigsarkivet) accepts 6 archival formats: TIFF, JPEG2000, MP3, WAV, MPEG2 and MPEG4.

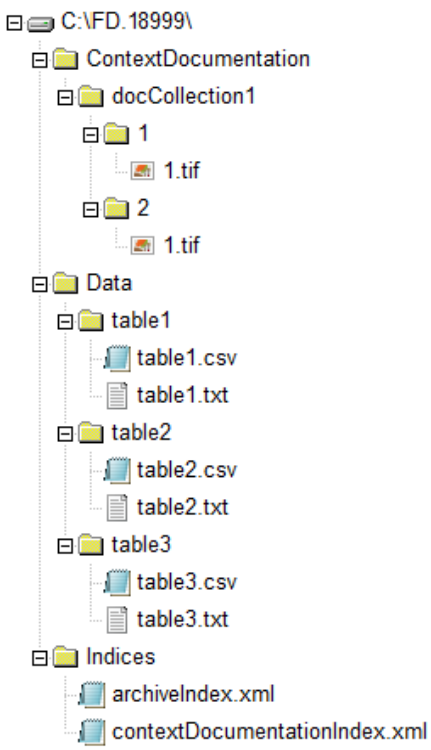
1. What is an information package?

As part of the National Archives' requirements for the submission of statistical datasets, data from the original statistical files must be extracted to an information package and tested before submission to the archives.

The information package must follow some specific requirements regarding folder structure, naming folders, files, data types, variable names, variable labels, format names, missing values and so on. All provisions for an information package are listed on Schedule 9, in the Executive Order on Information Packages.

An example of an information package is given in Figure 1.1.

Figure 1.1 Graphic overview of the elements and structure of an information package



The **ContextDocumentation** folder contains documents converted to preservation format, for example TIFF, that describe the data contained in the information package (E.g. how data is collected, methods report or a questionnaire).

The **Data** folder contains both a data file and a metadata file, both of which comply with Schedule 9. Data extracted from the original statistical files must be delivered in a semicolon-separated csv file **(table1.csv).** Metadata extracted from the statistical file (e.g. variable descriptions, response categories and missing value codes) must be submitted as a metadata file in txt format (**table1.txt**)

The **Indices** folder contains two index files with metadata at a more general level. For example, **the archiveIndex.xml** file contains information, such as the name of the dataset being delivered, the name of the researcher who collected data, the data coverage period, access restrictions to data, etc. The **contextDocumentationIndex.xml** file contains information about the context documents placed in the ContextDocumentation folder, such as the documents title, author and a subject categorization of the document.

2. How to manually create an information package

A description of how to create an information package by extracting data from one or more spreadsheets can be seen below. See section 3 if you need to create an information package from data that is already in CSV format.

E. Check list - steps before creating the information package

You should complete the following steps before you create the information package.

Nr.	Tasks	Finished
	General	
1. New folder with copy of material	Create a new folder on your computer with a copy of all data and documents to be provided, that is, statistical files to be submitted, context documents, shown in the archival provisions (*afleveringsbestemmelsen*) and two index files. Rename the folder with the information package serial number, shown in the archival provisions (*afleveringsbestemmelsen*).	
	Data	
2. Data files	The data files must be spreadsheets (e.g. . xls or .xlsx) or CSV file extracted from another program. This section of the guide primarily deals with extracts from spreadsheets. If you have a CSV file extracted from another program, see section 3.	
3. Variable in data files	Variable names (column names) must be placed in the first row of the spreadsheet. Variable names cannot be longer than 128 characters, start with a number or include blank spaces or special characters, such as %.	
4. The Variable category/type	Column variable categories/data type must be assigned according to the type of data they display in a cell/column. For example, if a cell contains dates, select the category "Date" with the yyyy-mm-dd format. In Excel, you can change the data type by selecting all relevant cells in the column and right clicking. Choose "Format Cells" > the "Number" tab > Select the category/data type that matches the contents of the cells. Here you can also define custom formats. See Figures 2. 1 - 2.7 for a description of allowed data formats/categories.	
5. Value labels	If there are codes among values in the data, these must be explained in a metadata file in the information package. Therefore, have these code descriptions ready. A description on how to specify the code in the metadata file is found in section M 'Create the metadata file (.txt)'.	
6. Missing values	If you have used missing value codes in your data, such as 9 = undisclosed, 10 = irrelevant, or 11 = dropout, these values should be described in a metadata file in the information package. A description on how to manually set codes for missing values in the metadata file is found in section M 'Create metadata file (.txt)'. Codes for missing values may only be used in numerical and categorical variables. Contact the archives if you are applying	

	codes for missing values in variables with the text, time or date types. All assigned codes for missing value should also be assigned a value in a code list in the metadata file.	
7. References (in submissions of more than one dataset)	When there are several datasets in the information package that have a relation with each other, the reference (merging variable) should have the same type/format and length across the datasets	
CONTEXT DOCUMENTATION		
8. Approval from the Archives	The context documentation file contextDocumentationIndex.xml should be approved by the Archives prior to their inclusion in the information package. The file specifies all the additional documents that are to be included in the information package. See 'Guide to the program Skab contextDocumentationindex'.	
9. TIFF documents	Each document registered in the context documentation file should be converted to TIFF or another format accepted by the Archives. See 'Guide to converting documents to TIFF'.	
ARCHIVE DESCRIPTION FILE		
10. Approval from the Archives	The archival description file should be approved by the Archives prior to their inclusion in the information package. The file includes the general metadata related to the delivered data. See 'Guide to the program Skab archiveIndex'.	

Data formats allowed in the data file

See Figure 9.3 in Schedule 9 of the Executive Order on Information Packages, the following data formats are permitted in an information package including statistical data.

Integer (int) as referred to in DS/ISO 6093:1985 (NR1) standard, corresponding to the ***Number*** category with no decimals in Excel. Note that the integer is a number with no decimal, so that the number of decimals should be set to 0.

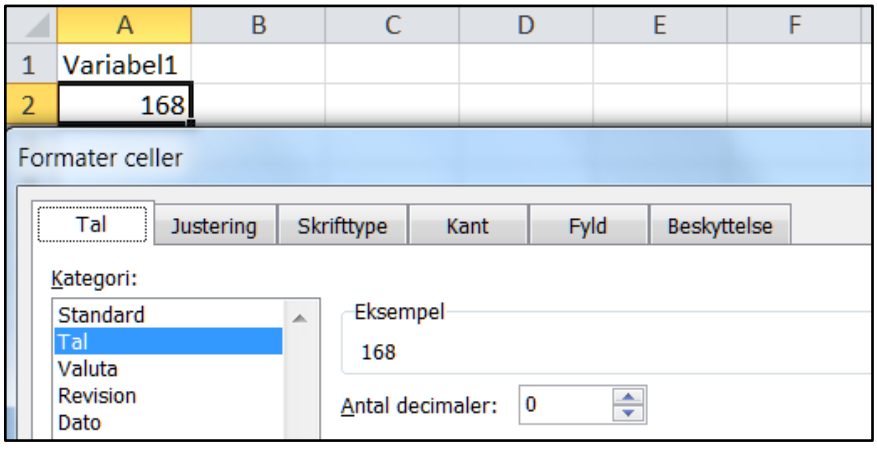


Figure 2.1 Choice for category 'Number' in Excel

Decimal (decimal) as referred to in DS/ISO 6093:1985 (NR2) standard, corresponding to ***Number*** in Excel
Note that the decimals is a number with decimals, but the data type may also include an integer with no decimals.

Note that the 1000-separator *must not* be applied.
Note that the decimal separator can be either '.' or ', '

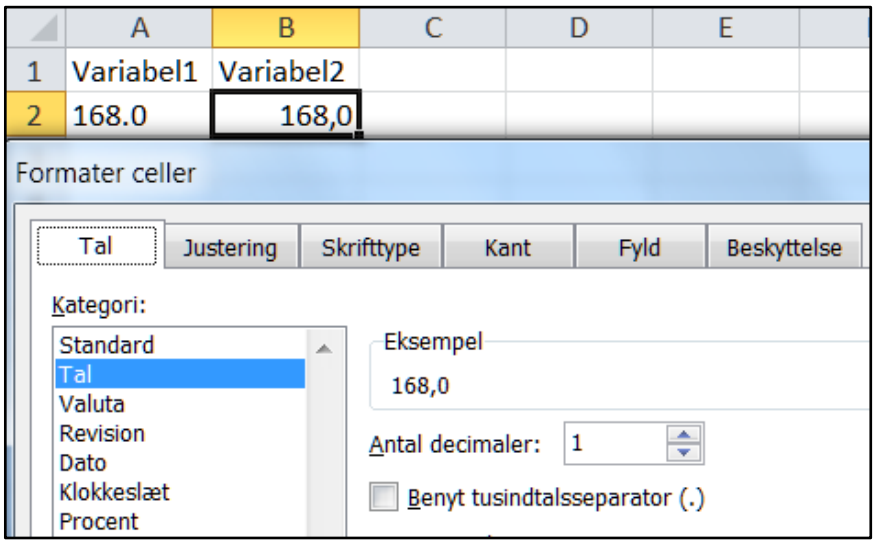


Figure 2.2 Choice for category decimal number in Excel

Text (string) corresponding to ***Text*** in Excel.

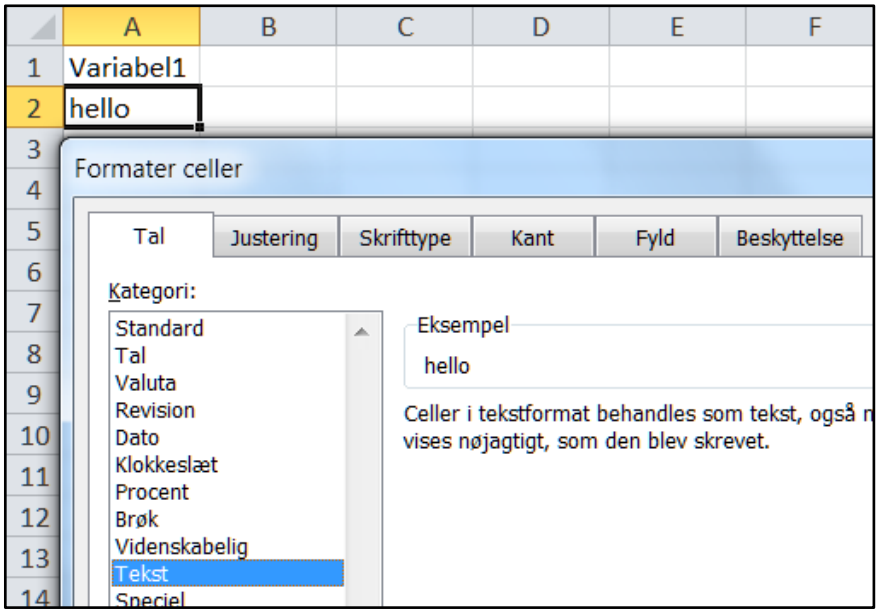


Figure 2.3 Choice of category text in Excel

Date (date), as referred to in DS/ISO8601:1993 standard, corresponding to ***Date*** in Excel
Note that the format is CCYY-MM-DD, e.g. 2019-11-30

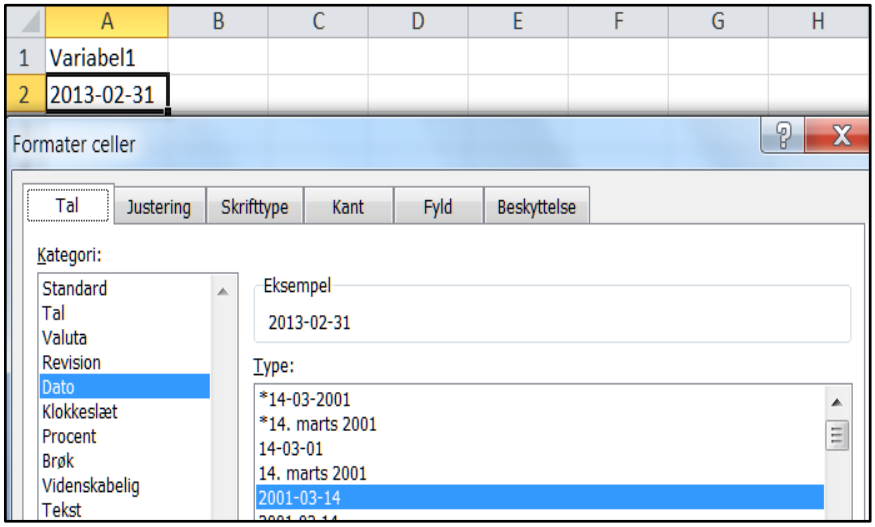


Figure 2.4 Choice of category date in Excel

Time (hour), as referred to in DS/ISO8601:1993 standard, corresponding to ***Time*** in Excel.
Note that the format is HH:MM:SS, e.g. 13:30:55

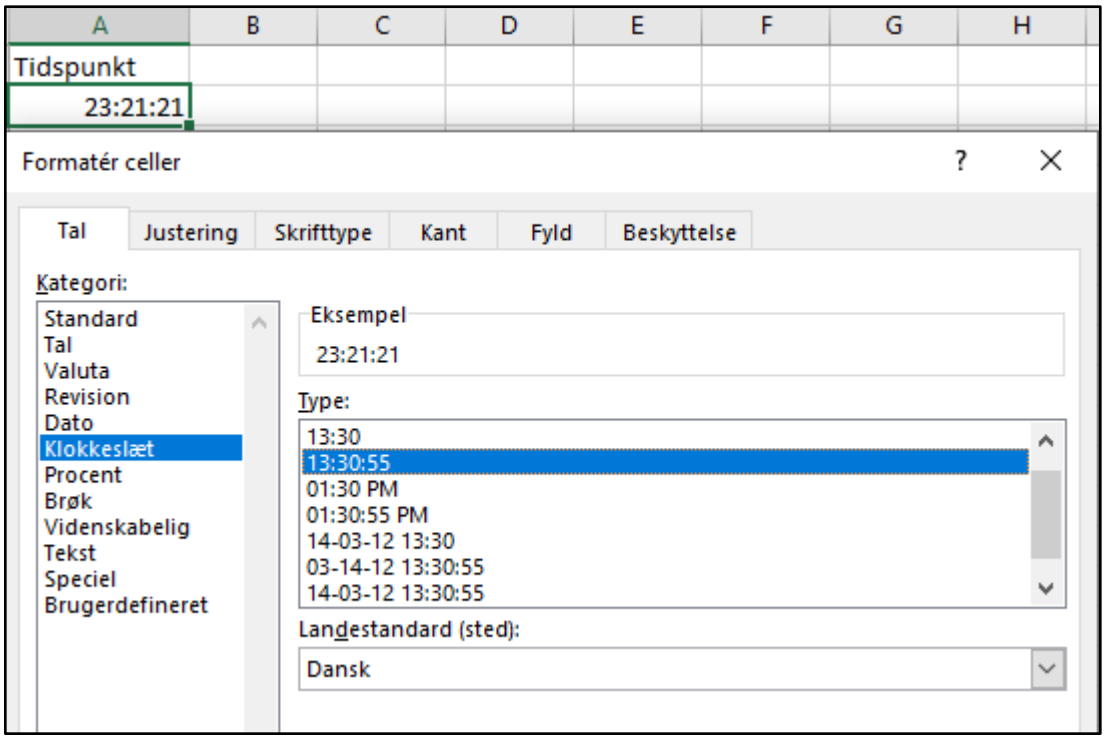


Figure 2.5 Choice for category Time in Excel

Timestamp (datetime), as referred to in DS/ISO8601:1993 standard.

Note that the format is CCYY-MM-DD HH:MM:SS.ssssss or CCYY-MM-DDTHH:MM:SS.ssssss

Note that entering milliseconds is optional (.ssssss).

Note that Excel does not have a pre-defined category that complies with the allowed timestamp format. Therefore, you must custom define a category yourself, one that meets this format, as shown below.

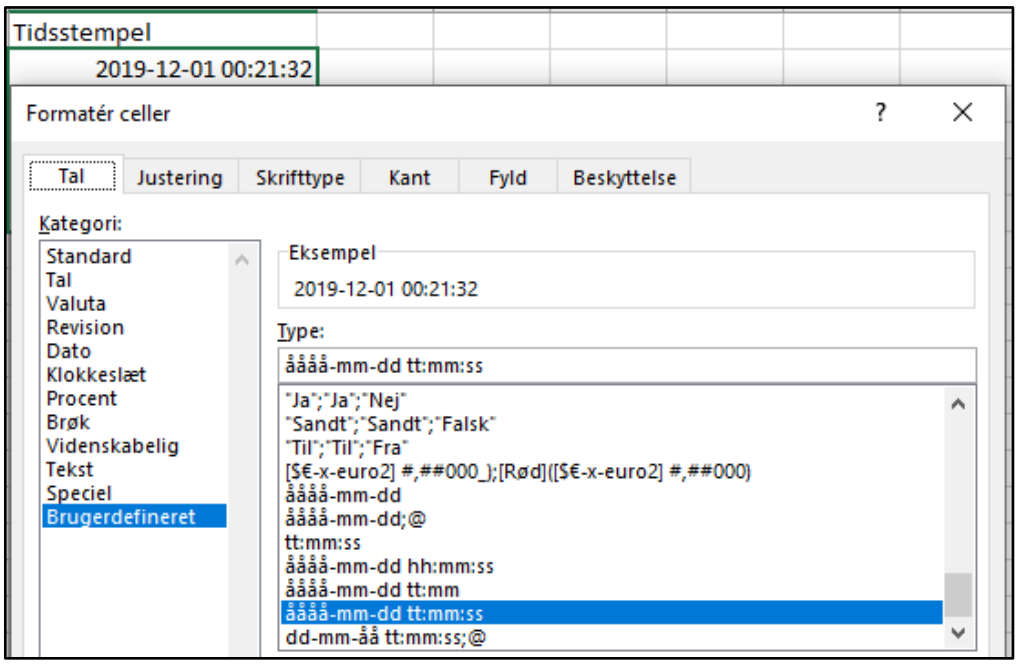


Figure 2.6 Definition of a user-defined timestamp category (with both date and time) in Excel

Note that Excel can specify a maximum of 2 digits in milliseconds, yyy-mm-dd tt:mm:ss.ss. You specify it as follows:

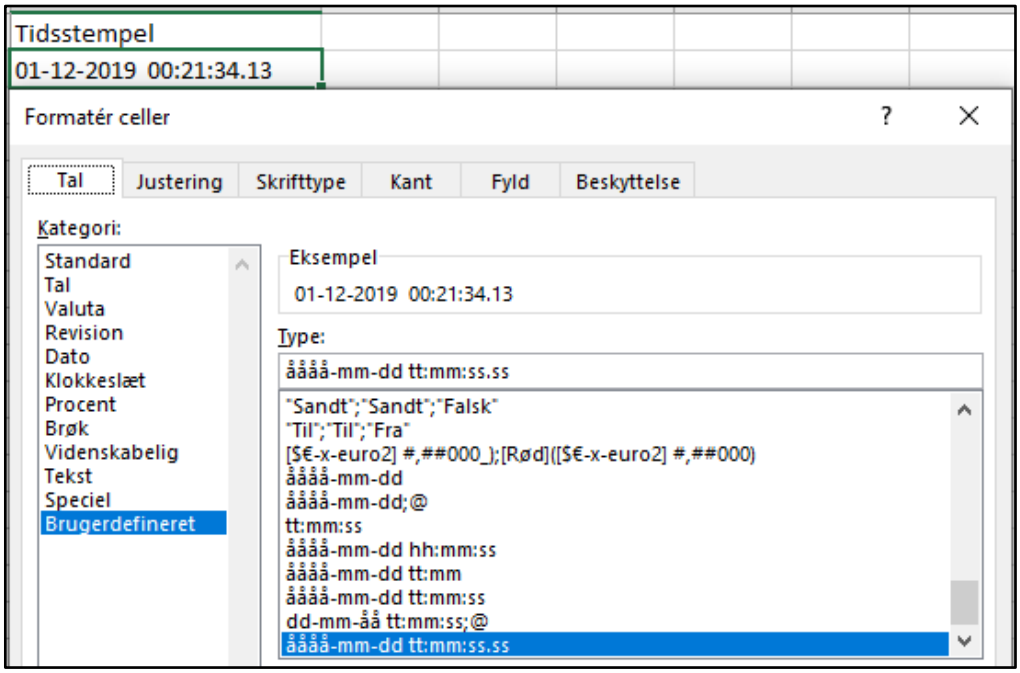


Figure 2.7 Definition of a user-defined timestamp category with milliseconds (both date and time) in Excel

F. Information package contents

An information package must consist of the following elements, as shown in Figure 1.1:

- Folder structure
- Index files archiveIndex.xml and contextDocumentationIndex.xml, located in the Indices folder
- Context documentation folder with context documents converted to TIFF
- Data files saved as CSV file(s), located in the data folder
- Metadata files with defined structure saved as txt file(s), located in the Data folder

You can create the information package's folder structure and place the index files and context documents in the information package using the ASTA program. When you use ASTA, this program makes sure to create folders in the appropriate structure and correctly name folders and files according to information package requirements. You can download ASTA from the National Archives' homepage www.sa.dk. Then, extract the data files, create the metadata files and place those files correctly in the information package in the Data folder.

1. Download ASTA program from the National Archives' homepage www.sa.dk
2. Click on AstaInstaller.exe to start ASTA

The sections below describe step-by-step how to create the different parts of the information package.

G. Create the information package folder structure

Use ASTA to create the outer folder structure in the information package. Select the menu item "Create information package and complete ***ONLY*** the 'Mappestruktur' (translation: Folder structure) session, see Figure 2.8.

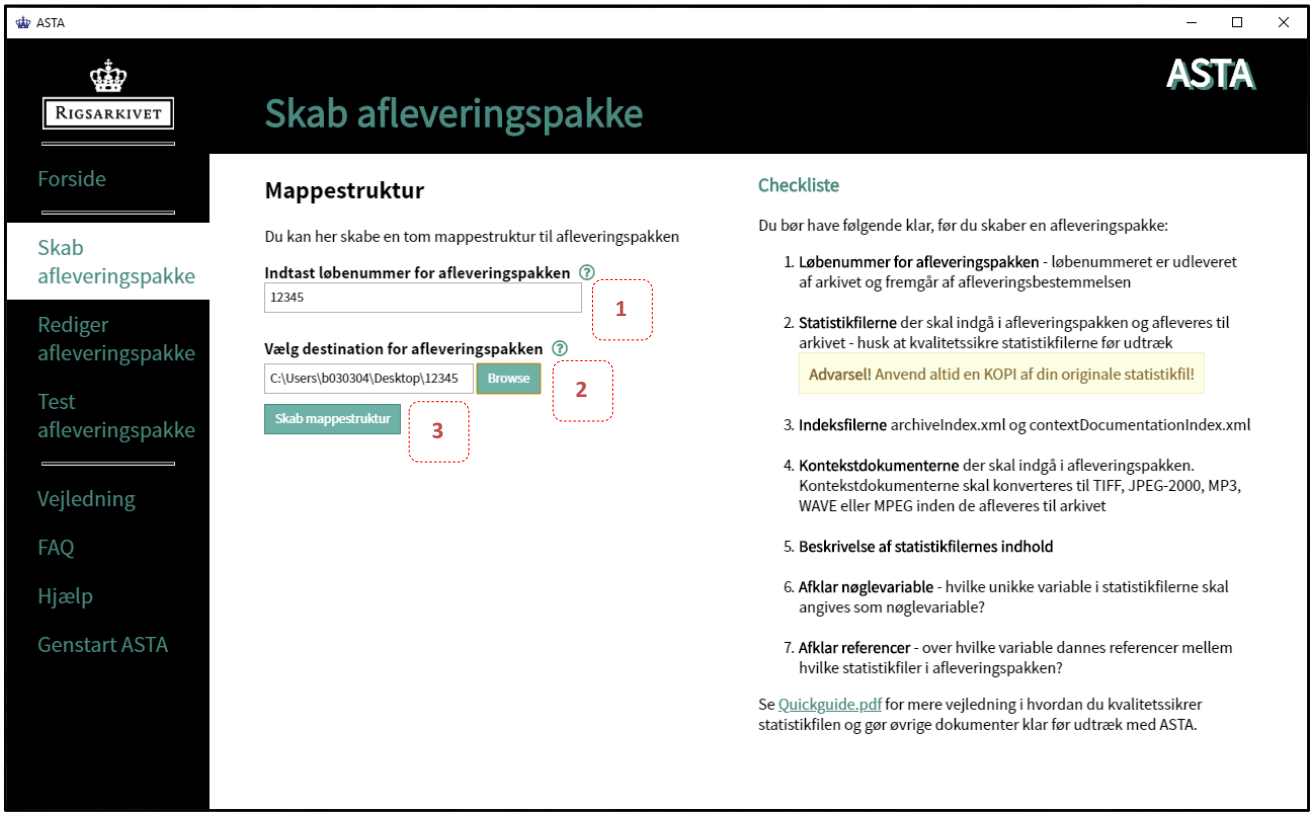


Figure 2.8: "Mappestruktur" (translation: Folder structure) page under the "Skab afleveringspakke" (translation: Create information package) in ASTA

1. Enter the **serial number** of the information package, in accordance with the serial number assigned by the Archives and shown in your Archival provisions (*afleveringsbestemmelse*),e.g. 12345
2. Select the destination where you want to save the information package by clicking on the '**Browse**' button and finding the correct folder.
3. Click the '**Skab mappestruktur'** (translation: create folder structure) button to create the submission folder structure.
   NOTE that you should not continue with the other steps in the program after this point.

After you click the "Skab mappestruktur" (translation: create folder structure) button (Figure 2.8), select the menu item "**Rediger afleveringspakke**" (translation: edit the information package) to place index files and context documents in the information package. The following page in Figure 2.9 appears, where you can choose the information package you have just created e.g. FD.12345.

Figure 2.9: "Tilføj indeksfiler og kontekstdokumenter" (translation: add index files and context documents" page under "Rediger afleveringspakke" (translation: edit the information package) in ASTA

1. Select the information package to be edited (e.g. FD.12345) by clicking on the '**Browse**' button
2. Click the '**Næste'** (translation: next) button to go to the next page

H. Place the index files in the information package

By clicking '**Næste'** (translation: next) in Figure 2.9, the page shown in Figure 2.10 appears. In this page you can place the index files *archiveIndex.xml* and *contextDocumentationIndex.xml* in the information package.

Figure 2.10: "Placer indeksfiler i afleveringspakken" (translation: place the index files in the information package*) under the menu item "Rediger afleveringspakke" (translation:* edit the information package*) in ASTA*

1. Select the archiveIndex.xml file using the '**Browse**' button
2. Select the contextDocumentationIndex.xml file using the '**Browse**' button.
3. Click the **'Næste'** (translation: next) button to place the selected files in the information package.

I. Place the context documents in the information package

By clicking **'Næste'** (translation: next) in Figure 2.10, the next page appears, as shown in Figure 2.11.

Here you can place context documents in the information package or replace existing context documents with new ones.

1. All context documents registered in the selected contextDocumentationIndex.xml are displayed in a document list. Click the '**Browse**' button next to each document to select the context document that corresponds to the document title in the document list.
   NOTE: You can place the documents, even if they are not converted to required file format e.g. TIFF. You can convert them later from the information package folder structure. You can also leave ASTA at this point and add them later, after you have converted them.
2. When you have finished placing all the context documents, we recommend that you print the document list on the **'Udskriv dokumentliste'** (translation: print the document list) button and verify whether the contents of the selected documents match the document title. When you click the **'Udskriv dokumentliste'** (translation: Print the document list) button, your document list appears in your default browser, e.g. Edge (see Figure 2. 12). Use the browser print function to print the list (see Figure 2.13). The document list is also automatically saved in an html version, where the selected information package is located when you click the **'Udskriv dokumentliste'** (translation: Print the document list) button.
3. Click the **'Næste'** (translation: next) button to place the selected documents in the information package. A screen with the confirmation that the information package is finished appears.
   NOTE: Your information package is not finished yet. You must place data and metadata files in the information package, before testing it with the ASTA **"Test afleveringspakke"** menu option.

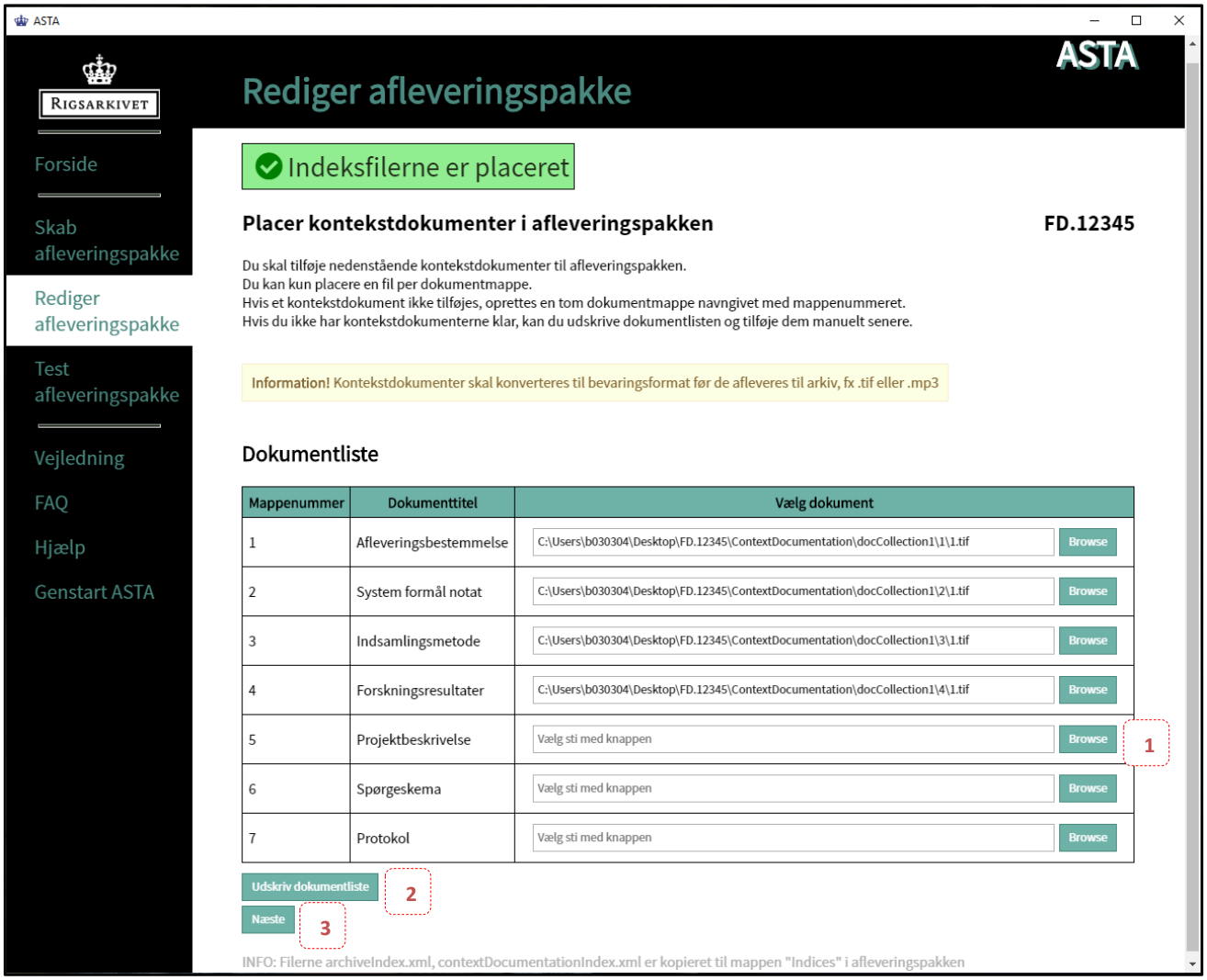


*Figure 2.11: Page "Placer indeksfiler i afleveringspakken" (*translation: place the index files in the information package*) under the menu item "Rediger afleveringspakke" (*translation: edit *the information package) in ASTA*

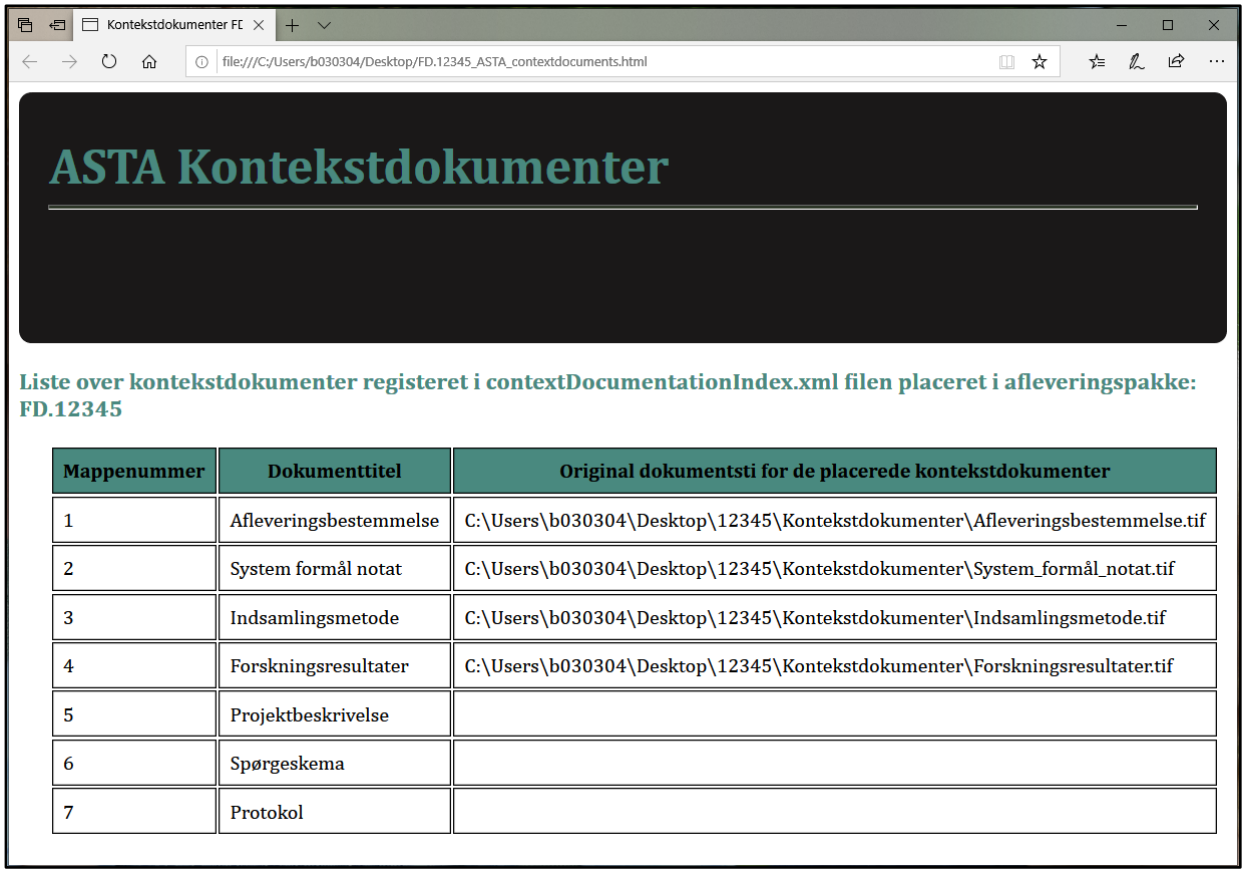
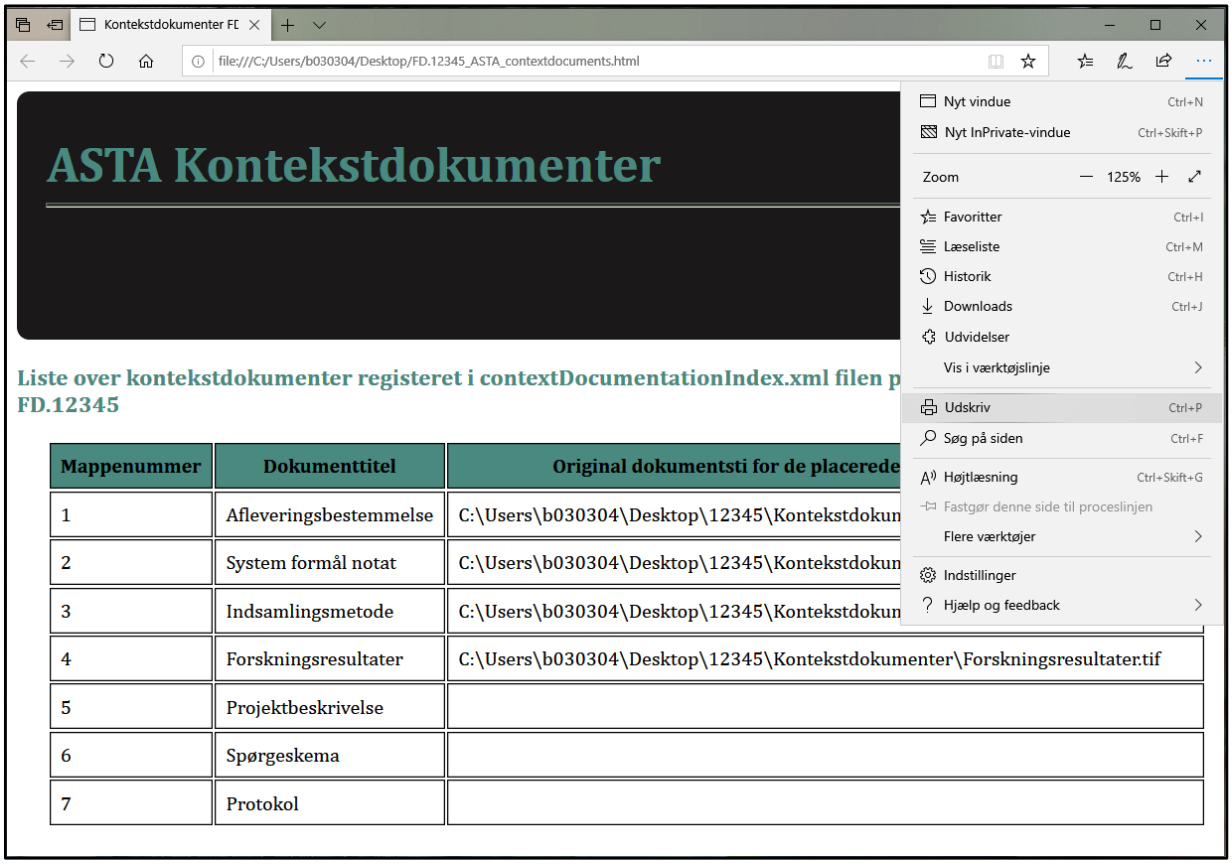


Figure 2.12: "Dokumentliste" (translation: document list), page *that appears when you click on "Udskriv dokumentliste" (*translation: Print the document list) *in ASTA*



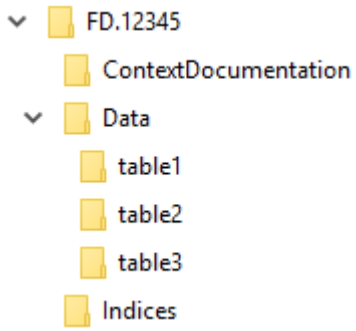
Mappenummer	Dokumenttitel	Original dokumentsti for de placerede
1	Afleveringsbestemmelse	C:\Users\b030304\Desktop\12345\Kontekstdokun
2	System formål notat	C:\Users\b030304\Desktop\12345\Kontekstdokun
3	Indsamlingsmetode	C:\Users\b030304\Desktop\12345\Kontekstdokun
4	Forskningsresultater	C:\Users\b030304\Desktop\12345\Kontekstdokumenter\Forskningsresultater.tif
5	Projektbeskrivelse	
6	Spørgeskema	
7	Protokol	

*Figure 2.13: Printing the "**Dokumentliste**" (translation: document list) via the browser's print function*

You now only need to create and place the data and metadata files in the "Data" folder.

J. Create sub-folders within the Data folder

Start by creating the sub-folder structure and naming it 'table1' under the 'Data' folder. There must be a 'table' folder for each dataset to be submitted. If you submit three spreadsheets, then 'table1', 'table2' and 'table3' folders must be created under the 'Data' folder. If you only submit one dataset, only 'table1' should be created.



K. Extract data from the spreadsheet to a data file (.csv)

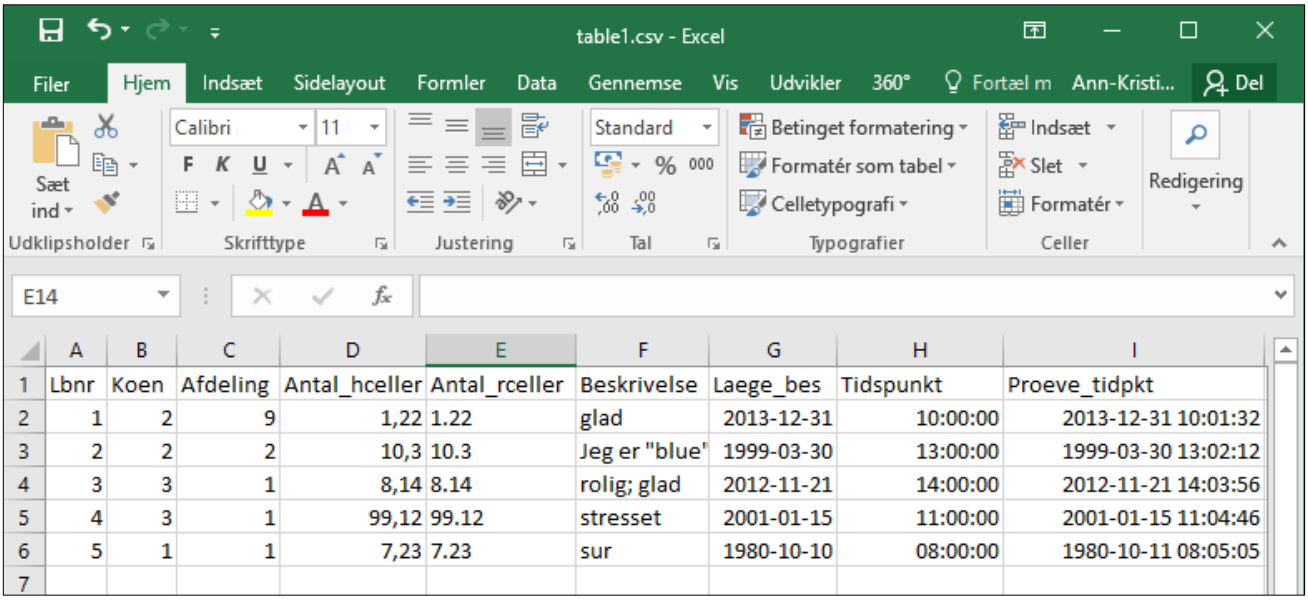
Before you extract data to a CSV file from Excel, make sure that all data formats (formats on contents in the cells) match the allowed data formats, see item 4 of the check list in section "E. Check list - steps before creating an information package" in the start of this document and illustrated in Figure 2.1.

To extract data from the spreadsheet to a CSV file, which comply with the format of CSV files in the information package, do the following:

1. Open the file in Excel. Click on 'File' > 'Open' > and select the spreadsheet you want to extract from.
2. Check that column names appear in the first row of the spreadsheet, see item 3 of the check list in section "E. Check list - steps before creating an information package" in the start of this document.
3. Check that all columns have data formats that are allowed, see item 4 of the check list in section "E. Check list - steps before creating an information package" and illustrated in Figure 2.1. Check special cells with time, dates and timestamps (both time and date)
4. Save the file as CSV by clicking 'File' > 'Save As'
5. Select the location where you want to save the CSV file e.g. directly in the information package in the 'Data' folder> 'table1'
6. Select the file type "CSV (Semicolon Separated) (*.csv)" in dropdown menu, which results from clicking on small arrow to the right of the area “Save as Type"
7. Enter file name of CSV file (e.g. "table1.csv"), if you want the extract to be placed in the 'table1' folder
8. Click 'Save'

Data from the selected spreadsheet is now extracted to a semicolon-separated file and placed in the information package.

If you have multiple spreadsheets to be submitted in the same information package, perform steps 1-8 above for each data file and place them on folders table2, table3, and so on.



	A	B	C	D	E	F	G	H	I
1	Lbnr	Koen	Afdeling	Antal_hceller	Antal_rceller	Beskrivelse	Laege_bes	Tidspunkt	Proeve_tidpkt
2	1	2	9	1,22	1.22	glad	2013-12-31	10:00:00	2013-12-31 10:01:32
3	2	2	2	10,3	10.3	Jeg er "blue"	1999-03-30	13:00:00	1999-03-30 13:02:12
4	3	3	1	8,14	8.14	rolig; glad	2012-11-21	14:00:00	2012-11-21 14:03:56
5	4	3	1	99,12	99.12	stresset	2001-01-15	11:00:00	2001-01-15 11:04:46
6	5	1	1	7,23	7.23	sur	1980-10-10	08:00:00	1980-10-11 08:05:05
7									

Figure 2.14 Example of an Excel-file with allowed data formats

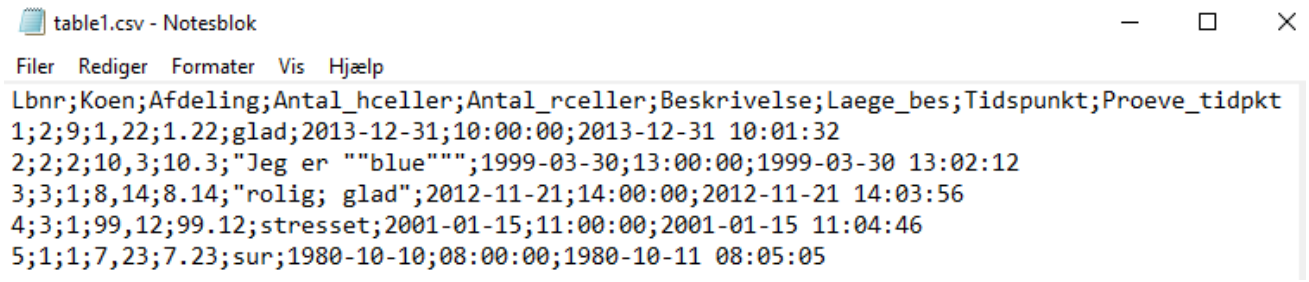


Figure 2.15 Example of the contents of a CSV exported from an Excel spreadsheet shown in figure 2.14

L. Make sure that extracting data to CSV file occurs without errors

It is your responsibility to ensure that the extract made from the original file with Excel, or other tools, is without errors. Therefore, you should perform a visual check and compare data in the original file with data in the extracted CSV file.

M. Create the metadata file (.txt)

You must now create the metadata file (table1.txt) that describes the contents of the data file (e.g. table1.csv). The content of the metadata file must comply with specific rules regarding structure and contents. These rules are defined in Figure 9.11 of Schedule 9 to the Executive Order on Information Packages. They are explained below.

Metadata file rules:

1. The metadata file must be in text format with extension .txt
2. The character set in the metadata file must be encoded as UTF-8 (see UTF-8 Guide)
3. The metadata file must be named 'table1.txt' if it is placed in the 'table1' folder in the Data folder in the information package
4. The text in the metadata file must have a left-sided orientation and no tabulations must be present
5. The metadata file must contain the following 9 headings/labels indicated in upper-case and in Danish as shown here: SYSTEMNAVN, DATAFILNAVN, DATAFILBESKRIVELSE, NØGLEVARIABEL, REFERENCE, VARIABEL, VARIABELBEKSRIVELSE, KODELISTE, BRUGERKODE
6. Rules and formats for contents under headings/labels are described in Figure 2.9

See an example of the contents of a metadata file in Figure 2.10.

You can also use the example of a sample information package available on the National Archives homepage www.sa.dk as inspiration for how to create a metadata file that meets the requirements shown in Figure 2.9.

You must name the metadata file with the same name as the data file it describes, e.g. table1.txt, and place in the same folder as that data file, see item 7 of section K.

You must create a metadata file for each data file that is included in the information package.

N. Place the data and the metadata files in the information package

Name the data and metadata files and copy them into the table folders in the Data folder.

The *Table1* folder must contain the data file named *table1.csv* and the metadata file named *table1.txt,* etc. See Figure 1.1.

Once you have named and placed all extracted data and metadata files correctly in the information package, it is finished and ready for testing.

Figure 2.9 Metadata file contents

Label	Label content description	Udfaldsrum	Fore-arrival	Obliga-torisk	Example
SYSTEMNAVN	Name of the program from which data is extracted or the original format of the data.	SPSS SAS Been Excel or fritekst	1	Yes	SYSTEMNAVN Excel
DATAFILNAVN	The name of the data file as it is called in the context of use. Note that data file name should not include the extension of the file(e.g. .xls)) Data file names must be unique within the package of submission.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99), that is: The data file name cannot be *ikke* longer than 128 characters, start with a number, or include small spaces or special characters such as %.	1	Yes	DATAFILNAVN Sundhedsundersøgelsen
DATAFILBESKRIVELSE	Description of the contents of the data file. Fx the full title of the statistics file, an existing project description, or an abstract that accurately covers the contents of the statistics file. Note that the data file description must not contain characters for line breaks.	Open text field	1	Yes	DATA FILE DESCRIPTION Health Survey of Danes living in Rostock, Germany

NØGLEVARIABEL	Data file unique key variable, provided with the name(s) of the variable(s) that compose the key variable. A key variable can be composed of multiple variables. Multiple variables in a key variable are separated by spaces. The key variable must uniquely identify a row in the data file. That is, two values in the variable cannot have the same value.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99). Dvs. at: Variable names that are part, of a key variable must *not* be longer than 128 characters, starte with a number, or include smallspaces or specialcharacters such as %.	0-1	Yes, if unique key variable exists	`KEYVARIABLE` `date of birth lbnr`
REFERENCE	References to other data files in the information package. A reference to another data file in the information package is indicated by the name of the referenced data file (foreign data file). That is followed by the variable name of the foreign data file key variable (foreign variable). Then it is followed by the variable name of the variable (reference variable) that is in the data file that refers to the foreign variable in the foreign data file. There must be full match between the data type and length of the key variables included in the reference.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99). Dvs. at: The data file name and variable names included in the reference , must *not* be longer than 128 characters, start with a number, or include gaps or special character(s) such as %.	0-m	Yes, if reference to other data file in the information package is found	`REFERENCE` `Keyfile 'nglefil_id' 'hovedfil_id'`

	If merge variables exist that can link two data sets (data files) in the information package, these must be specified. It should only specify relationships where the merging results in a dataset where variables are merged together into a dataset (horizontal merge). That is, in the case, where there are two sets of different variable but same individuals that can be merged into a dataset over a common merge variable. Relationships with any key files must also be specified here. For example, if you submit a dataset in which sensitive information from the dataset is extracted to a key file, the relationship between the dataset and the key file must be defined in a reference.				

Variable	Specifies all variable/columns in the data file. Each variable/column is specified on a separate line.
	
	A variable/column is specified as a set of variable names, followed by notation for the variable's data format and any code list reference.
		Variable naming must comply with ISO/IEC 9075:1999 - Database Language SQL (SQL-99). That is, that:
		
		Variable names included in the reference must *not* exceed 128 characters, start
			1-m	Yes	`Variable`
					`Lbnr int`
					`Koen int Koen.`
					`Int Department.`
					`Antal_hceller decimal`
					`Antal_rceller decimal`
					`Description string`
					`Laege_bes date`
					`Time hour`
					`Proeve_tidpkt datetime`

	For each variable, 2-3 pieces of information is given: ***The first piece*** of information is the variable name. Variable names must be unique within the same metadata file. Variable names must begin with a letter and must consist of a mixture of letters, numbers and underscores of a maximum of 128 characters. ***The second piece of information*** is the data type of the variable. The following xml data types are allowed: • **string** (used for text) • **int** (used for integers) • **decimal** (used for decimal) • **date** (used for dates) • **hour** (used for time/time) • **datetime** (used for timestamp - both date and time) ***The third piece of information*** is the indication of any code list reference. If the data type is numeric integer or decimal, the code list reference is indicated by the selected code list name and a subordinated period '.' If the data type is text, the reference is entered with the	with a number, or include , for example, spaces or special characters such as %. If the variable name is a reserved word in SQL:1999, it must be either renamed or encircled by double-apostrophe.			

	selected code list name with a leading dollar sign '$' and a subordinated dot '.'				
VARIABEL-BESKRIVELSE	Descriptions of the content of the variables(s) indicating the variable name followed by the description surrounded by single apostrophes. Each description should be written on a separate line. All variables/columns must have a description that is written as fully as possible so that a future user has the opportunity to understand which information is recorded in the column. All abbreviations in the description must be explained and any units of measurement must be specified (e.g. meters, centimeters, kilograms). If there are dates, there must be an explanation for what the date is about e.g. wedding date or date of birth. If information about several individuals in a dataset is available, e.g. data on a mother and child, the variable description should clearly indicate which variable information is linked to which individuals, for example, the	Open text field	1-m	Yes	VARIABELBESKRIVELSE Lbnr 'Serial Number' The cow 'The Patient's Gender' Department 'Undersøgelsesafdeling' Antal_hceller 'Number of White Cells' Antal_rceller 'Number of red cells' Description 'Words that best describe your mood' Laege_bes 'Date of medical visit (CCYYMMDD)' Time 'Time doctorexamination'' Proeve_tidpkt'Tidspunkt blood test'

	gender of the child or the gender of the respondent. For surveys, the Danish National Archives recommends that all the question texts be inserted into the variable/column descriptions. This will make it easier for future users to search and use the data. The question text should be inserted in its original formulation and full length.				
KODELISTE	Code lists, also called value labels, indicated by the code list name, followed by several sets of codes and code descriptions. Each information must be encircled by single apostrophe and separated by spaces. If you use codes in the data file that are not also used in other variables/columns, you can enter them here. Furthermore, remember to enter a code list reference note for all the variables/columns where the code is used named under the VARIABLE label (described above in this table). In surveys, code lists (known also as value labels) correspond to answer categories to a question.	Naming code lists must comply with ISO/IEC 9075:1999 - Database Language SQL (SQL-99). That is, that: Code list names must *not* be longer than 128 characters, starte with a number, or include smallspaces or specialcharacters such as %.	0-m	Yes, if code lists exist	KODELISTE Koen '1' 'Man' '2' 'Female' Section '1' 'Celile Blodbank' '2' 'Marie Vang Blood Bank' '9' 'Undisclosed'

BRUGERKODE	Codes for missing values indicated by the name of the column in which the missing values is used, followed by the missing code encircled by single apostrophes and separated by spaces. Examples of missing codes for missing values: 9 = Undisclosed 10 = Irrelevant 11 = Do not participate Missing codes must also appear in the code list under the KODELISTE label associated with the column in which they are used, also with a code list reference under the VARIABLE label. Missing codes for missing values can only be used in categorical and numerical variables with the integer, decimal and text data types.	Variable naming must comply with ISO/IEC 9075:1999 - Database Language SQL (SQL-99). That is, variable names *cannot* be longer than 128 characters, start with a number, or include small spaces or special characters such as %.	0-m	Yes, if special codes exist for missing values	`BRUGERKODE` `Afdeling '9'`

table1.txt - Notesblok

Filer Rediger Formater Vis Hjælp

```
SYSTEMNAVN
Excel

DATAFILNAVN
Sundhedsundersøgelsen

DATAFILBESKRIVELSE
Sundhedsundersøgelse af danskere som bor i Rostock, Tyskland

NØGLEVARIABEL
Lbnr

REFERENCE

VARIABEL
Lbnr int
Koen int Koen.
Afdeling int Afdeling.
Antal_hceller decimal
Antal_rceller decimal
Beskrivelse string
Laege_bes date
Tidspunkt time
Proeve_tidpkt datetime

VARIABELBESKRIVELSE
Lbnr 'Løbenummer'
Koen 'Patientens køn'
Afdeling 'Undersøgelsesafdeling'
Antal_hceller 'Antal af hvide celler i 1mm3 blod'
Antal_rceller 'Antal af røde celler i 1mm3 blod'
Beskrivelse 'Ord som bedst beskriver dit humør'
Laege_bes 'Dato for lægebesøg (CCYYMMDD)'
Tidspunkt 'Tidspunkt for lægeundersøgelse (hh:mm:ss)'
Proeve_tidpkt 'Dato og tidspunkt for blodprøvetagning (CCYYMMDD hh:mm:ss)'

KODELISTE
Koen
'1' 'Mand'
'2' 'Kvinde'
Afdeling
'1' 'Celile Blodbank'
'2' 'Marie Vang Blodbank'
'9' 'Uoplyst'

BRUGERKODE
Afdeling '9'
```

Figure 2.10 Example of a folder structure and the contents of a metadata filel

3. Submission of CSV file as information package

If you have CSV files extracted from programs other than Excel, you can also make an information package with these as follows:

1. Create the folder structure of the information package and place index files and context documents as described in section G-I.
2. Create the sub-folders in the Data folder as described in section H.
3. You must ensure that the CSV file complies with the requirements for data files in regards to the provision on Schedule 9 session 9.F-9. H in the Executive Order on Information Packages. With the program ASTA, you can test whether your CSV file complies with data file provisions. Data file provisions are that:
    a. The CSV file must be semicolon separated
    b. First line of the data file must specify column names/variable names
    c. The CSV file must comply with escape and semicolon rules in text values. That is, if the semicolon is included in a value for a variable, the whole value must be enclosed with double apostrophe '"' (U+0022). If double apostrophe is included in a variable value, the double apostrophe must be preceded by a double apostrophe and the entire value shall be enclosed with double apostrophe.
    d. the data formats in the CSV file must comply with the formats of data types described in Figures 2.1-2.7
    e. missing values in the data file must be indicated as no value »« or a space » «
4. Create the metadata files manually as described in section M

4. How to test the information package

It is your responsibility that the structure and contents of the information package comply with the information package requirements set out in Schedule 9 of the Executive Oder on Information Packages. Therefore, it is important that you test the submission package and correct any errors, before you submit the information package to the archive. The program ASTA can help you test the information package.

1. Download the ASTA from the National Archives homepage www.sa.dk
2. Click AstaInstaller.exe to start executing the program ASTA
3. Select the menu item 'Test information package', see Figure 4. 1
4. Select the information package you want to test (e.g. FD.12345) by clicking the '**Browse**' button in Figure 4.1
5. Click on **'Start Test'** button
6. The test has now begun and the test status is displayed in the test **log**, see also Figure 4.2

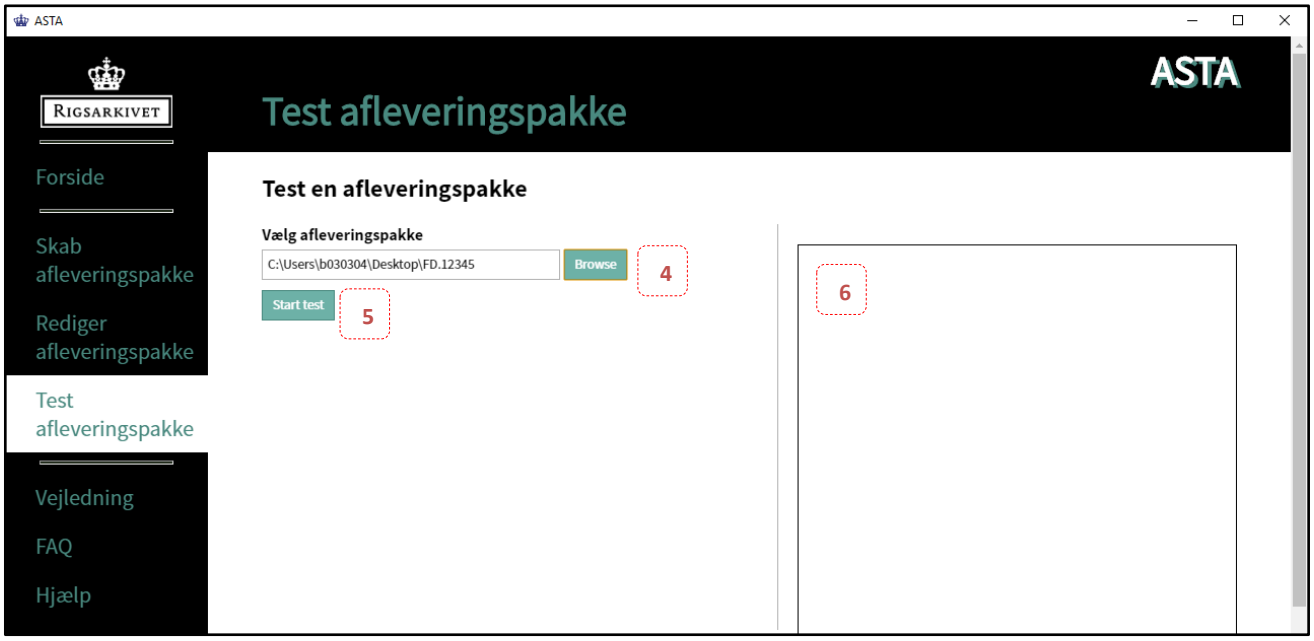


Figure 4.1: 'Test afleveringspakke' (translation: Test information package) page in ASTA before starting the test.

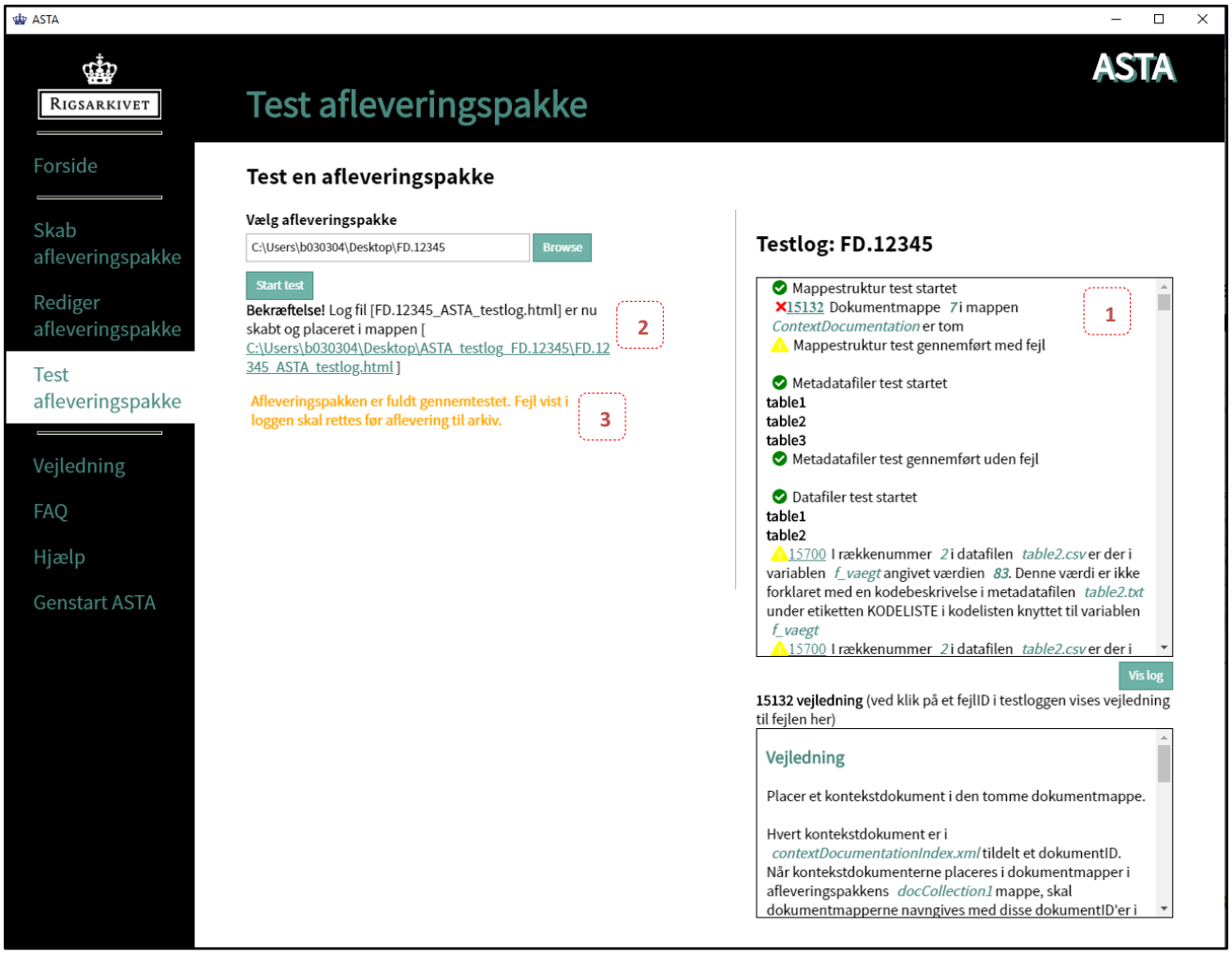


Figure 4.2: 'Test afleveringspakke' page in ASTA when the test is completed

Test status and correction of errors in test log

1. Carefully review the generated **error messages** marked with red crosses ✖ in the test log and always correct them before submitting them to the archive. **Hints** appear as warnings in a yellow exclamation point ⚠ . Hints are not always errors and should only be corrected if they are judged to be errors in the information package.

2. In addition to the test log that appears on the screen, a **html version of the test log** is also generated automatically after the test in the same location as the selected information package (see Figure 4.3). By clicking on the link to this test log, it opens in your default browser and can be printed using the browse print function (see figure 2.13). The HTML test log also provides instructions on how to understand and correct error messages in the test log (see Figure 4.4).

3. When the test is finished, one of the following three **status messages** appears:

   The information package has been tested without errors and can now be delivered to the archive

   This indicates that an information package is ready to be submitted to the National Archives.
   NOTE: Since ASTA does not test for all the provisions on Schedule 9, the Archives may contact you with further errors or missing elements that need to be corrected.

   The information package has been fully tested. Errors shown in the log must be corrected before submission to the Archives.

   This indicates that the test has been completed, but there are still errors that need to be corrected. After you correct the errors shown in the test log and run a new test, no more new errors should appear.

   The test has been aborted and the information package has not been tested. Errors shown in the log must be corrected, before the test can continue.

   This indicates that there are errors that need to be corrected before the test can continue. After you correct the errors shown in the test log and run a new test, several new errors may occur. For multiple tests, a maximum of 40 of the same type of error appears in the test log. This means that after you correct the 40 errors shown and run a new test, more of the same type may occur.

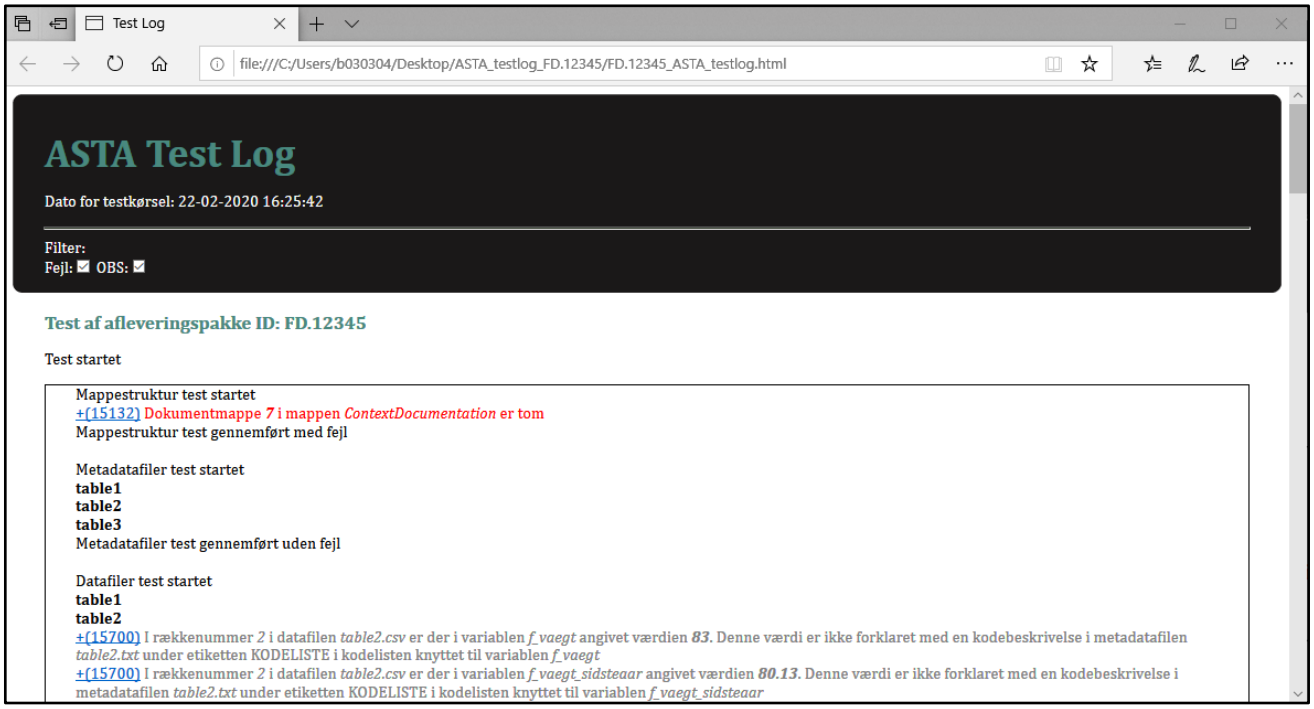


Figure 4.3: Html-version of the test log from 'Test afleveringspakke' in ASTA

Html version of test log with guide

1. The html test log that appears when you click the link from ASTA to this test log (see 2 in Figure 4.2) displays an error ID (e.g. 15132) based on each error found. Clicking on this error ID will provide a guide for how to understand and correct the error. An example of data and metadata without errors is also shown, as well as an indication of the specific wording of the provisions, as it is written on Schedule 9 of the Executive Order on Information Packages (see Figure 4.4)

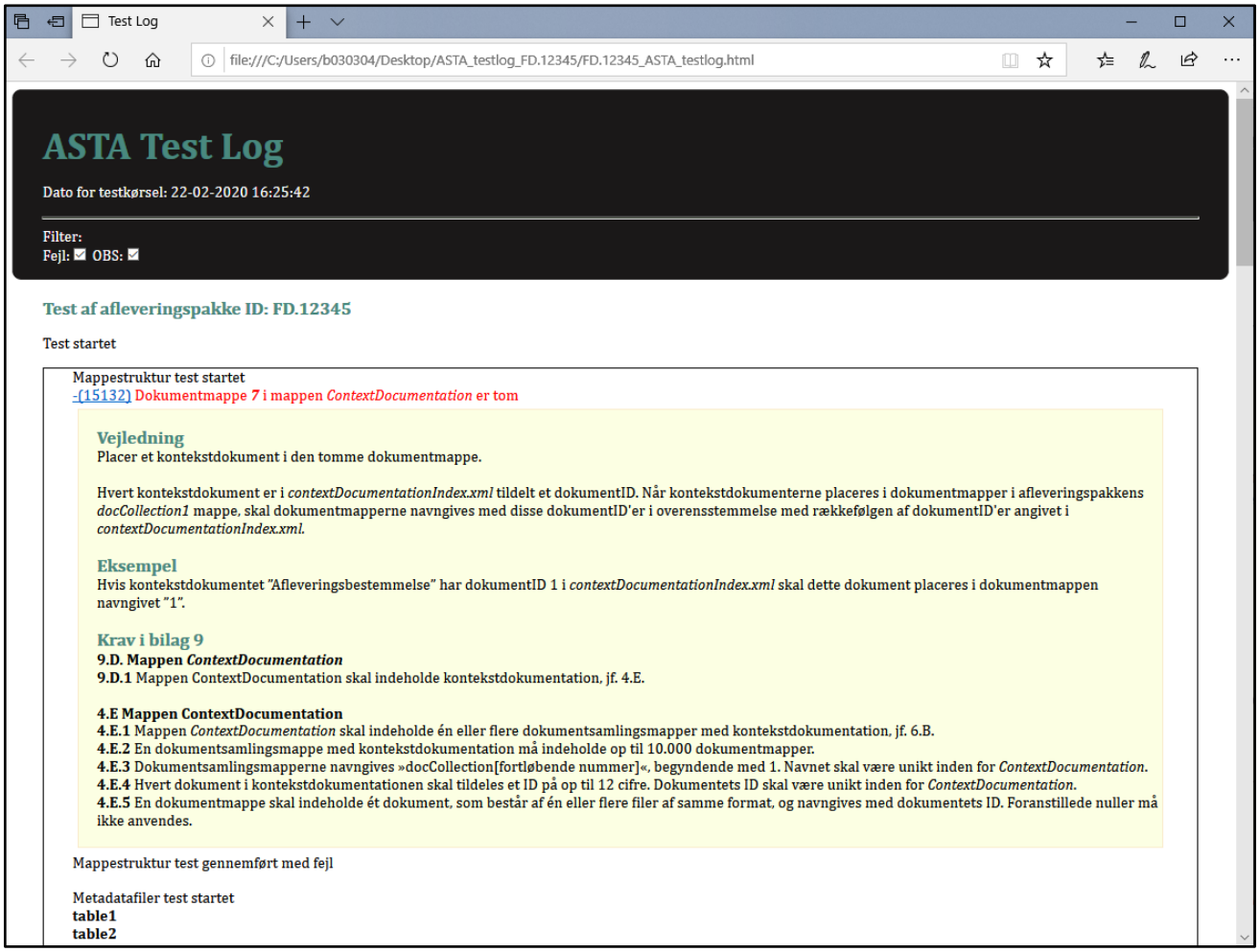


Figure 4.4: Html-test log with the guide after one click on errorl-ID 15132

5. Support in the National Archives

If you experience problems creating and testing the information package, please contact the data manager for research data in the National Archives at the following e-mail: mailbox@sa.dk.