

# Quick guide

## - production and testing of an information package with ASTA

*The National Archives, July 2020*

*Version 2.0*

### What is ASTA?

The National Archives has created ASTA (Aflevering af Statistikfiler Til Arkiv – translation: Statistical File Submission to the Archives) tool that you can use to create and test an information package containing data extracted from statistical files that comply with the National Archives' requirements for an information package. These requirements are presented on Schedule 9 of the Executive Order on Information Packages. The Executive order and its instructions can be found on the National Archives' homepage [www.sa.dk](http://www.sa.dk).

When you create an information package with ASTA, the program ensures that the information package meets the structural requirements for folder structure and data and metadata files. On the other hand, you must ensure that data and metadata contents in the statistical files comply with the requirements set out by Schedule 9 before the extraction with ASTA.

### How to ensure the quality of the statistical file before extraction

The "Skab afleveringspakke" (translation: Create information package) function in ASTA simply extracts the data and metadata that is in the statistics file. It is, therefore, important, that you ensure that the quality of data and metadata in the statistical file meets the Archives' provisions listed in Schedule 9 of the Executive Order on Information Packages before you extract with ASTA.

Data and metadata quality requirements are listed below:

#### Data file

- The character set for the statistical file must be encoded as UTF-8. See the instructions on UTF-8.

#### Variable

- There must be no line breaks in the content of a variable.
- Variable names must be specified for all variables.
- Variable names may never begin with a number. However, they can be a mixture of letters and numbers.
- Variable names must not be over 128 characters long.
- Variable descriptions [Variable labels] must be specified for all variables.

### **Data types (Formats)**

- Data types for variables in the statistical file must comply with the allowed data types in an information package specified in Figure 9.3 of Schedule 9 of the Executive Order on Information Packages. If this is not the case, you must change the data types to the allowed formats yourself before extraction with ASTA.

### **Value labels/Formats**

- All values codes must be described with a code label [Value labels].
- Value labels can only be associated with numeric and categorical variables with numeric integer, decimal, and text data types.
- Values must be unique within the same value labels associated with a variable.

### **Missing values codes**

- Special codes for missing values (SAS and Stata) can only assume values A-Z or .a-.z., respectively.
- Special codes (SAS and Stata) for missing values may only be used for numeric and categorical variables with numerical integer and decimal data types.
- User-defined codes (SPSS) for missing values may only be used for numeric and categorical variables with numeric, decimal, and text data types.
- Consistent only special codes (.a-.z or A-Z) or custom codes for missing values may be used.
- If user-defined codes (SPSS) for missing values are applied to variables with data type date or timestamp, it must be described in a context document in the information package.
- Both special codes and user-defined missing values codes in a variable must always appear in the value labels for the given variable.
- If value labels are only specified for some of the values that appear on a scale, this should be described in the variable description for the variable.

### **Merge variables (references)**

- There must be data type-and-length consistency between all the merge variables that are part of a reference between multiple statistical files in the information package

Read Schedule 9 of the Executive Order on Information Packages for a detailed description of what an information package is and what the data and metadata files requirements are.

### **This must be ready before you create an information package with ASTA**

As it is not possible to undo or go back when using ASTA, you should have the files and information below ready before you create an information package with ASTA.



You can add or replace existing index and context documents in an information package afterward with the "Rediger afleveringspakke" (translation: Edit Information package) menu item.

- **Identify the statistical files** to be delivered to the Archives.

**Information!** If sensitive information is in a key file, placed separately from the statistical file(s), extracts from both the statistical file and the key file must be included in the information package.

- **Identify the serial number** of the information package. The serial number is provided by the National Archives and appears in the archival provisions (*afleveringsbestemmelse*).
- **Fill in the** `archiveIndex.xml` and `contextDocumentationIndex.xml` index files. The National Archives has developed tools for producing these index files. See the instructions for `Create archiveIndex` and `Create contextDocumentationIndex` on the National Archives' homepage [www.sa.dk](http://www.sa.dk).

**Information!** The Archives must approve the filled-out index files before placing them in the information package.

- **Identify the context documents** to be included in the information package. These are listed in the archival provisions (*afleveringsbestemmelse*).
- **Convert the context documents** to the format accepted by the Archives, e.g. `.tif` (documents and images), `.mp3` (audio) and `.mpg` (video). See instructions for TIFF conversion on the National Archives' homepage [www.sa.dk](http://www.sa.dk).

**Information!** You can also convert the context documents after the information package is created.

- **Determine if there is a unique** variable(s) in the statistical files that must be specified as the key variable(s) during extraction with ASTA.
- **Determine if there are merge** variables across multiple statistical files in the information package to be specified when reference(s) is informed under the extraction with ASTA, as well as the, which variables are under the reference(s).



## These parts of the information package can be created with ASTA

Use the "Skab afleveringspakke" (translation: Create Information package) menu item to do the following:

- Name the information package with the correct serial number.
- Create an empty folder structure for the information package with the structure and naming of folders that meet the Schedule 9 requirements.
- Extract data and metadata from your original statistical files to data and metadata files that meet the structural requirements for these files as specified in Schedule 9.
- Add archiveIndex.xml and contextDocumentationIndex.xml documents to the information package. See the instructions in 'Guide to Skab archiveIndex' and 'Guide to Skab contextDocumentationIndex'.
- Add context documents to the information package.

## These parts of the information package you must do yourself

In ASTA, you cannot:

- Fill in the index files
- Convert the context documents to the accepted preservation format, e.g. .tif or .mp3

## How to check if extraction with ASTA is without errors

It is important that the data submitted to the National Archives is authentic and corresponds to the content of the original statistical file(s). It is the responsibility of the contracting authority to ensure that ASTA extraction, or extraction made with other tools, has no errors. This can be done in many ways, both automatically and visually. Below you will find a list of examples of what needs to be checked and how it can be done.

### Automatic control of an extraction from an SPSS dataset

In SPSS, this quality control can be performed by running a compare syntax in the statistics program that compares the original statistical file to the output statistical file that ASTA creates. The compare syntax of SPSS is incorporated into the export script for SPSS that ASTA forms. When running the export script created by ASTA, quality control of the extracts from SPSS made with ASTA is automatically performed. The result of this comparison is automatically stored in a log file, named Exportscriptlog.spv, at the same destination where the selected original statistical file is located. You should open this log file in SPSS and check for the differences between the original file and the output file. The result of the comparison is at the bottom of the log under the heading "Compare datasets".

Compare syntaxes for SAS and Stata that are not included in the export scripts in ASTA.

### Visual control



In addition to the automatic check, you should also visually check data before and after the extraction, as well as check whether metadata, in the metadata file (e.g. table1.txt), is extracted from the data file correctly and with no errors.

### To test an information package

Use the "Test afleveringspakke" (translation: Test Information package) menu option this way:

- Select the information package that you want to test and click the "Start test" button
- Read any error messages shown in the test log
- Click the error ID in the html version of the test log, which results from clicking the "Vis log" (translation: Show log) button to see instructions on how to correct the error(s)
- Correct the error(s) in the test log and test the information package again
- Once the information package has been tested completely and it has no errors, it can be submitted to the Archives, according to the submission rules set out on the National Archives homepage [www.sa.dk](http://www.sa.dk).

### How to fix errors in an existing information package

- **Errors in the data and metadata files** are corrected directly in the statistical file and a new extraction needs to be made with ASTA.

If the error cannot be corrected in the statistical file, it is then allowed to make corrections directly in the data file (e.g. table1.csv) and the metadata file (e.g. table1.txt) extracted from the statistical file. However, these errors should be corrected as a very last step, when one is sure that there are no more errors that need to be corrected by making new extractions from the statistical file.

- **Errors in the index files** can be corrected in the National Archives' tools that create them. The tools validate whether the files meet the index file requirements. Changes to the index files must be approved by the Archives.
- **Errors in context documents** are corrected by replacing the existing context document that has an error with the correct one.

**Information!** If new context documents are added to the information package and they are not included in the archival provisions (*afleveringsbestemmelse*), these provisions must be updated to include these documents and approved by the Archives.