User Guide to ASTA version 2.1.1

User Guide to the program ASTA (Submission of Statistical data to the Archives) that can be used in regards to the production and testing of an information package including research data. The submission to the Danish National Archives should conform to the requirements listed under the schedule 9 in the Executive Order on Information Packages (Executive Order no.128, of 12 February, 2020)

The Danish National Archives, June 2020

Version 2.0

Contents

0. Reading guide to ASTA User Guide 2

A. User guide's target audience and its application 2

B. Reference to other guides 2

C. Legislation and regulation 2

D. Definitions 3

1. ASTA – program to submit statistical files to the Danish National Archives 3

A. What is an information package? 3

B. Where do I find ASTA? 3

D. Checklist: Important steps before using ASTA 4

E. Review of ASTA user interface 6

F. Generate the information package 7

G. Testing the information package 15

H. Control that ASTA extraction has no losses 18

I. Editing the information package 19

J. ASTA support in the National Archives 21

Appendix 1 22

A. SPSS – preparing the statistical file for extraction 22

B. Stata – preparing the statistical file for extraction 26

C. SAS – preparing the statistical file for extraction 30

D. R/RStudio – forberedelse af statistikfilen før udtræk ..... 35~~34~~

0. Reading guide to ASTA User Guide

Public authorities, including research institutions, are required to provide a copy of the research data and documents deemed worthy of preservation. If the research data to be submitted are statistical files or similar (for example, spreadsheets), the submission format is called a submission information package. State authorities must submit them to the Danish National Archives (Rigsarkivet). Municipalities and regions can decide whether to submit to the Archives or to create their own archive.

The National Archives has established a series of provisions for an information package due to the preservation and future use of the data. All authorities must comply with those provisions when delivering data. These provisions are described in the National Archives' Executive Order on Information Packages, Schedule 9: Information package of certain types of research data.

The National Archives developed a tool called ASTA, which can be used for the production and testing of an information package. ASTA can be downloaded from the National Archives' homepage www.sa.dk.

You can extract data and metadata files automatically from statistical programs such as SAS, Stata, SPSS and R/RStudio with ASTA. Appendix 1 provides a guide on how to ensure your statistical files from the above mentioned statistical programs comply with the requirements listed in Schedule 9, before extracting them with ASTA.

The ASTA user guide describes how to use the program, as well as what you need to know before beginning to generate an information package with ASTA.

A. User guide's target audience and its application

The ASTA user guide is directed to those that produce information packages to be delivered to the Archives with data from statistical files with SAS, Stata,SPSS and R/RStudio formats. Those can be, for instance, the authority's own data manager, the IT department, a researcher or an external IT- consultant.

B. Reference to other guides

In addition to the ASTA User Guide, the National Archives has prepared other relevant guides for the production and submission of information packages:

- Quickguide – for production and testing of an information package with ASTA
- User Guide to Schedule 9 about information packages of certain types of research data, in the Executive Order on Information Packages
- User Guide to the production of an information package with data from spreadsheet or csv-files
- User Guide to the program Skab archiveIndex
- User Guide to the program Skab contextDocumentationIndex
- User Guide to converting a document to TIFF
- User Guide to UTF-8
- Sample information package with statistical data FD.18005

All guide materials, along with their programs, can be found in the National Archives' homepage www.sa.dk.

C. Legislation and regulation

Information about legislations can be found on the Danish National Archives' (Rigsarkivet) homepage www.sa.dk.

D. Definitions

Information package with data from statistical files in general consists of context documents - which should be delivered in the archival format designated by the Archives -, extracts of data and metadata from statistical files, as well as two index files in xml format containing metadata about the data and the context documents submitted.

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

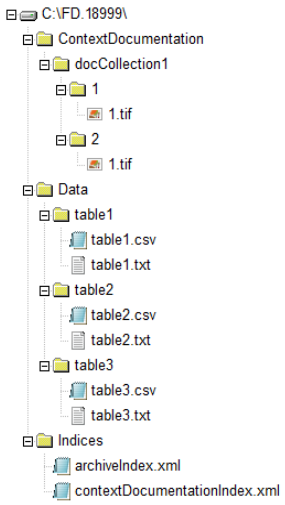
1. ASTA – program to submit statistical files to the Danish National Archives

A. What is an information package?

As part of the Archives' requirements for the submission of statistical data, data from original statistical files must be extracted into an information package and tested prior to submission to the Archives. ASTA is a tool designed to help in creating and testing the information package.

The information package must follow some specific requirements in regards to folder structure, folder names, files, data types, variable names, variable labels, format names, missing values, etc. All requirements for an information package can be found in the Schedule 9 in the Executive Order on Information Packages.

As an example, an information package can look like this:



B. Where do I find ASTA?

Versions of ASTA for Windows and Mac can be downloaded from the Archives' homepage.

C. System technical requirements, when you use ASTA

You can find below the operational systems, internet browsers and versions of statistical programs that ASTA supports and which versions it has been tested for.

If you use other operational systems, internet browsers or statistical programs and versions, we cannot guarantee that ASTA will be error-free.

Operational systems:
ASTA is tested for the following operational systems:

- Windows 7 and 10
- macOS High Sierra

Internet browsers:
The test logs in ASTA, showing information and guide under an error-ID, work on the following browsers:

- Google Chrome, Version 77.0.3865.90 (Official version) (64-bit)
- Mozilla Firefox, version 69.0.2 (64-bit)
- Microsoft Internet Explorer 11 [Remember to accept blocked contents to see error-IDs, if necessary]
- Microsoft Edge, version 17.17134
- Safari version 11.1.2 (13605.3.8)

Statistical programs:
Export scripts to extract data and metadata from statistical files can be executed in the following versions of the statistical programs:

- SAS v. 9.4
- SPSS v. 24
- Stata v. 14.2
- R v. 3.6.2 /Rstudio v 1.20.5033

Additional system technical requirements:
Optimal screen definition is 1920 x 1080 or 1920 x 1200. Lower definition use may prevent some buttons from appearing in ASTA.

As ASTA will be installed on your computer, you need to have administrator rights over your computer to run the program.

D. Checklist: Important steps before using ASTA

The following steps should be checked or performed prior to creating an information package with ASTA. Appendix 1 in the end of this document gives a thorough description of how you can do it using different statistical programs SPSS, Stata and SAS, to ensure that items 1-8 of this checklist are fulfilled.

NR.	TASK	DONE
GENERAL		
1. Check/install statistical program	Check that you have installed the statistical program that you use to open the file and you will be extracting data from with ASTA	
2. New folder with a copy of the material	Create a new folder in your computer with a copy of all data and documents that will be submitted to the Archives. In other words, the statistical files, context	

	documents cited in the archival provisions and the two index files that will be submitted. If possible, name the folder with the information package number that is stated in the archival provisions (afleveringsbestemmelse).	
STATISTICAL DATAFILES		
3. Data files	Data files should be either in SPSS, SAS or Stata format. Data files have to be encoded as UTF-8.	
4. Variables in the data files	Variable names should not: be more than 128 characters; start with a number; or include a blank spaces or special characters, such as %. All variables in the file should have variable labels that describe their contents.	
5. Variable type/format	All variables should have defined type/format that are accepted by the Archives, cf. Schedule 9, figure 9.3 in the Executive Order on Information Packages. They cannot be default formats.	
6. Value labels	All values in value labels must be unique and specified. In case some values are not specified (for example in the case of Likert scales, where only the first and last values have specific labels), the variable description (variable label) should describe this. It can e.g include the text ”Scale used. Not all values are described in the value label”	
7. Missing values	Missing values can only, be used in numeric and categorical variables and they should be consistent. Contact the Archives’ data manager for research data, if you use codes for missing values in character (text/string), time or date variables. All codes for missing values should, also be described in that specific variable’s value labels.	
8. References (in deliveries of more than one dataset)	When you deliver more than one dataset in the same information package, where there is a reference between these two files (a merging variable), the merging variable should have the same type/format and length across these datasets.	
CONTEXT DOCUMENTATION		
9. 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’.	
10. 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		
11. 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'.	

E. Review of ASTA user interface

Elements in the main menu (Box 1 shown in orange color bar in figure 1)

ASTA's main menu is placed on the left side under the National Archives logo. The first part of the menu is a front page introducing ASTA and showing how an information package structure and contents look like.

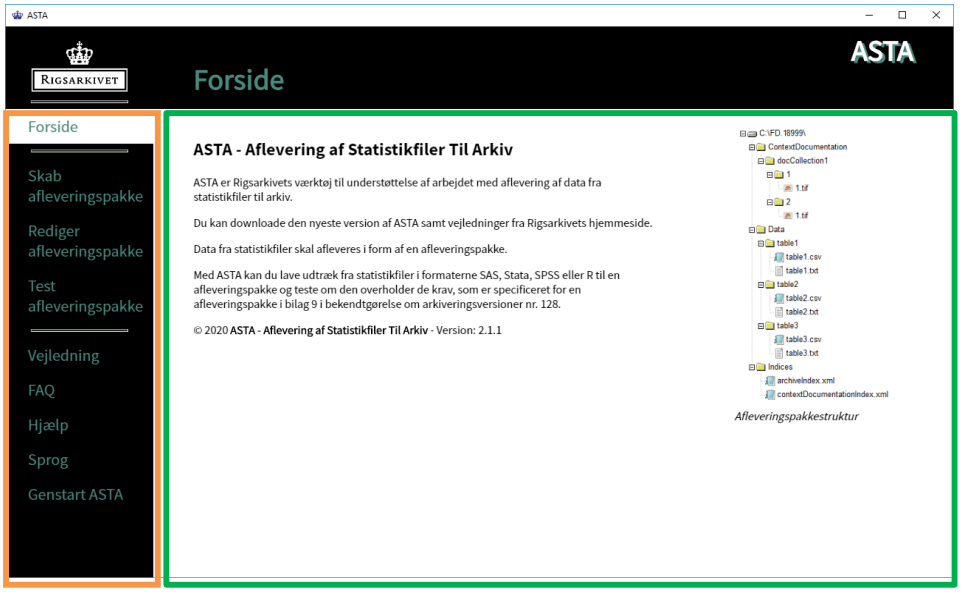


Figure 1: ASTA's front page

Box 1 — Main menu box with 7 different work levels.
Box 2 — Detailed sub-levels of the chosen main menu level.

After the front page, there are three main functions in the program:

- **Generate the information package** – By clicking on this menu item you can immediately start the process of creating an information package. The process is composed of 5 steps:
    1. Create the information package folder structure
    2. Extract data and metadata from statistical file(s)
    3. Fill in information about the statistical file(s)
    4. Include index files
    5. Include context documents

- **Edit the information package** – Under this menu item, you can upload new or missing index file(s) or context document(s) to an existing information package.
- **Test the information package** – This menu item can be used to test the information package, after the information package has been generated. This is an important step prior to submitting the information package to the archives, because this function validates that the information package structure and contents conform with the provisions specified in "Schedule 9 about information packages of certain types of research data" in the Executive Order on Information Packages. When the Archives receives an information package, it uses the same tool to test the submission and, if the provisions are not followed, the information package creator is requested to correct and re-submit the information package. Hence, we recommend performing the test and adjusting for eventual errors or missing element(s), shown in the test log. This has to be done until the test runs with no errors.

The third part of the main menu includes support information about ASTA and other relevant information regarding the submission of data from statistical file(s) to the Archives.

F. Generate the information package

Preparation prior to using the function Generate the information package

See the checklist in item D. Among other things, you should have the following ready prior to beginning to generate the information package with ASTA:

- Information package number, as stated in the Archival provisions (afleveringsbestemmelse).
- A new folder with a copy of all the material that will be included in the information package, including a copy of the statistical file(s) to be submitted to the Archives.
- You need to make sure that the statistical file(s) follow the provisions in Schedule 9, see 3-8 in paragraph 1, item D.
- Approved versions of index files archiveIndex.xml and contextDocumentationIndex.xml.
- All context documents registered in the contextDocumentationIndex.xml file must be converted to one of the archival format: TIFF, JPEG-2000, MP3, WAVE or MPEG.
- A description of the dataset.
- Information on key variables in the file (if there are any).
- If there are several data files that can be merged together in the information package, you should describe in a separate document which datasets can be merged together and which variables are the merging variables (keys).

Folder structure

When you click on the menu item '**Create information package**' a view page appears, as shown in figure 2. In that part of the program, you can create a folder structure for the submission information package.

1. Type the **serial number** for the information package, in accordance with the information that is provided by the Archives in the archival provisions. If the information package name provided is, for instance, FD.12345, then the number to type in that field is 12345.
2. Choose the destination, where the information package shall save files, by clicking on the button '**Browse**' and choosing the correct folder.
3. Click on '**Create folder structure**' to create a folder structure for the information package. Note that in the next page, at the bottom, there is a link - that can be clicked on – linking to the folder structure that was created (see figure 3).

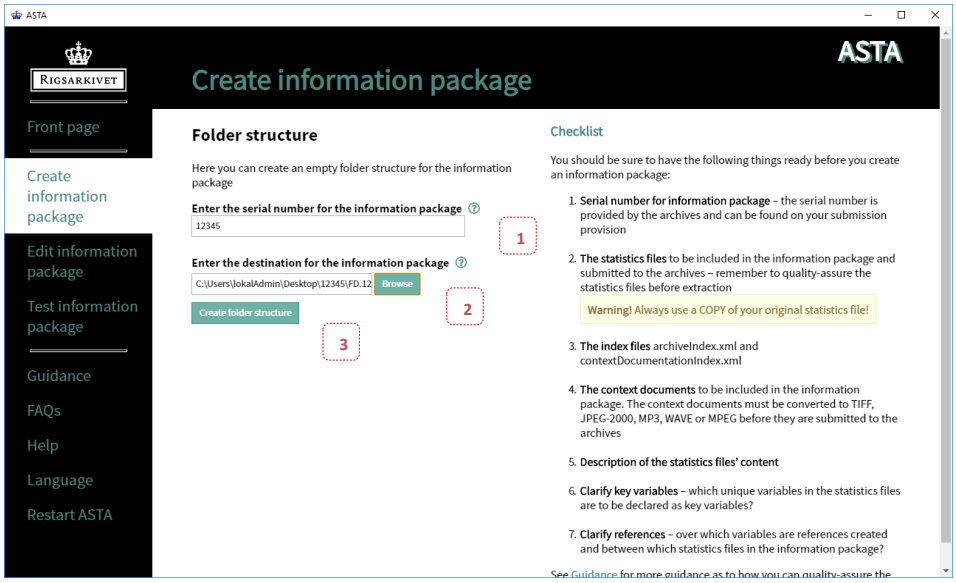


Figure 2: Page "Folder structure" under "Create information package" in ASTA

Extraction of data from statistical file(s)

When you click on the button 'Create folder structure' (see the previous paragraph) the next page opens with a function to extract data and metadata from statistical files shown in figure 3.

Extraction of data and metadata from statistical file is composed of three steps.

In STEP 1, in figure 3, you choose the statistical file where data will be extracted from:

1. Choose the copy of your statistical file you will be extracting from with the button **'Browse**'
2. Click on '**Next**' to move to STEP 2.

STEP 2: Run the export script

Data and metadata is extracted from a statistical file to a data file (e.g. table1.csv) and to a metadata file (e.g. table1.txt), that conforms with the provisions about a structure from Schedule 9 by running an export syntax from the script that ASTA automatically generated based on the statistical file. In step 2, in figure 4, you should run that export script. In this manner:

1. **Do not close the program ASTA** while you run the export script, which you find when you click on '**here**' in the green box with the text "**Export script has been created and is located her**". By clicking on 'here' an explorer window opens up with the place where the export script is placed. The export script is named after the statistical file name, with a script extension that is either SPSS (.sps), Stata (.do), SAS (.sas) or R/RStudio (.rds, .RDS, .RData. Note that at the bottom of the program screen the exact name of the script is presented (for instance, generationsundersøgelsen.sps). Double click on

the script file to open it in its respective statistical program, which you should have installed in your computer. Run the syntax in that export script.

2. When the script is finished running, you should return to ASTA and click on the button "**Next**" to go to STEP 3.

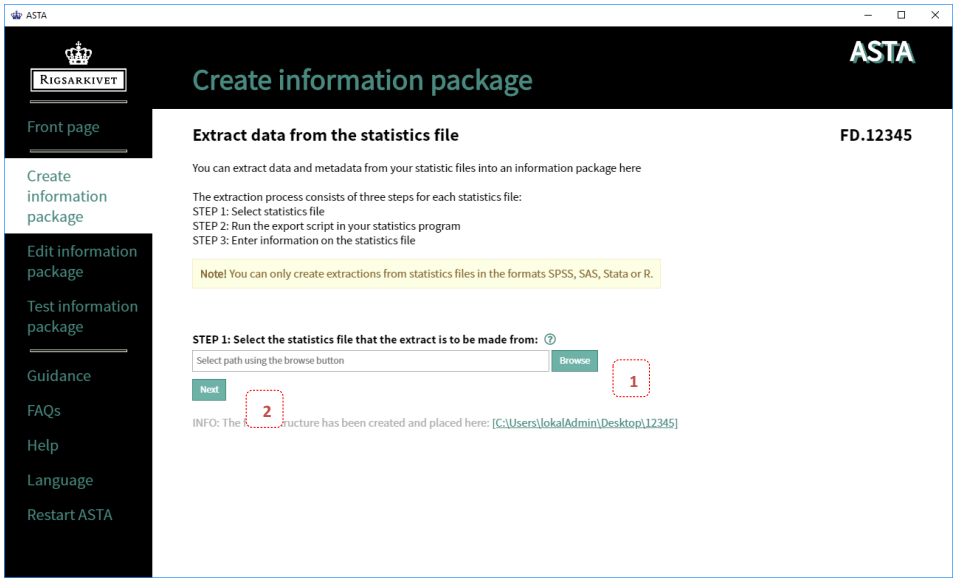


Figure 3: Page "Extract data from the statistics file" under "Create information package" in ASTA

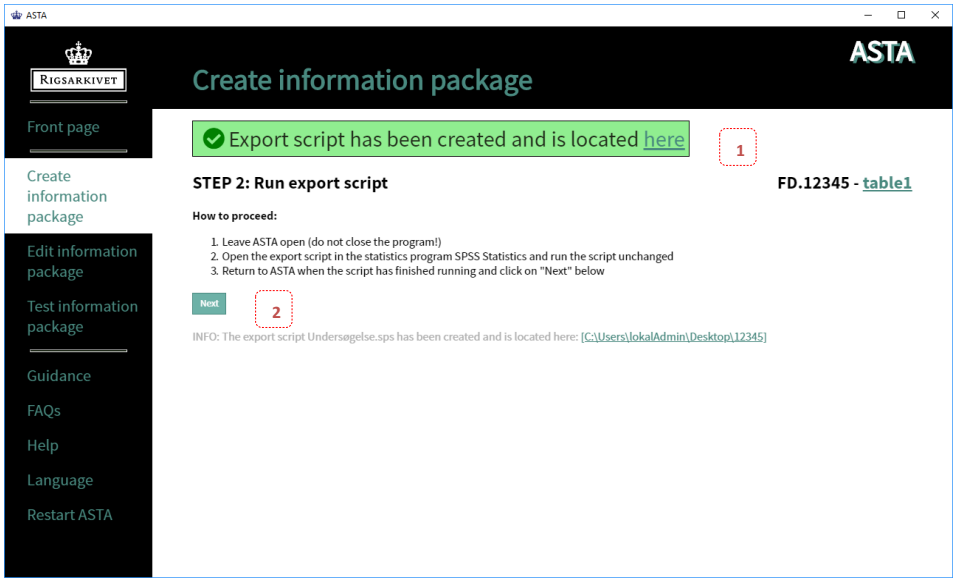


*Figure 4: Page "*Step 2: Run the export script*" under "Create information package" in ASTA*

STEP 3: Enter the information about the statistical file

In step 3, in figure 5, you should enter the information about the chosen statistical file.

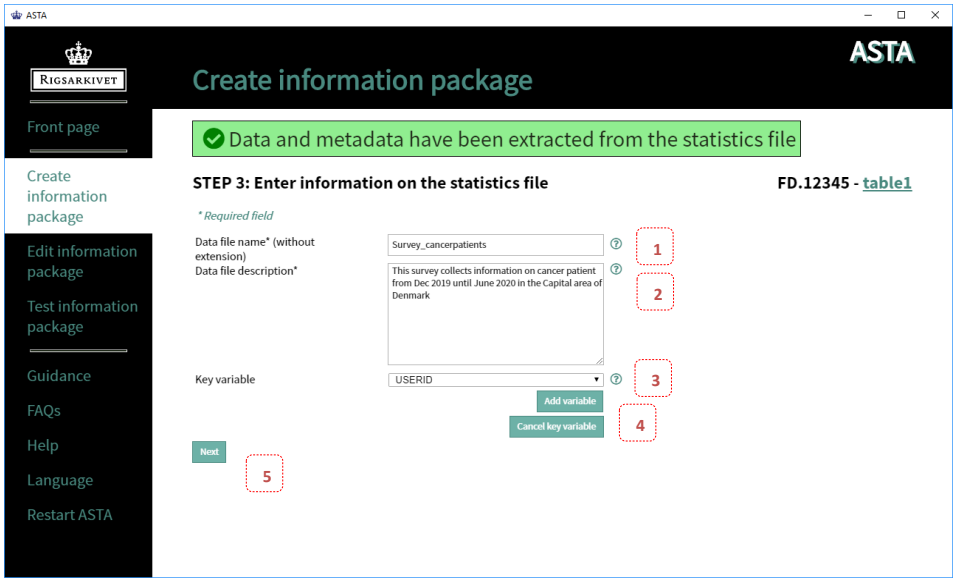


Figure 5: Page "STEP 3: Enter information about the statistics file" under "Create information package" in ASTA

1. Provide the dataset name, as it is referred to when it is used (mandatory)
2. Provide a description of the dataset's contents, max. 4096 characters (mandatory)
   If you have a unique key variable in the file, you should follow step 3 and 4. Otherwise, go to step 5.
3. Click on the small arrow and choose the key variable from the dropdown menu that appears
4. Click on '**Add variable**' to include the chosen key variable
   If your key variable is composed of several variables, go back to step 3. and 4, until all variables are chosen.
   *The chosen key variable can be deleted by clicking in the button '***Cancel key variable***'.*
5. Click on '**Next**' to go to the next page

Data extraction is finished

By clicking on **"Next"** in the previous STEP 3, the next page is shown as seen in figure 6. It shows a confirmation that, the data extraction is finished for the chosen file. Note that under INFO, a link is shown to the generated data and metadata, e.g. table1.csv and table1.txt.

You can choose to make an extraction with other statistic file(s) that will be delivered in the same information package.

1. Click on the button "**New extraction**", if you wish to make an extraction from other statistical file(s).

This step will send one back to STEP 1 on the page ”Enter information on statistics file”, where you will choose the statistical file that you want to extract data from.

2. Click on the button ”**Next**”, if the information package will not include other statistical file(s).

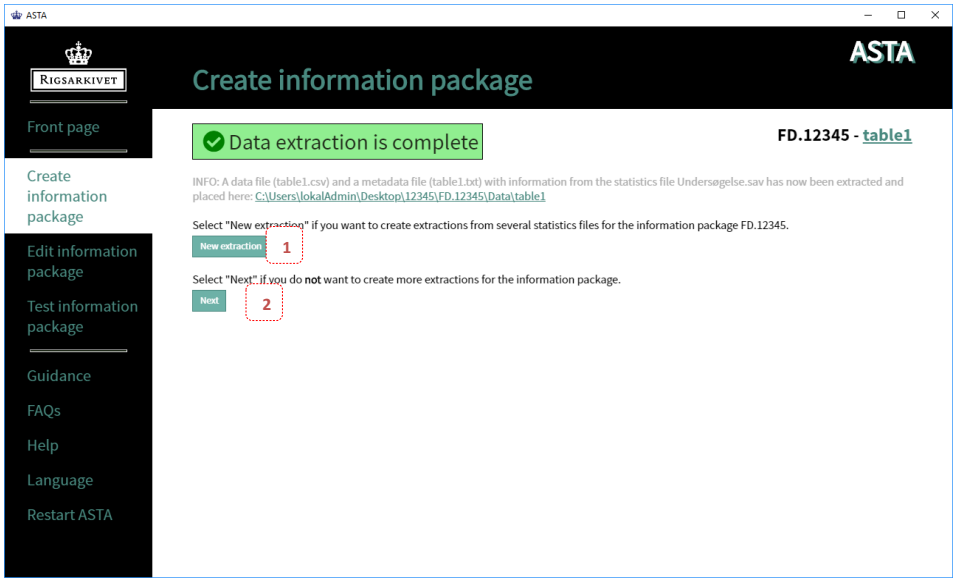


Figure 6: Page ”Data extraction is complete” under ”Create information package” in ASTA

Copying the files for controlling the extraction

You can make a copy of the files used in order to control the extraction with ASTA by clicking on the button **‘Next’**. The files will be copied into the folder ’ASTA_kontrolfiler_FD.XXXX’. That folder should be submitted to the archives along with the submission information package.

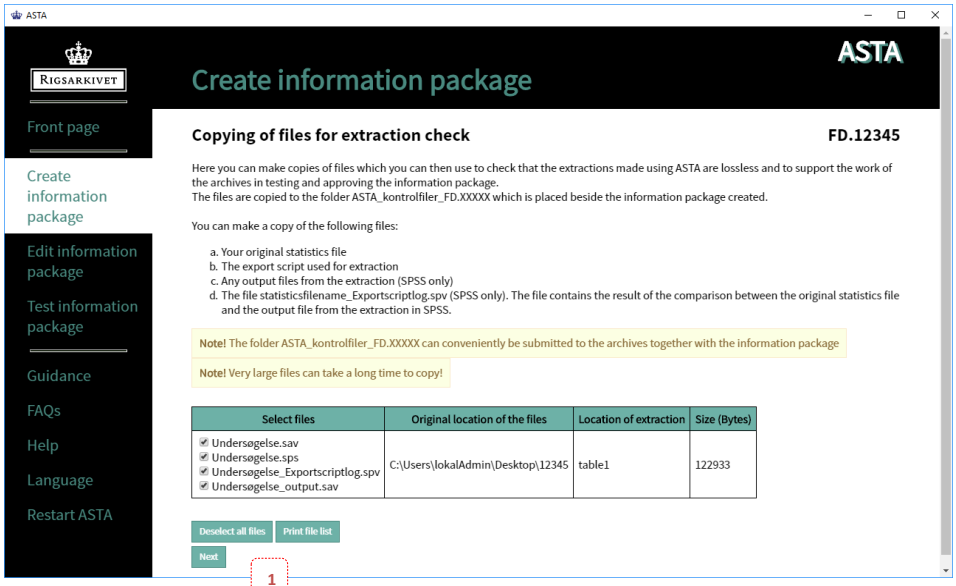


Figure 7: Page "Copying of files for extraction check" under "Create information package" in ASTA

Placing index files in the information package

You must place the index files in the information package, see figure 7.

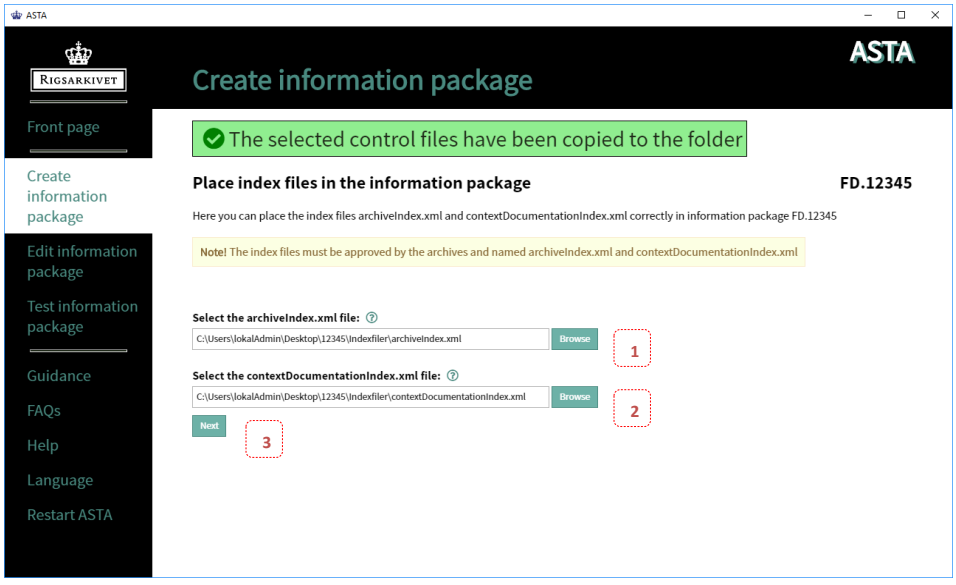


Figure 8: Page "Place index files in the information package" under "Create information package" in ASTA

1. Choose the file *archiveIndex.xml* by clicking on the '**Browse**' button.
2. Choose the file *contextDocumentationIndex.xml* by clicking on the '**Browse**' button.
3. Click on the button '**Next'** to upload the chosen files.

Placing context documents in the information package

You must now place the context documents that are registered in the file contextDocumentationIndex.xml in the information package, see figure 9a.

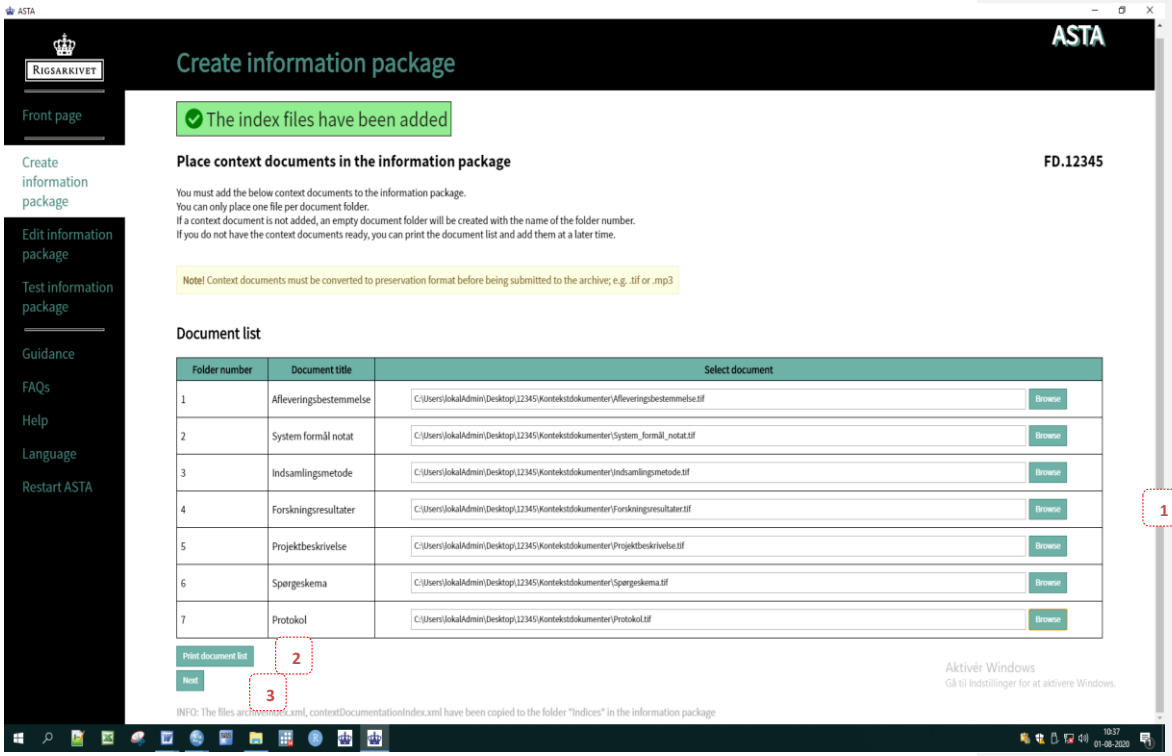
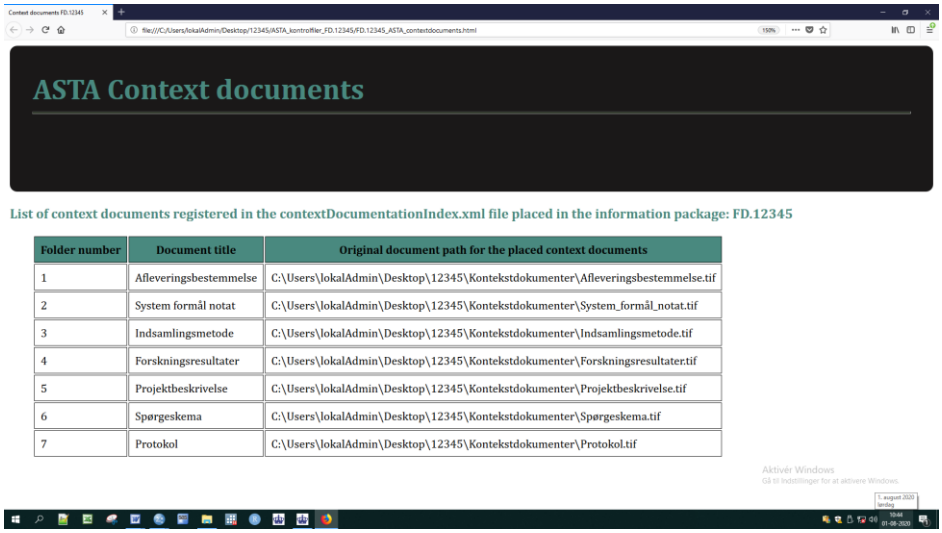


Figure 9a: Page "Place context documents in the information package" under "Create information package" in ASTA

1. All context documents registered in the contextDocumentationIndex.xml file are in the 'document list' . Click on the button '**Browse**' for each of the document and choose the context document that refers to the document title in the document list. If you do not have all the context documents ready, you will be able to include them later under the menu item '**Edit information package**'.
2. We recommend that you print the document list by using the button '**Print document list**' and control that the content of the documents corresponds with their title. When you click on the button "Print document list" you will see the document list in your standard browser, e.g. Edge (see figure 9b). Use the browser's print function to print the list (see figure 9c).
3. Click on '**Next**' to upload the chosen document(s).

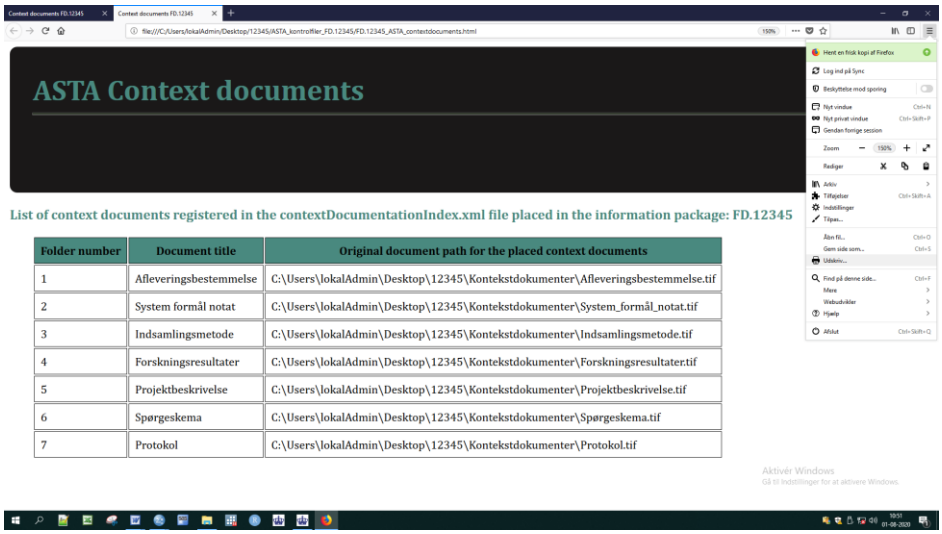


ASTA Context documents

List of context documents registered in the contextDocumentationIndex.xml file placed in the information package: FD.12345

Folder number	Document title	Original document path for the placed context documents
1	Afleveringsbestemmelse	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Afleveringsbestemmelse.tif
2	System formål notat	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\System_formål_notat.tif
3	Indsamlingsmetode	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Indsamlingsmetode.tif
4	Forskningsresultater	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Forskningsresultater.tif
5	Projektbeskrivelse	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Projektbeskrivelse.tif
6	Spørgeskema	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Spørgeskema.tif
7	Protokol	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Protokol.tif

Figure 9b: "Document list" that appears when you click on "Print document list" in ASTA



ASTA Context documents

List of context documents registered in the contextDocumentationIndex.xml file placed in the information package: FD.12345

Folder number	Document title	Original document path for the placed context documents
1	Afleveringsbestemmelse	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Afleveringsbestemmelse.tif
2	System formål notat	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\System_formål_notat.tif
3	Indsamlingsmetode	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Indsamlingsmetode.tif
4	Forskningsresultater	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Forskningsresultater.tif
5	Projektbeskrivelse	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Projektbeskrivelse.tif
6	Spørgeskema	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Spørgeskema.tif
7	Protokol	C:\Users\lokalAdmin\Desktop\12345\Kontekstdokumenter\Protokol.tif

Figure 9c: Print the "Document list" by using the browser's print function

The Information package is created

The information package is now created and must be tested.

1. Click on the button '**Test FD.XXXXX**' that is a shortcut to the test of this created information package. You can also test the information package by choosing the menu item 'Test information package'.

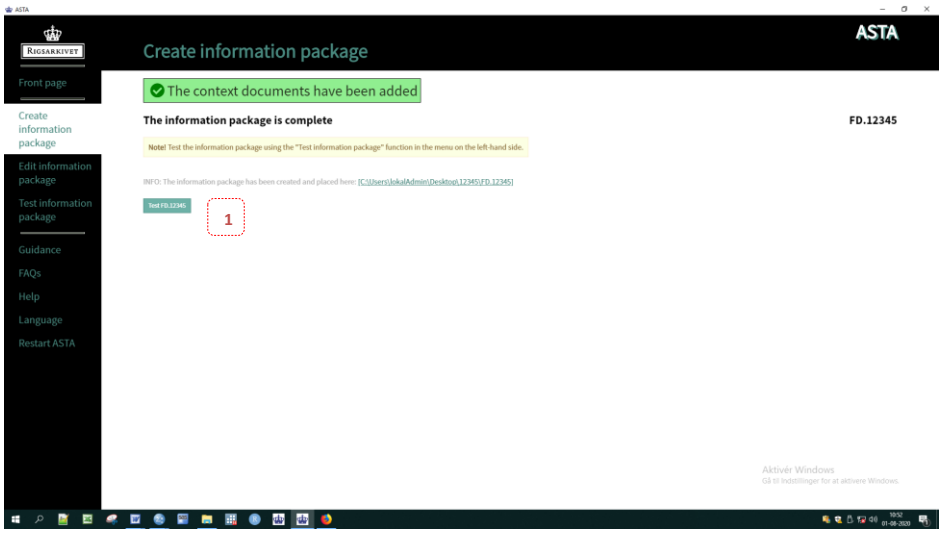


Figure 10: Page "The information package is complete" under "Create information package" in ASTA

G. Testing the information package

Test the information package

It is your responsibility to make sure that the information package fulfills all the provisions in Schedule 9 in the Executive Order on Information Packages. Hence, it is important for you to test the information package and adjust eventual errors, prior to delivering the information package to the Archives.

You can test the information package under the menu 'Test information package', see figure 11.

1. Choose the information package you will test (e.g. FD.12345) by clicking on the button '**Browse**'
2. Click on the button **'Start test'**
3. The test has now started and the test status is shown in the test log, se figure 11.

Test status and correction of errors in the test log

1. Thoroughly review the error messages marked with a red cross ✗ in the test log (4 in figure 11) and always adjust the errors before submitting to the Archives. **Hints** are shown as warnings with a yellow warning sign ⚠. Hints are not always errors and should only be corrected if they are in fact an error or missing element in the information package.

2. Apart from the test log that is shown in the screen, a **html-version of the test log** is also created and, in the end of the test, it is saved automatically in the same place as you have chosen to place the information package (see figure 12). If you click on the link to this test log (5 in figure 11), it will open in your standard browser and it can be printed with the help of the browser's print function (see

figure 9c). The html-test log includes also instructions as to how the error message should be interpreted and corrected (see figure 13).

3. When the test is finished, the following **status message** is shown:
   The information package is tested with no errors and can be delivered to the Archives
   This indicated that the information package is ready to be delivered to the Archives. NOTE: As ASTA does not test for all the provisions in Schedule 9, the Archives can later communicate other errors and missing elements that need to be corrected.

   The information package is tested. Errors shown in the log should be corrected prior to submitting to the Archives.
   This indicates that the test is finished, but there are still some errors to be adjusted. When you have adjusted for those errors in the test log, you need to run a new test and make sure that there are no other errors

   The test was interrupted and the test of the information package is not finished. Errors shown need to be corrected before the test can continue.
   This indicates that there are errors to be corrected before the test can continue. When you have corrected the errors in the test log and run the new test, other errors may occur. In severe tests, there is a maximum of 40 of the same type of errors shown in the test log. That means that when you correct those 40 errors and run a new test, there can be more errors of the same type.

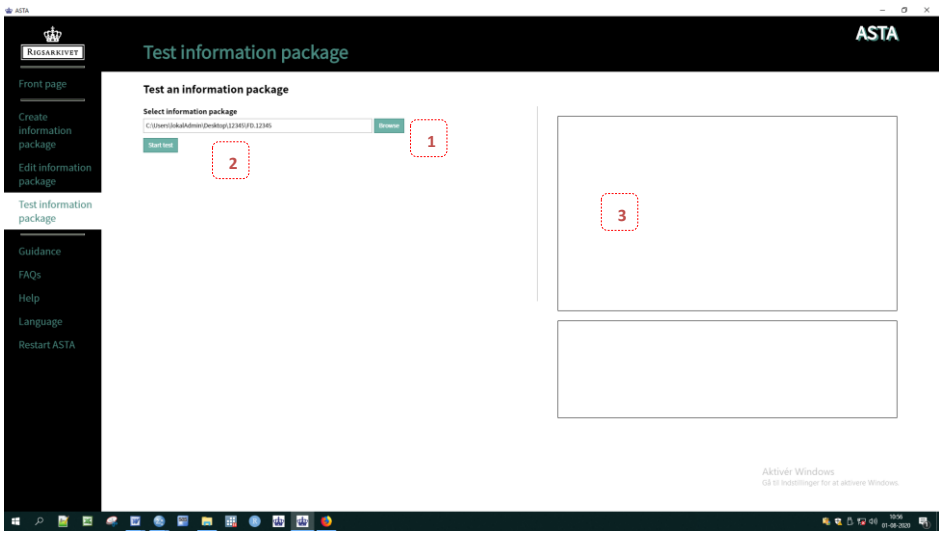


Figure 11: Page 'Test an information package' in ASTA before the test starts

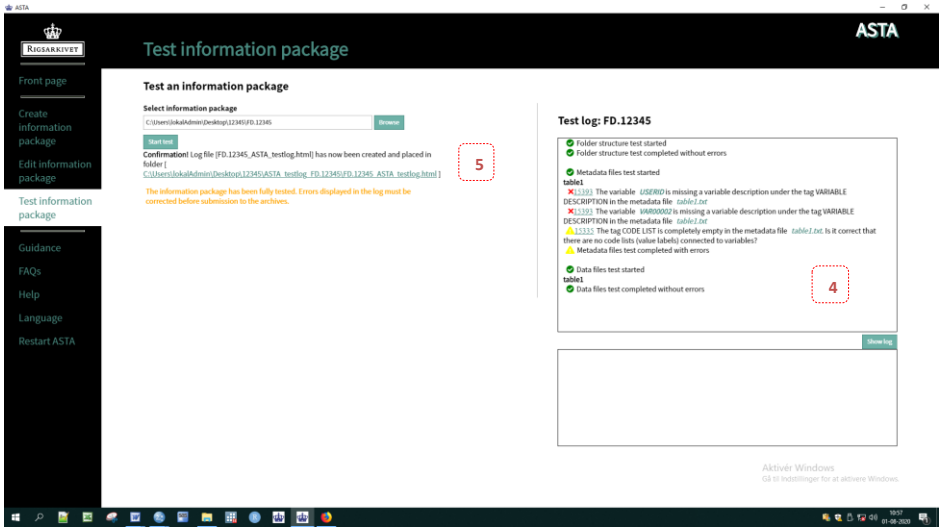


Figure 12: Page 'Test an information package' in ASTA when the test is completed

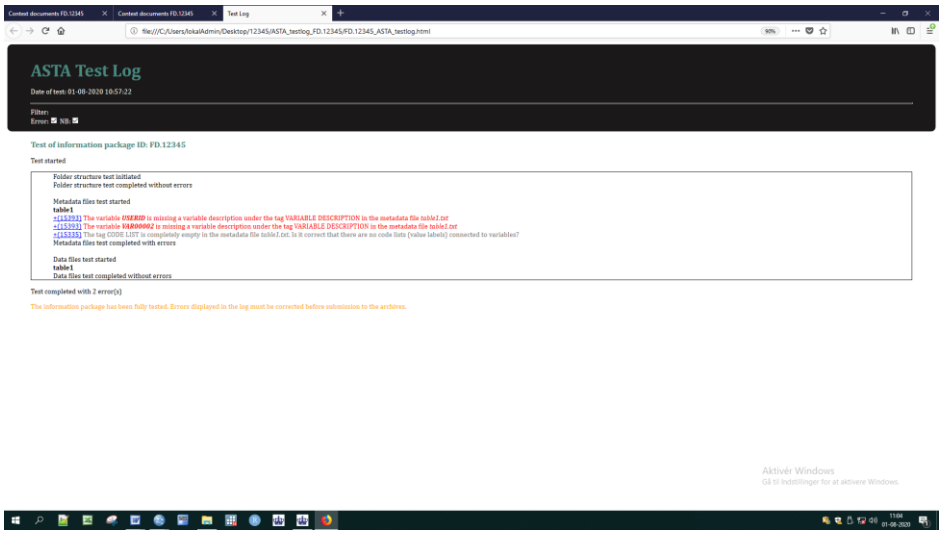


Figure 13: Html-version of the test log from 'Test an information package' in ASTA

Html-version of the test log with instructions

1. In the html-test log that opens when you click on the link in ASTA to this test log (see 5 in figure 12) an error-ID is shown (e.g. 15393) for each of the errors. By clicking in the error-ID, instructions appear about how the errors should be interpreted and corrected. It also shows an example of data and metadata without error, as well as the exact requirements as they are written in the provisions in Schedule 9 in the Executive Order on Information Packages (see figure 14).

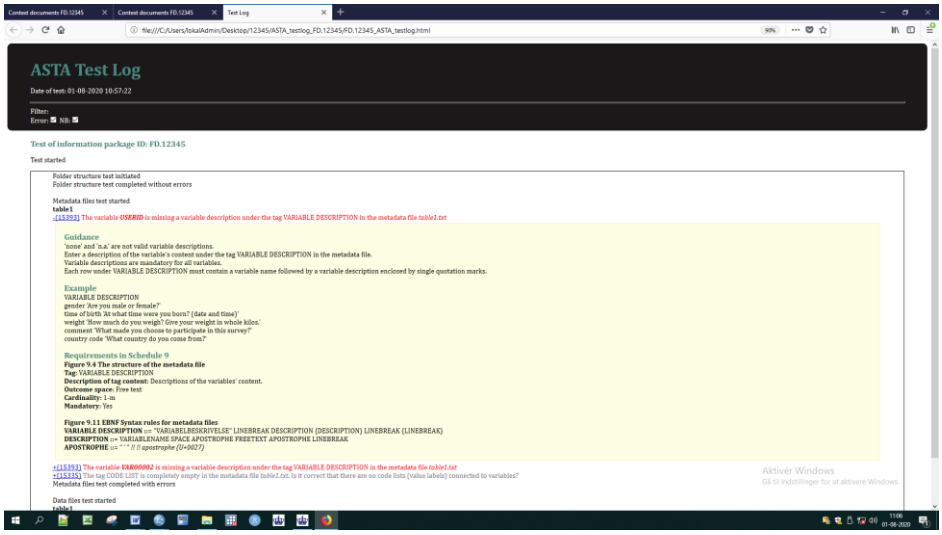


Figure 14: the Html-test log with the instructions after clicking on the error-ID 15393

H. Control that ASTA extraction has no losses

It is important that the data delivered to the Archives is authentic and corresponds to the contents in the original statistical files. It is the responsibility of the records creator to make sure that the extraction that is made with ASTA or other tools is complete, without losses. This could be done in many different ways, both automatically and visually. Below there are listed examples of what should be controlled along with how to do it.

Automatic control of the extraction in SPSS

This quality control can be done in SPSS by clicking on running the compare syntax, in the statistical program that compares the original statistical file with the output statistical file that ASTA has created. The compare syntax in SPSS is included in the SPSS export script that ASTA created. The export script executes an automatic quality control of the extraction from that ASTA makes from an SPSS file. The results of this comparison is saved automatically in the logfile called *statistikfilnavn_Exportscriptlog.spv*, in the same location where the original statistical file is on your computer. You should open that log file in SPSS and control the differences between the original file and the output one. The comparison results are in the bottom of the log file under the heading ”Compare datasets”.

Syntax to quality control the extraction from SAS and SPSS are not included in the export script in ASTA.

Visual control

Aside from the automatic control, one should also control visually the data before and after the transformation, as well as check that the metadata in the metadata file (e.g. table1.txt) is extracted from the data file correctly and with no errors.

Copying the control files

Under ASTA extraction, you have the possibility to copy all relevant files that can support the quality control of the extraction in the folder 'ASTA_kontrolfiler_FD.XXXXX'. Among those file is the chosen (original) statistical file that the extraction is made from, along with eventual output files and logs that ASTA generated under the extraction.

Control of the placing of context documents

It is recommended that under the extraction with ASTA that you print the document list of documents included (see figure 9c) and use it to control that all the context documents are placed in the correct document folder.

I. Editing the information package

Under the menu item 'Edit information package', you can edit the information package regarding the placement of the index files and the context documents, see figure 15.

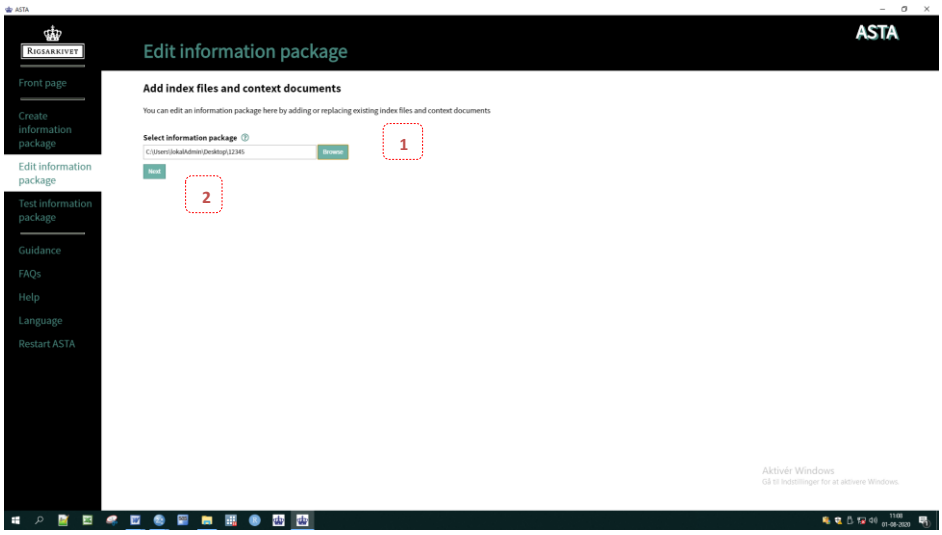


Figure 15: Page "Add index files and context documents" under the menu item "Edit information package" in ASTA

1. Choose the information package (ie FD.12345) to be edited by clicking on the button '**Browse**'
2. Click on the button '**Next**' to go to the next page

Placing index files in the information package

By clicking on the button 'Next' on the previous page, a new page opens, as shown in figure 14. You can place the index files *archiveIndex.xml* and *contextDocumentationIndex.xml* in the information package if they are missing or replace them with newer updated version of the files (1 and 2).

1. Choose the file *archiveIndex.xml* by clicking on the button '**Browse**' if that document is missing or you want to replace it with a new updated version.
2. Choose the file *contextDocumentationIndex.xml* by clicking on the button '**Browse**' if that document is missing or you want to replace it with a new updated version.
3. Click on the button '**Next'** to upload the selected file(s). Otherwise, you can continue to use the already existing index files.

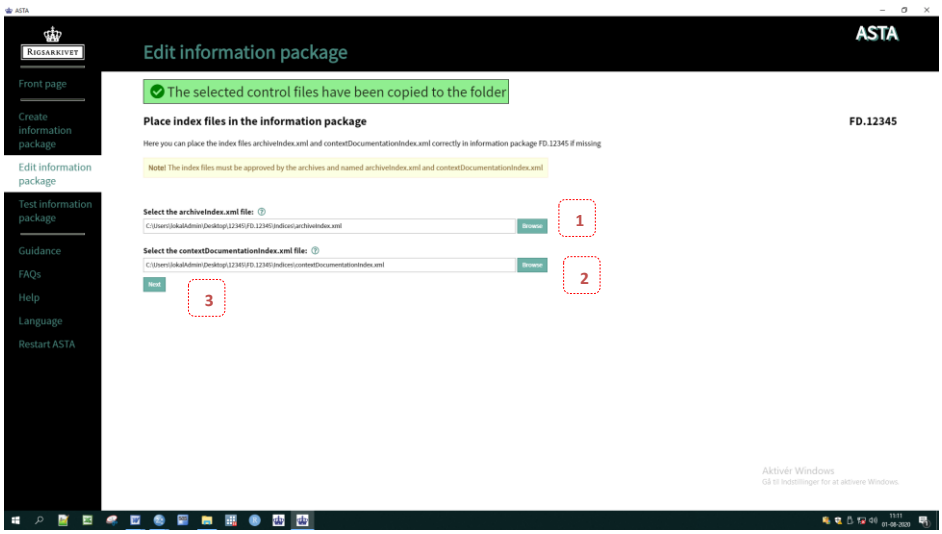


Figure 16: Page "Place index files in the information package" under the menu item "Edit information package" in ASTA

Placing context documents in the information package

By clicking on the button 'Next' on the previous page, a new page appears, as shown in figure 17.

You can place context documents in the information package if they are missing or replace them with newer updated ones.

1. All context documents that are included in the *contextDocumentationIndex.xml* are shown in the document list. Click on the button '**Browse**' beside each document and choose the context document that is related to the document title in the document list. If there already is a document in the folder, a link to its placement is shown. Click on the button 'Browse' if you wish to replace the existing context document in the folder with a new one.
2. When you have finished placing all context documents, we recommend that you print out the document list by clicking on the button '**Print document list**' and check that the content of the

chosen documents corresponds with the document title. When you click on the button 'Print document list' the document list is shown in your standard browser, e.g. Edge (see figure 8b). Use the browser print function to print the list (see figure 8c). When you click on the button 'Print document list', the document list is saved automatically in an html-version in the same area where the information package is located.

3. Click on the button '**Next**' in order to upload the chosen document(s). The screenshot on figure 10 shows the confirmation that the process of creating the information package is finished. Click on the button 'Test FD.XXXXX' to test the information package.

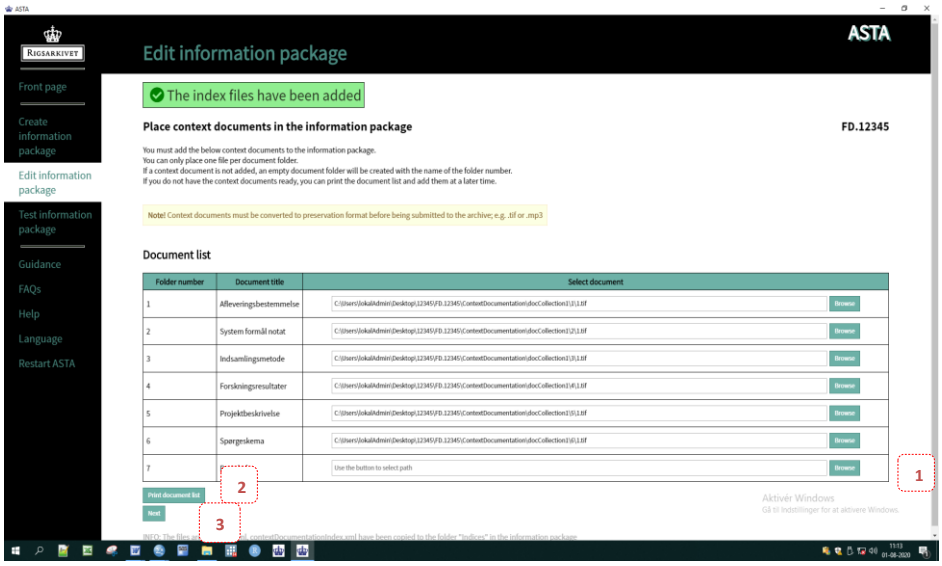


Figure 17: Page "Place context documents in the information package" under the menu item *"Edit information package" in ASTA*

J. ASTA support in the National Archives

If you experience problems while you create and test the information package with the help of ASTA, please contact the data manager for research data in the Danish National Archives at the following e-mail address: mailbox@sa.dk.

Appendix 1

This appendix gives a detailed overview of how you can make sure that the check items 1-8 are present when using specifically SPSS, Stata or SAS statistical program

A. SPSS – preparing the statistical file for extraction

SPSS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
1. Check/install statistical program	Check that you have installed SPSS on the PC where ASTA is being used.		You cannot extract data from an SPSS file with ASTA if you do not have SPSS installed in your PC.
2. New folder with a copy of the material	Create a new folder on your computer with a copy of all data and documents that will be submitted to the Archives. In other words, the statistical files, context documents cited in the archival provisions and the two index files that will be submitted. If possible, name the folder with the information package number that is stated in the archival provisions (afleveringsbestemmelse).	As the preparation for submission involves changes in variables, adjustments in formats and conversion of files, it is important to preserve the integrity of the original file(s) by separating them from the submission material.	The original data can be overwritten or even deleted.
SPSS DATA FILES			
3. Data files	Check that the SPSS files have the extension .sav	Data must be in a data format that ASTA can extract from. For SPSS files, the extension that can be used is .sav.	ASTA can only extract data from SPSS, SAS, Stata and R/RStudio formats
	Check that the data files are encoded as UTF-8	All data files should be encoded as UTF-8. Tip: On the lower right side of SPSS window, you can see “Unicode: ON” or “Unicode: OFF”. Adjust the encoding so that it is ”ON” (This means SPSS is using UTF-8). You should further check that your dataset shows all the special characters, such as æ, ø and å correctly.	If the data is not encoded as UTF-8, some characters, such as æ, ø and å will not be properly visualized. After the submission to the Archives, the information package will be tested. If the data is not correctly encoded as UTF-8, other errors may occur in the submission.

SPSS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
		See the guide about UTF-8 in the Archives homepage.	Those errors can only be adjusted by creating a new extraction from the original statistical files. Hence, to save time, one should control the encoding before submission.
4. Variable in data files	Check that all variable names are in accordance with the archive provisions for naming variables.	Variable name must not be longer than 128 characters, start with a number or include a blank character (space) or other characters other than letters or numbers. Use "Variable View" tab to see all your attributes such as variable names, labels, types, width and decimals. Adjust each variable that does not follow the Archives submission rules.	ASTA will show an error under the testing of the information package.
	Check that all variables have labels.	All variables must have variable labels that describe their contents. Use "Variable View" tab, under column "Label", to make sure that all variables have labels.	ASTA will show an error under the testing of the information package.
5. Variable type/format	Check that all variables have type/format that are allowed by the National Archives and not "default formats"	All variables should have defined type/format that are allowed by the National Archives and not "default formats". Use "Variable View" tab, under column "Type" to ensure that all the variables have the allowed type/format.	ASTA will show an error under the testing of the information package.
6. Value labels	Check that all value labels are unique and specified.	All values in the value labels must be unique and specified. In case some values are not specified (for example, in the	After submission to the Archives, the information package will be tested to ensure

SPSS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
		case of Likert Scales where only the first and the last values have specific value labels), the variable description (variable label) should describe this. It can, for example, include the text ”Scale used. Not all values are described in the value labels”. Variable labels can be changed in the “Variable View” tab, in the column ”Label” in SPSS.	that value labels are in place.
7. Missing values	Control that only numeric and categorical variables have missing codes in the ”Missing” column.	Missing values can only be used in numeric and categorical variables. Contact the Archives’ data manager, if you use codes for missing values in character (text/string), time or date variables.	ASTA will show an error under the testing of the information package.
	All missing value codes should be described in that specific variable value labels where the missing value is used.	It is important to understand the missing value code assigned to understand the data. Hence, all assigned codes for missing values should also be assigned a value in that specific variable value labels. e.g. 9 = Missing response 10 = Irrelevant 11 = Dropout	ASTA will show an error under the testing of the information package.
8. References (in submissions of more than one dataset)	Check that eventual merging variables have the same type/format and length across these related datasets.	Merging variables should be pointed out in the information package in order to describe the possibility of merging the related datasets. When there are several datasets in the information package that have a relation with each other, the	Missing conformity in type and length of merging variables will affect the possibility to merge submitted datasets. After submission to the archives, there will be control for the

SPSS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
		reference (merging variable) should have the same type/format and length across the datasets. Check that the format, type and length are identical in the ”Variable view” tab for each merge variable.	compliance of this rule. Save yourself time by doing it beforehand.

B. Stata – preparing the statistical file for extraction

Stata			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
1. Check/install statistical program	Check that you have Stata installed on the PC where ASTA is being used.		You cannot extract data from a Stata file with ASTA if you do not have Stata installed on your PC.
2. New folder with a copy of the material	Create a new folder on your computer with a copy of all data and documents that will be submitted to the Archives. In other words, the statistical files, context documents cited in the archival provisions and the two index files that will be submitted. If possible, name the folder with the information package number that is stated in the archival provisions (afleveringsbestemmelse).	As the preparation for submission involves changes in variables, adjustments in formats and conversion of files, it is important to preserve the integrity of the original file(s) by separating them from the submission material.	The original data can be overwritten or even deleted.
STATA DATA FILES			
3. Data files	Check that the Stata files have the extension .dta	Data must be in a data format that ASTA can extract from. For Stata files, the extension that can be used is *.dta*.	ASTA can only extract data from SPSS, SAS, Stata and R/RStudio formats
	Check that the data files are encoded as UTF-8	All data files should be encoded as UTF-8. You should further check that your dataset shows all the special characters, such as æ, ø and å correctly. See the guide about UTF-8 in the Archives homepage.	If the data is not encoded as UTF-8, some characters, such as æ, ø and å will not be properly visualized. After the submission to the Archives, the information package will be tested. If the data is not correctly encoded as UTF-8, other errors may occur in the submission. Those errors can only be adjusted by creating a new extraction from the

Stata			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
			original statistical files. Hence, to save time, one should control the encoding before submission.
4. Variable in data files	Check that all variable names are in accordance with the archive provisions for naming variables.	Variable name must not be longer than 128 characters, start with a number or include a blank character (space) or other characters other than letters or numbers. Use ”Property” window (on the bottom right of the main window in Stata) to see all your variables, their attributes such as names, labels and formats. Adjust each variable that does not follow the Archives submission rules.	ASTA will show an error under the testing of the information package.
	Check that all variables have labels.	All variables must have variable labels that describe their contents. Look at all the variables to make sure that all variables have labels. Use ”Variables Manager” option under the menu option ”Data” to change or to include variable labels.	ASTA will show an error under the testing of the information package.
5. Variable type/format	Check that all variables have type/format that are allowed by the National Archives and not “default formats”	All variables should have defined type/format that are allowed by the National Archives and not “default formats”. Use ”Properties” window to ensure that all the variables have the allowed type/format. In case they do not, use ”Variables Manager” option under the menu option ”Data” to change or include variable labels.	ASTA will show an error under the testing of the information package.

Stata			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
6. Value labels	Check that all value labels are unique and specified.	All values in the value labels must be unique and specified. In case some values are not specified (for example, in the case of Likert Scales where only the first and the last values have specific value labels), the variable description (variable label) should describe this. It can, for example, include the text ”Scale used. Not all values are described in the value labels”. Variable labels can be changed in ”Variables Manager” under the menu option ‘Data’ in Stata.	After submission to the Archives, the information package will be tested to ensure that value labels are in place.
7. Missing values	Control that only numeric and categorical variables have missing codes.	Missing values can only be used in numeric and categorical variables Control that only numeric and categorical variables have missing codes that are described in the “value labels” column in the “Variables Manager” option under the menu item ”Data” in Stata. Codes for missing values in Stata (user defined missing codes) must fall within the range between .a-.z. Contact the Archives’ data manager, if you use codes for missing values in character (text/string), time or date variables.	ASTA will show an error under the testing of the information package.

Stata			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
	All missing value codes should be described in that specific variable value labels where the missing value is used.	It is important to understand the missing value code assigned to understand the data. Hence, all assigned codes for missing values should also be assigned a value in that specific variable value labels. e.g. .u = Missing response .i = Irrelevant .d = Dropout You can include value labels for missing code in the value labels under ”Variables Manager” in Stata.	ASTA will show an error under the testing of the information package.
8. References (in submissions of more than one dataset)	Check that eventual merging variables have the same type/format and length across these related datasets.	Merging variables should be pointed out in the information package in order to describe the possibility of merging the related datasets. 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. Check that the format, type and length are identical in ”Properties” option in Stata for each of the files submitted.	Missing conformity in type and length of merging variables will affect the possibility to merge submitted datasets. After submission to the archives, there will be control for the compliance of this rule. Save yourself time by doing it beforehand.

C. SAS – preparing the statistical file for extraction

SAS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
1. Check/install statistical program	Check that you have SAS installed on the PC where ASTA is being used.		You cannot extract data from an SAS file with ASTA if you do not have SAS installed on your PC.
2. New folder with a copy of the material	Create a new folder on your computer with a copy of all data and documents that will be submitted to the Archives. In other words, the statistical files, context documents cited in the archival provisions and the two index files that will be submitted. If possible, name the folder with the information package number that is stated in the archival provisions (afleveringsbestemmelse).	As the preparation for submission involves changes in variables, adjustments in formats and conversion of files, it is important to preserve the integrity of the original file(s) by separating them from the submission material.	The original data can be overwritten or even deleted.
SAS DATA FILES			
3. Data files	Check that the SAS files have the extension *.sas7bdat*	Data must be in a data format that ASTA can extract from. For SAS files, the extension that can be used is *.sas7bdat to data files and .sas7bcat to catalogue files (format files)* Note that both data as their respective catalogue files must have the same name in order to extract the formats from the catalogue file with ASTA.	ASTA can only extract data from SPSS, SAS, Stata and R/RStudio formats
	Check that the data files are encoded as UTF-8	All data files should be encoded as UTF-8. You should further check that your dataset shows all the special characters, such as æ, ø and å correctly.	If the data is not encoded as UTF-8, some characters, such as æ, ø and å will not be properly visualized. After the submission to the Archives, the

SAS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
		See the guide about UTF-8 in the Archives homepage.	information package will be tested. If the data is not correctly encoded as UTF-8, other errors may occur in the submission. Those errors can only be adjusted by creating a new extraction from the original statistical files. Hence, to save time, one should control the encoding before submission.
4. Variable in data files	Check that all variable names are in accordance with the archive provisions for naming variables.	Variable name must not be longer than 128 characters, start with a number or include a blank character (space) or other characters other than letters or numbers. Use 'Proc Contents' in SAS to see all your variables, their attributes such as names, labels and formats. Adjust each variable that does not follow the Archives submission rules.	ASTA will show an error under the testing of the information package.

	Check that all variables have labels.	All variables must have variable labels that describe their contents.
		
		Use 'Proc Contents' in SAS to control that all the variables have labels. In case a variable does not have a value, a "Label statement" can include it when running the following syntax in SAS:
		
		`'LABEL variable = "label" … ;'`
		
		e.g.
		`'LABEL koen = "Respondentens køn" alder = "Respondentens alder";'`
			ASTA will show an error under the testing of the information package.

SAS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
5. Variable type/format	Check that all variables have type/format that are allowed by the National Archives and not "default formats"	All variables should have defined type/format that are allowed by the National Archives and not "default formats". This step is particularly important for SAS users, as the standard for SAS is NOT to include the formats. Hence, typical standard formats (default formats) are assigned, unless the user explicitly designates a format. In order to check the format, run a 'Proc Contents' procedure in SAS and visually control that the output format in the "Format" column is not a default format. That column shall show a format that is assigned in the format catalog, assigned via a 'Proc Format' procedure) or a format assigned via a 'Format statement', such as this: 'FORMAT variable-1 <. . . variable-n> <format>;' e.g. 'FORMAT koen koen. alder f3.;'	ASTA will show an error under the testing of the information package.
6. Value labels	Check that all value labels are unique and specified.	Check that all value labels are unique and specified.dataset and to control that all the levels and missing values are specified in the value label. If that is not the case, include the missing value labels, in this way: 'Proc format library=xxx.yyy select x:;'	After submission to the Archives, the information package will be tested to ensure that value labels are in place.

SAS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	

		e.g.: If you would like to check the format for ”age” variable that is in the “Formats” catalogue file in the library ”MyLib”, run the following syntax:
		`'Proc format library=MyLib.Formats select age:;'`
		
		e.g.: If you would like to check the formats in all variables in the “Formats” catalogue in the library ”Mylib”, you can run the following syntax:
		`'Proc format library=MyLib.Formats select a-z:;'`	

7. Missing values	Control that only numeric and categorical variables have missing codes.	Missing values can only be used in numeric and categorical variables Control that only numeric and categorical variables use codes for missing values that are described in the “value labels” by running the following syntax in SAS: `'Proc format library=xxx.yyy select x:;'` See the example under item 6 of this table. Code for missing values in SAS (user defined missing codes) must fall within the range between A-Z. Contact the Archives' data manager, if you use codes for missing values in character (text/string), time or date variables.	ASTA will show an error under the testing of the information package.

SAS			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
	All missing value codes should be described in that specific variable value labels where the missing value is used.	It is important to understand the missing value code assigned to understand the data. Hence, all assigned codes for missing values should also be assigned a value in that specific variable value labels. e.g. U = Missing response I = Irrelevant D = Dropout You can include value labels for missing code in the value labels via a 'Proc Format' procedure in SAS.	ASTA will show an error under the testing of the information package.
8. References (in submissions of more than one dataset)	Check that eventual merging variables have the same type/format and length across these related datasets.	Merging variables should be pointed out in the information package in order to describe the possibility of merging the related datasets. 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. Check that the format, type and length are identical by running a 'Proc Contents' procedure in SAS and control the type of the variables in the "Format" column and the length in the "Len" column.	Missing conformity in type and length of merging variables will affect the possibility to merge submitted datasets. After submission to the archives, there will be control for the compliance of this rule. Save time by doing it beforehand.

D. R/RStudio – preparing the statistical file(s) before extractionIn order to run in R/RStudio the syntax generated in ASTA, you need to install some plugins. Though, before installing them, you need to make sure that Rtools is installed because it is Rtools that allows you to install plugins. You can download Rtools here: https://cran.r-project.org/bin/windows/Rtools/.

Formateret: Engelsk (USA)

The following plugin needs to be installed before running your script from ASTA:

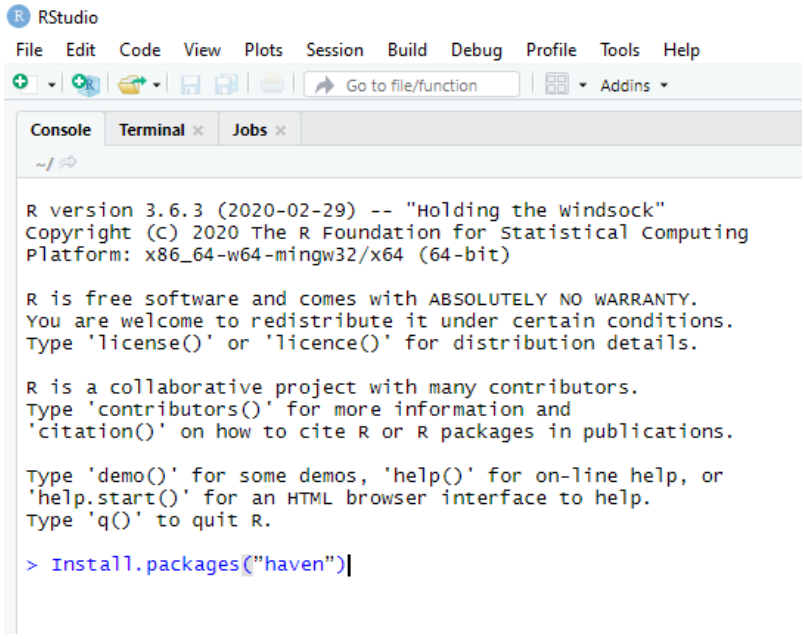
- Haven

This plugin allows you to convert time and date formats in R/RStudio to the formats allowed in regards to the Executive Order 128, Schedule 9 (see figure 9.3):

- Chron

Dates need to be converted to the formats allowed, prior to making the extraction with ASTA and the R/Rstudio export script.

To install these plugins (packages) in R/RStudio, you should write this command/syntax in your console area: `Install.packages ("plugin-name ")`, for example `Install.packages("haven")`, after it is downloaded from the internet. If your computer is not online, you can download the packages at https://cran.Rstudio.com/src/contrib/PACKAGES and then you must click on 'Install Packages' in the window to the right lower side of RStudio. That will access that package and run it into the program, by running the syntax `Library(plugin-name)`, e.g. `Library(haven)`.



We recommend that you extract data from R files, that are originated from R language with the R export script from ASTA. Even though R can import SAS, SPSS and Stata filer and save them as R format, there are

some transformation that happen under the importing of data into R that change the data, e.g. all user defined codes and special codes for missing values are changed to NA or space. Generally, there is a risk for a conversion mistake and loosing important data when other formats are read into R/RStudio.

R/RStudio			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
1. Check/install statistical program	Check that you have R/RStudio installed on the PC where ASTA is being used.		You cannot extract data from an R/RStudio file with ASTA if you do not have SAS installed on your PC.
2. New folder with a copy of the material	Create a new folder on your computer with a copy of all data and documents that will be submitted to the Archives. In other words, the statistical files, context documents cited in the archival provisions and the two index files that will be submitted. If possible, name the folder with the information package number that is stated in the archival provisions (afleveringsbestemmelse).	As the preparation for submission involver changes in variables, adjustments in formats and conversion of files, it is important to preserve the integrity of the original file(s) by separating them from the submission material.	The original data can be overwritten or even deleted.
R DATA FILES			
3. Data files	Check that R files have extension *.rds.*	Data must be in a data format that ASTA can extract from. For R files, the extensions that can be used are *.rds.*	ASTA can only extract data from SPSS, SAS, Stata and R/RStudio.

R/RStudio			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
	Check that the data files are encoded as UTF-8	All data files should be encoded as UTF-8. You should further check that your dataset shows all the special characters, such as æ, ø and å correctly. See the guide about UTF-8 in the Archives homepage.	If the data is not encoded as UTF-8, some characters, such as æ, ø and å will not be properly visualized. After the submission to the Archives, the information package will be tested. If the data is not correctly encoded as UTF-8, other errors may occur in the submission. Those errors can only be adjusted by creating a new extraction from the original statistical files. Hence, to save time, one should control the encoding before submission.

4. Variable in data files	Check that all variable names are in accordance with the archive provisions for naming variables.	To check all the variable names in the dataset(s), one can use the following syntax in R/Rstudio :
		
		`str(objekt,…)`
		
		Example of syntax, where the dataset analysed is called *mydata*:
		
		`str(mydata)`
		
		This syntax shows all the variables in the dataset mydata, their names, labels, factor levels (codes and code descriptions) and the first few values in the variables.
		
		You can change variabel names in R/RStudio - so that they conform to the provisions
			ASTA will show an error under the testing of the information package.

R/RStudio			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	

		regarding naming of variables - with the following syntax:
		
		`names(datasetname) [variableposition] = ”newvariablename”`
		
		Example of syntax where the da taset *mydata* has the first *[1]* va riable is called *cpr_nr*:
		
		`names(mydata) [1] = 'cpr _nr'`
		
		Remember to save the changes you make in R/RStudio, e.g. in a new file, you can use the following syntax to save the dataset with changes as a new R datafile:
		
		`saveRDS(object=datasetna me, file="path/datasetna me.extension")`
		
		Example of syntax, where the dataset *mydata* is saved with a new name *mydata2*:
		
		`saveRDS(object=mydata, f ile="C:/Users/lokalAdmin /Desktop/Testdata/mydata 2.rds")`	

	Check that all variables have labels.	All variables must have variable labels that describe their contents. It is a provision in Schedule 9. A statistical datasets that are originally generated in R/Rstudio most likely do not have variable labels. Therefore, you have to include these variable labels to the dataset before the submission. You can include variable labels with the following syntax:	ASTA will show an error under the testing of the information package.

R/RStudio			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			

		`attributes(datasetname$variablename)$label <- "Description"`
		
		Example of syntax, where value labels *Respondentens cpr-nummer* is given to variable *cpr_nr* in the dataset *mydata*:
		
		`attributes(mydata$cpr_nr)$label <- "Respondentens cpr-nummer"`	

5. Variable type/format	Check that all variables have type/format that are allowed by the National Archives and not "default formats"	All variables should have defined type/format that are allowed by the National Archives and not "default formats". To check all the type/formats of the variables in the datasetone can run the `str(object,…)` syntax in R/Rstudio. **Example of syntax**: `str(mydata)` to see all the variable attributes in the file 'mydata' If one or more data format notations do not follow provisions shown in figure 9.3 in Schedule 9 in the Executive Order 128, the variable format notation(s) need to be changed to one that is accepted. All dates, timestamps and time variables in an R/RStudio dataset are character (string) variables. In other words, these format notations are not necessarily the correct ones for these variables. You must then be extra careful in checking these data types and including the correct format notation	ASTA will show an error under the testing of the information package.

R/RStudio			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
		when doing an extraction with ASTA. If those variables are string, they must be changed to the correct notation (e.g. xml notation of time, date and datetime depends on the contents of the dataset. If the content of the dataset does not comply with the given accepted data format notation, the variable, for instance date, must be converted to the correct date format notation in R dataset before the submission. See figure 9.3 in Schedule 9 about the accepted data types in a dataset. Plugin 'chron' to R/RStudio can be used to convert time- and date formats.	

6. Value labels	Check that all factor levels (value labels: codes and code descriptions) are unique and specified.	Check that all factor levels (value labels: codes and code descriptions) are unique and specified.
		
		Statistical datasets that are originally generated in R/RStudio most likely do not have factor levels/value labels defined. Therefore, you have to include the codes and code descriptions to the variables in the dataset before the submission. You can use the syntax below for that:
		
		`x <- c(0,1)`
		`names(x) <- c("X", "Y")`
		`attributes(datasetname$variablename)$labels <- x`
		
		Example of syntax, where factor levels will be included for the variable *sex* in the dataset *mydata* :
		
		`x <- c(1,2)`
			After submission to the Archives, the information package will be tested to ensure that value labels are in place.

R/RStudio			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
		GENERAL	
		`names(x) <- c("Male", "Female")` `attributes(mydata$sex)$labels <- x`	
7. Missing values	In R files it is not possible to provide values for missing values	Statistical datasets that are originally generated in R/RStudio most likely do not have codes for missing values. When an original dataset from SPSS, Stata or SAS is imported into R/RStudio, the existing codes for missing values are substituted for 'NA' or an empty space. Therefore, we do not recommend that these imported-and-saved as R datasets are extracted with ASTA.	ASTA will show an error under the testing of the information package.
8. References (in submissions of more than one dataset)	Check that eventual merging variables have the same type/format and length across these related datasets.	Merging variables should be pointed out in the information package in order to describe the possibility of merging the related datasets. 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. Check that the format, type and length are identical by running a `str(objekt, …)` in R/RStudio and controlling the type of the variables formats in the list in the output. **Example of syntax**:	Missing conformity in type and length of merging variables will affect the possibility to merge submitted datasets. After submission to the archives, there will be control for the compliance of this rule. Save yourself time by doing it beforehand.

R/RStudio			
NO.	**TASK**	**GUIDE**	**IF YOU DO NOT DO IT**
GENERAL			
		`str(mydata)` to see all the atributtes, among them formats, in the dataset mydata	