This is an English translation of the Executive Order on Information Packages (*Bekendtgørelse om arkiveringsversioner nr. 128*). In the event of a conflict between the Danish version and the English translation, the Danish version shall supersede.

Executive Order on Information Packages

Pursuant to Sections 5(1), 13(1) and 14 of Executive Order no. 591 of 26 June 2003 on public archives and the activities of public archives and following consultation with the municipal and regional parties, the following is stipulated:

Scope

1. The provisions of this Executive Order shall apply to records created by the public administration and the courts and determined by the National Archives to be worthy of preservation.

2. The preservation of digital data must take place in the form of information packages.
(2) Audio and video worthy of preservation are digitised prior to submission to the public archive and preserved as information package, cf. Section 2(1)
(3) Where this is required in the interest of preservation, the National Archives may require that archives worthy of preservation are digitised prior to submission to the public archive and preserved as information packages, cf. Section 2(1)

3. An information package of data worthy of preservation shall be produced in accordance with the instructions as indicated by schedules 2-8 of this Executive Order.
(2) Data created in connection with research using scientific methods and created or processed in statistical programs or similar shall be submitted as an information package in accordance with the rules in Schedule 9.
(3) The National Archivist may designate forms of preservation other than information packages, prepared in accordance with Schedule 2-9, if this is required in the interest of preservation.

Production and submission

4. The deadlines for the production and submission of information packages of data from the state administration and the courts' IT systems shall be fixed by the National Archives.

5. Information packages of data from the IT systems of the municipal and regional authorities containing personal data shall be produced and submitted before data is deleted. The individual municipal or regional authority may make agreements on earlier deadlines with the receiving archive.
(2) The production of information packages of data from other IT systems shall take place before data is deleted or when the IT system is decommissioned.
(3) In addition, the National Archives may decide that an information package of data from an IT system must be produced when this is required in the interest of preservation.

6. If the submitting authority, in connection with the production of the information package, finds errors or deficiencies in the document conversion or in the conversion of table contents, an agreement shall be made with the receiving archive for rectification thereof before the information package is submitted.

7. Information packages of state authorities' data shall be approved by the National Archives. Information packages of municipal and regional authorities' data shall be approved by the receiving archive.

(2) Information packages of municipal and regional data worthy of preservation shall be approved by the same public archives which receive the authorities' information packages that are subject to a submission obligation.

(3) Data transferred to an information package shall not be deleted by the authorities before the information package has been approved.

(4) For research data submitted as an information package according to Schedule 9, data may only be deleted once the receiving archive has produced and approved an information package in accordance with the rules in Schedule 2-8 of the submitted information package.

Commencement etc.

8. This Executive Order shall enter into force on 1 March 2020
(2) Executive Order no. 1007 of 20 August 2010 on information packages shall be revoked.

Schedule 1 Schedules

Schedule 2 Graphical overview of elements and structure in an information package according to Schedule 3-8

Schedule 3 Information package of digital data and any documents
    3.A. General rules on information packages
    3.B. Data structure
    3.C. Data content
    3.D. Metadata describing the information package

Schedule 4 Data structure
    4.A. General rules on data structure
    4.B. Location of folders and files
    4.C. *Indices* folder
    4.D. *Tables* folder
    4.E. *ContextDocumentation* folder
    4.F. *Schemas* folder
    4.G. *Documents* folder

Schedule 5 Data content
    5.A. Table content
    5.B. Data types
    5.C. Conversion of table content to digital documents, audio, video or geodata
    5.D. Text format
    5.E. Digital documents
    5.F. Audio and video
    5.G. Geodata
    5.H. Compression
    5.I. Optimisation
    5.J. No deterioration

Metadata describing the information package
    6.A. Archive description file
    6.B. Context documentation
    6.C. Metadata describing the tables of the information package
    6.D. SQL queries

Schedule 7 Submission medium

Schedule 8 Forms

Schedule 9 Information package for certain types of research data
    9.A. Submission of research data
    9.B. Folder structure of an information package
    9.C. Indices folder
    9.D. Context Documentation folder

Graphical overview of elements and structure in an information package according to Schedule 3-8

Figure 2.1 Graphical overview of possible elements and structures of an information package according to Schedule 3-8

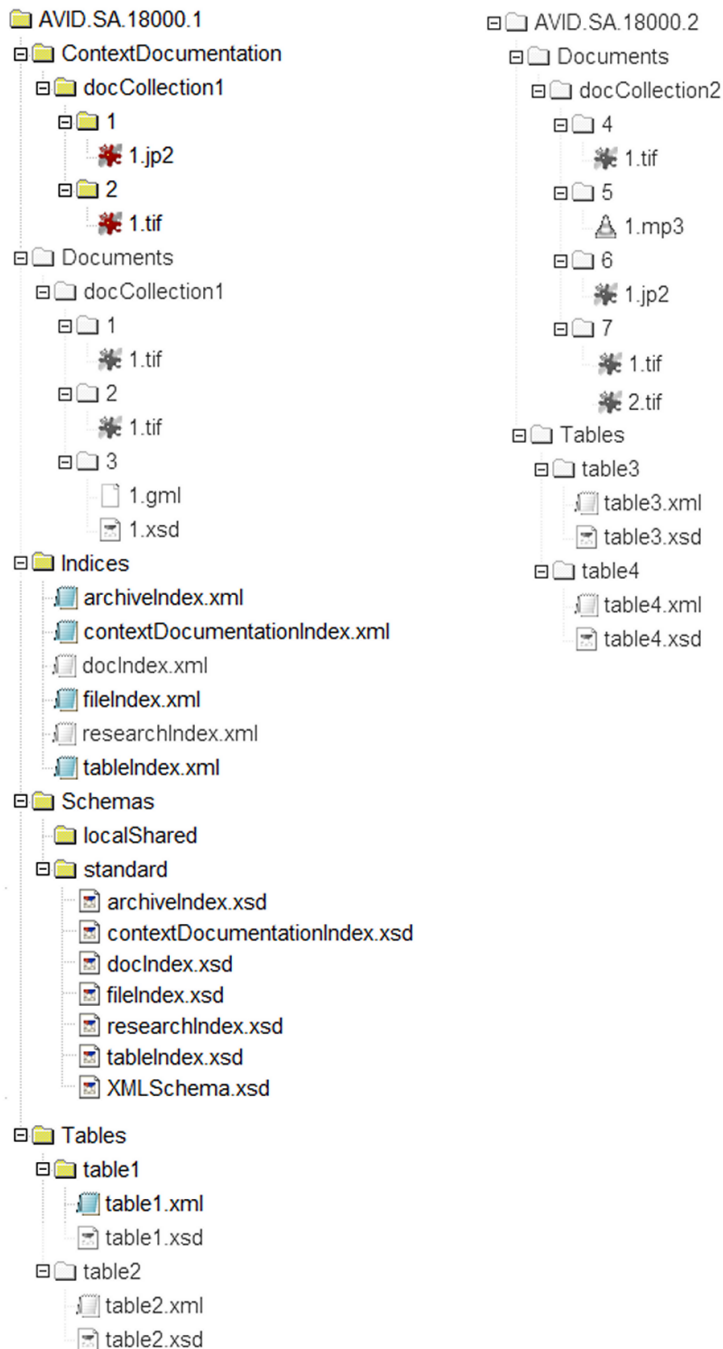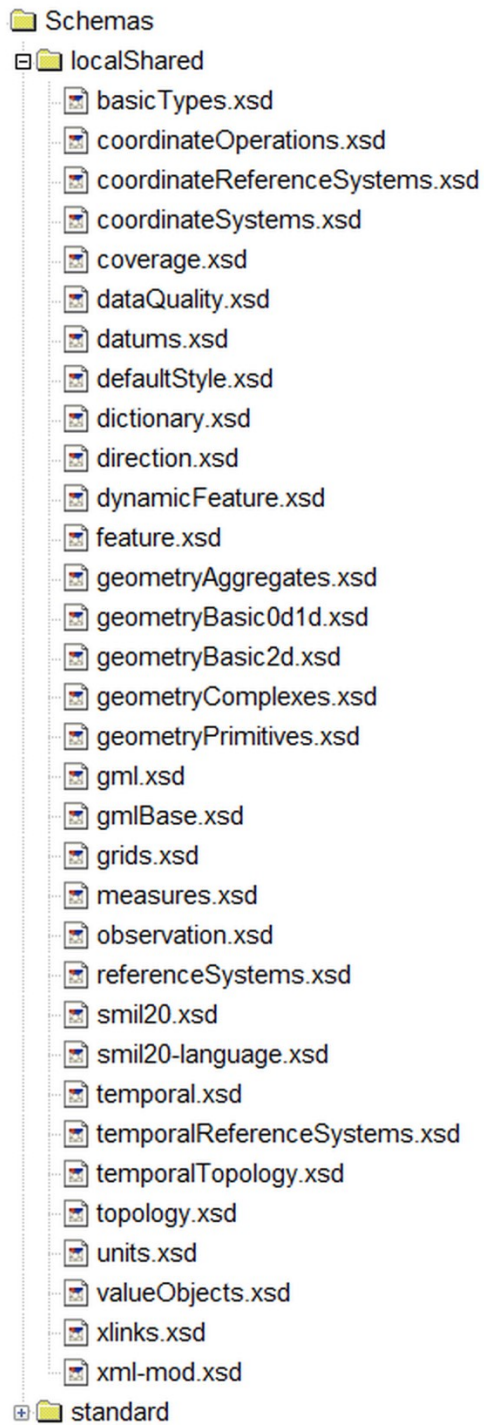Elements that are not grey are mandatory in an information package.

```
AVID.SA.18000.1                              AVID.SA.18000.2
   ContextDocumentation                         Documents
      docCollection1                               docCollection2
         1                                            4
            1.jp2                                        1.tif
         2                                            5
            1.tif                                        1.mp3
   Documents                                        6
      docCollection1                                   1.jp2
         1                                            7
            1.tif                                        1.tif
         2                                               2.tif
            1.tif                                   Tables
         3                                            table3
            1.gml                                       table3.xml
            1.xsd                                       table3.xsd
   Indices                                           table4
      archiveIndex.xml                                 table4.xml
      contextDocumentationIndex.xml                    table4.xsd
      docIndex.xml
      fileIndex.xml
      researchIndex.xml
      tableIndex.xml
   Schemas
      localShared
      standard
         archiveIndex.xsd
         contextDocumentationIndex.xsd
         docIndex.xsd
         fileIndex.xsd
         researchIndex.xsd
         tableIndex.xsd
         XMLSchema.xsd
   Tables
      table1
         table1.xml
         table1.xsd
      table2
         table2.xml
         table2.xsd
```

Figure 2.2 Location of schemas from OGC to validate GML version 3.1.1

- 📁 Schemas
  - 📁 localShared
    - 🗎 basicTypes.xsd
    - 🗎 coordinateOperations.xsd
    - 🗎 coordinateReferenceSystems.xsd
    - 🗎 coordinateSystems.xsd
    - 🗎 coverage.xsd
    - 🗎 dataQuality.xsd
    - 🗎 datums.xsd
    - 🗎 defaultStyle.xsd
    - 🗎 dictionary.xsd
    - 🗎 direction.xsd
    - 🗎 dynamicFeature.xsd
    - 🗎 feature.xsd
    - 🗎 geometryAggregates.xsd
    - 🗎 geometryBasic0d1d.xsd
    - 🗎 geometryBasic2d.xsd
    - 🗎 geometryComplexes.xsd
    - 🗎 geometryPrimitives.xsd
    - 🗎 gml.xsd
    - 🗎 gmlBase.xsd
    - 🗎 grids.xsd
    - 🗎 measures.xsd
    - 🗎 observation.xsd
    - 🗎 referenceSystems.xsd
    - 🗎 smil20.xsd
    - 🗎 smil20-language.xsd
    - 🗎 temporal.xsd
    - 🗎 temporalReferenceSystems.xsd
    - 🗎 temporalTopology.xsd
    - 🗎 topology.xsd
    - 🗎 units.xsd
    - 🗎 valueObjects.xsd
    - 🗎 xlinks.xsd
    - 🗎 xml-mod.xsd
  - 📁 standard

Information package of digital data and any documents

3.A. General rules on information packages

3.A.1 An information package consists of:
– data structure, cf. 3.B
– data content, cf. 3.C
– metadata describing the information package, cf. 3.D

3.A.2 An information package must not contain encrypted information.

3.A.3.a An information package must contain all data worthy of preservation and any documents from a defined period of time during which data are no longer corrected or added or must be in the form of a 'snapshot' containing all data worthy of preservation and any documents from a specific time. See, however, Section 3.A.3.b

3.A.3.b Provided that checks have been carried out that closed cases have been properly closed, the receiving archive may approve that an information package contains only documents from closed cases. However, metadata for all cases and documents, both completed and pending, must be included in the information package.

If the information package of an IT system with documents is prepared without changing the archive period, or if, in connection with a change of archive periods, documents are transferred to a new period, the documents included in the information package must be identified in the IT system to enable exclusion from subsequent information packages.

3.A.5 Smaller collections of documents created in connection with research using scientific methods can be submitted as part of the institution's Records Management System upon agreement with the receiving archive.

3.B. Data structure

3.B.1 The data structure in the information package consists of:
– a folder structure, cf. 4.B, illustrated in Figure 2.1
– a relational database structure on the first normal form or higher, specified in the index file tableIndex.xml, cf. 4.C.5.a
– other XML index files, cf. 4.C, structuring metadata describing the general content of the information package, all files in the information package, its context documentation and any digital documents

3.C. Data content

3.C.1 The data content consists of:
– table content in standardised data types, cf. 5.B
– any digital documents, audio, video and geodata in standardised data formats, cf. 5.E - 5.G - 5.G.

3.D. Metadata describing the information package

3.D.1 Metadata describing the information package consists of descriptions of administrative use, data content and IT system, cf. Schedule 6.

Data structure

4.A. General rules on data structure

4.A.1 All keys in an information package must have a unique identifier. Situations where it is necessary to extract parts of key fields to understand the content or function of the IT system must be avoided. A field that is part of a primary key must not consist solely of white space characters or be NULL.

4.A.2 When the value of a field is a code representing a fixed and unique value, the codes must be explained. If the value is not specified in a code or look-up table in the IT system, one or more tables must be generated in the information package representing the values of the code fields. Alternatively, either a field should be added to the data with the value that the code covers or the code value in table data must be replaced by the actual content when extracting data.

4.A.3 If the documents in an IT system with documents are stored using a significant structure, this structure must be converted to one or more tables in the information package.

4.B. Location of folders and files

4.B.1 In the root of the file system on the submission medium, cf. Schedule 7, a folder must be placed with the medium name. The medium name consists of the unique information package ID, with a suffix ».n« added, indicating the order of the media, where n is a consecutive media serial number starting with 1.

4.B.2 The content of the information package is placed in folders as specified in Figure 4.1.

Figure 4.1 Information package's folders

Folder name	Description	Manda-tory	Location
Indices	Index files, cf. 4.C	Yes	Always on the first medium
Tables	Table content, cf. 4.D	Yes	Optional location, may be distributed on several media
ContextDocumenta-tion	Context documentation, cf. 4.E	Yes	Always on the first medium
Schemas	XML schemas, cf. 4.F. The schema folder must have two subfolders: *standard* and *lo-calShared*	Yes	Always on the first medium
Documents	Any documents, audio, video and geodata, cf. 4.G	No	Optional location, may be distributed on several media

4.B.3 The folders must be named as specified in Figure 4.1.

4.B.4.a An information package ID consists of the prefix AVID, a 2-4 letter code indicating the receiving archive and an information package serial number. The elements are separated by dots.

4.B.4.b The information package ID is provided by the National Archives.

4.B.5.a If an information package is too large for one medium, cf. Schedule 7, it may be distributed on several media.

4.B.5.b The folders *ContextDocumentation, Indices* and *Schemas* must always be on the first medium submitted.

4.B.5.c The folder in the root of subsequent media must only contain the folders with content that must be distributed on several media. Such folders must always be named as described in Figure 4.1, without a suffix.

4.B.6 Each medium must only contain one of each of the folders specified in Figure 4.1.

4.C. Indices *folder*

4.C.1.a The Indices folder must contain the following index files with metadata describing the information package and its content:

– fileIndex.xml
– archiveIndex.xml
– contextDocumentationIndex.xml
– tableIndex.xml

4.C.1.b If the information package contains digital documents, audio, video or geodata, the *Indices* folder must also contain the following index file:

– docIndex.xml

4.C.1.c If the information package contains data created in connection with research using scientific methods and submitted in accordance with the rules in Schedule 9, the *Indices* folder shall also contain the following index file:

– researchIndex.xml

4.C.1.d All index files must conform to the relevant schema, cf. Schedule 8.

4.C.2.a **fileIndex.xml** must contain a complete list of all files in the information package. However, fileIndex.xml is exempt from this rule.

4.C.2.b The information shown in Figure 4.2 must be specified for each file in the information package.

Figure 4.2 Information in fileIndex.xml

Element name	Designation	Description	Cardinality per file	Man-datory
foN	Path	Specification of a path to the information package folder containing the file	1	Yes
fiN	Name	File name	1	Yes
md5	Check sum	The check sum of the file of the type MD5 according to *IETF RFC1321 - The MD5 Message-Digest Algorithm*: 128 bit (16 bytes) represented as 32 hexadecimal digits, all in *either* lower-case letters *or* capital letters (lower or upper case). The requirement for 32 hexadecimal digits means that prefixed zeros are required, depending on the value.	1	Yes

4.C.3 **archiveIndex.xml** must contain the information as shown in 6.A.

4.C.4.a **contextDocumentationIndex.xml** must contain an index of the documents in the context documentation of the information package.

4.C.4.b The information shown in Figure 4.3 must be specified for each document in the context documentation.

Figure 4.3 Information in contextDocumentationIndex.xml

Element name	Designation	Outcome space	Cardinality per document	Mandatory
documentID	Document ID	ID of up to 12 digits	1	Yes
documentTitle	Document title	Free text	1	Yes
documentDescription	Document description	Free text	0-1	No
documentDate	Date	Year, year-month, year-month-day according to the XML standard	0-1	No
authorName	Author name	Free text	0-m	No
authorInstitution	Author institution	Free text	0-m	No
documentCategory	Document category	Category according to Schedule 6, Figure 6.2	1-m	Yes

4.C.5.a **tableIndex.xml** must contain a specification of a relational database structure on the first normal form or higher. All tables in the information package must be specified.

4.C.5.b »tableIndex.xml« must conform to the general XML schema »tableIndex.xsd«, cf. 4.F.

4.C.5.c If a field can have the value NULL, tableIndex.xml must contain the value "true" for the column's appurtenant element "nullable".

4.C.6.a **docIndex.xml** must be the link between each document and its location. In addition, »docIndex.xml« must contain information on the original file names of the document, file type in the information package and any parent documents. »docIndex.xml« must not contain metadata describing the documents in the context documentation.

4.C.6.b The information shown in Figure 4.4 must be specified for each document in docIndex.xml.

Figure 4.4 Information in docIndex.xml

Element name	Designation	Description	Outcome space	Cardinality per document	Mandatory
dID	DocID	Unique ID identifying a document	ID of up to 12 digits	1	Yes
pID	Parent	ID of parent document	DocID	0-1	No
mID	MediaID	Name of the medium on which the document is stored	Cf. 4.B.1	1	Yes

dCf	SubFolder	Document collection folder in which the document is placed	Cf. 4.G	1	Yes
oFn	OrigFile-name	Specification of the file name (incl. extension) in the IT system	Free text	1	Yes
aFt	Archival-Format	Specification of file format in the information package	Extension, cf. 4.G.8	1	Yes
gmlXsd	Schema	For GML files, reference is made to the schema to be used for validation of the relevant GML file	Free text	0-1	No

4.C.7.a **researchIndex.xml** must contain specification of main tables and codes for missing values, cf. figure 4.5.

Figure 4.5 Information in researchIndex.xml

1. Metadata describing information packages					
	Element name	**Description**	**Cardinality per submission**	**Mandatory**	
1.a	mainTables	List of main tables	1	Yes	
2. Metadata describing the list of main tables					
	Element name	**Description**	**Cardinality per list of main tables**	**Mandatory**	
2.a	table	Main table	1-m	Yes	
3. Metadata describing the individual tables					
	Element name	**Description**	**Cardinality per table**	**Mandatory**	
3.a	tableID	The main table ID, as specified in the »folders« element of tableIndex.xml, cf. rules for naming tables in 5.A.	1	Yes	
3.b	source	Original format of the data or name of program from which the data originates	1	Yes	
3.c	specialNumeric	Boolean value that indicate if there are special codes for missing values	0-1	Yes, if codes for missing value are used	
3.d	columns	List of columns	0-1	Yes, if codes for missing value are used	
4. Metadata describing the column list					
	Element name	**Description**	**Cardinality per list of columns**	**Mandatory**	

4.a	column	Column	1-m	Yes
5. Metadata describing the individual columns				
	Element name	**Description**	**Cardinality per column**	**Mandatory**
5.a	columnID	Column ID as indicated in tableIndex.xml, cf. rules for naming columns in Figure 6.3, 4.b	1	Yes
5.b	missingValues	List of codes for missing values	1	Yes
6. Metadata describing the lists of codes for missing values				
	Element name	**Description**	**Cardinality per list of codes for missing values**	**Mandatory**
6.a	value	Code for missing value	1-m	Yes

4.D. *Tables* folder

4.D.1 The *Tables* folder must contain one folder for each table in the information package.

4.D.2.a The folder for a table must be named »table[consecutive number]«.

4.D.2.b The consecutive numbering starts with 1. Prefixed zeros are not allowed.

4.D.3 The folder for each table must contain a file: table[consecutive number].xml, cf., however, see 4.D.5

4.D.4 »table[consecutive number].xml« is an XML instance that contains data for table in question and its structure must be consistent with the XML instance »tableIndex.xml« so that data types in the XML instance are transformed from SQL: 1999 to XML data types in accordance with Schedule 5 Figure 5.1.

4.D.5 It is permitted in the folder for each table to include an XML schema that specifies only the structure of the table in question. The schema is named »table[consecutive number].xsd«. The schema must be in accordance with the XML instance »tableIndex.xml«, which specifies the structure of the entire relational database including all tables.

4.D.6 If a field in a table has a value NULL, the element of the XML instance (»table[consecutive number].xml«) must have the attribute xsi:nil="true" for that field, cf. the W3C standard for handling nil values in XML.

4.E. *ContextDocumentation* folder

4.E.1 The *ContextDocumentation* folder must contain one or more document collection folders with context documentation, cf. 6.B.

4.E.2 A document collection folder with context documentation can contain up to 10,000 document folders.

4.E.3 The document collection folders must be named »docCollection[consecutive number]«, starting with 1. The name must be unique within *ContextDocumentation*.

4.E.4 Each document in the context documentation must be given an ID of up to 12 digits. The document ID must be unique within *ContextDocumentation*.

4.E.5 A document folder must contain one document consisting of one or more files in the same format and be named with the document ID. Prefixed zeros are not allowed.

4.E.6 A document file (or files) must be named consecutively with a number, starting with 1, and the format extension, cf. 4.G.8.

4.F. *Schemas* folder

4.F.1 The *Schemas* folder must be divided into the subfolders *standard* and *localShared*.

4.F.2 The *standard* folder must contain schemas for the index files of the information package, cf. Schedule 8, and W3C standard for XML schema, cf. http://www.w3.org/2001/XMLSchema.xsd.

4.F.3 For the schemas fileIndex.xsd, archiveIndex.xsd, contextDocumentationIndex.xsd, tableIndex.xsd, docIndex.xsd and W3C's standard XML schema, the schemas provided by the National Archives must always be used. The schemas and their naming must not be changed in the information package.

4.F.4 The *localShared* folder must contain any GML schemas not placed with the relevant GML document, cf. 4.G.7.a.

4.G. *Documents* folder

4.G.1 The Documents folder must contain one or more document collection folders, up to a maximum of 10,000.

4.G.2 The document collection folders must be named »docCollection[consecutive number]«, starting with 1. The name must be unique within Documents.

4.G.3 A document collection folder can contain up to 10,000 document folders.

4.G.4 Each document in the information package must be given an ID of up to 12 digits. The document ID must be unique within *Documents*.

4.G.5 A document folder must contain one document consisting of one or more files in the same format and be named with the document ID. Prefixed zeros are not allowed.

4.G.6 A document file (or files) must be named consecutively with a number, starting with 1, and the format extension. Prefixed zeros are not allowed.

4.G.7 For GML files, the relevant schema must be stored in the same folder as the GML file and be named with a consecutive number followed by .xsd. See, however, 4.G.7.a. Prefixed zeros are not allowed.

4.G.7.a Alternatively, GML schemas may be stored in the schema folder named *localShared*, cf. 4.F. GML schemas in the *localShared* folder are named »localSchema[consecutive number]«, starting with 1.

4.G.8 Use of extensions

4.G.8.a Documents in the TIFF format must be given the extension tif.

4.G.8.b Documents in the MP3 format must be given the extension mp3.

4.G.8.c Documents in the MPEG-2 and MPEG-4 formats must be given the extension mpg.

4.G.8.d Documents in the JPEG-2000 format must be given the extension jp2.

4.G.8.e Documents in the GML format must be given the extension gml.

4.G.8.f Documents in the WAVE format must be given the extension wav.

4.G.9 The option of retrieving documents pertaining to the same case must be transferred to the information package in accordance with the instructions provided by the receiving archive.

Data content

The data content of the information package consists of table content in standardised data types and of any digital documents, audio, video and geodata in standardised formats.

5.A. Table content

5.A.1.a In accordance with the table structure, which in the XML instance »tableIndex.xml« is defined for each table, each table must be contained in an XML instance named »table[consecutive number].xml«.

5.A.1.b The consecutive numbering starts with 1. Prefixed zeros are not allowed.

5.A.2 The content of the individual fields must be cleared of any prefixed and suffixed white space characters.

5.B. Data types

5.B.1.a The standardised data types to be used for table content are specified in Figure 5.1. They represent an extract of the SQL:1999 standard represented as data types in W3C XML Schema Language 1.0.

5.B.1.b The W3C XML Schema Language 1.0 data type must be used.

The translation from SQL:1999 data types has been listed to show how the translation to data types in W3C XML Schema Language 1.0 must be carried out.

Figure 5.1 Permitted data types for table content

Data type	SQL:1999 (specified in tableIndex.xml)	XML data type (specified in table[consecutive number].xsd)
Text	CHARACTER, CHAR, CHARACTER VARYING, CHAR VARYING, VARCHAR, NATIONAL CHARACTER, NATIONAL CHAR, NCHAR NATIONAL CHARACTER VARYING, NATIONAL CHAR VARYING, NCHAR VARYING	string
Integer	INTEGER, INT, SMALLINT	integer
Decimal	NUMERIC, DECIMAL, DEC	decimal
	FLOAT	float
	REAL, DOUBLE PRECISION	double
Boolean	BOOLEAN	boolean
Hexadecimal	CHARACTER, CHAR, CHARACTER VARYING, CHAR VARYING, VARCHAR, NATIONAL CHARACTER, NATIONAL CHAR, NCHAR NATIONAL CHARACTER VARYING, NATIONAL CHAR VARYING,	hexBinary

Figure 5.1 Permitted data types for table content

Data type	SQL:1999 (specified in tableIndex.xml)	XML data type (specified in table[consecutive number].xsd)
	NCHAR VARYING	
Date	DATE	date
Time	TIME[WITH TIME ZONE]	time
Time stamp	TIMESTAMP[WITH TIME ZONE]	dateTime
Interval	INTERVAL	duration

5.B.2 The *string* data type must only contain non-marked-up text which must be interpreted directly.

5.B.3 According to W3C, the *boolean* data type may only have the values 1; 0 or *true*; *false*.

5.B.4 The *date, time* and *dateTime* data types may be used with or without *Time Zone*.

5.C. Conversion of table content to digital documents, audio, video or geodata

5.C.1 Table content must comply with the data types listed, cf. 5.B. This means that data content in table form from an IT system which is to be transferred to an information package and which cannot immediately meet this requirement must be converted as follows:

5.C.1.a to digital documents, audio, video or geodata, with the content being converted to the formats shown in 5.E - 5.G.

5.C.1.b to table content of the *string* data type, cf. 5.B, with all content other than that allowed for the data type being deleted.

5.C.2 The receiving archive gives instructions as to whether a given content should be handled according to 5.C.1.a or 5.C.1.b.

5.D. Text format

5.D.1 Unicode encoding

5.D.1.a Data in the index files and table content of the information package must be encoded as *well-formed* UTF-8, as provided in ISO/IEC 10646:2003 Annex D, and as described in *The Unicode Standard 5.1*, Chapter 3.

5.D.1.b The encoded characters must be valid Unicode scalar values. Surrogates/RC elements and Unicode *noncharacters* may not be used.

5.D.1.c No characters in *Private Use Area* may be encoded.

5.D.1.d The control code characters from and including #x00 to and including #x1F are not allowed with the exception of the control code characters tabulator (TAB) #x09, new line (LF) #x0A and return (CR) #x0D.

5.D.2 XML encoding

5.D.2.a According to the XML standard, the rules applying to the representation of the characters in question are listed in figure 5.2.

Figure 5.2 Representation of characters in the XML standard

Char.	Entity	Use
<	<	Mandatory
>	>	Recommended

&	&	Mandatory
'	'	Recommended
"	"	Recommended

5.D.2.b The characters from and including #x7F to and including #x9F must be specified with their decimal or hexadecimal character reference.

5.D.2.c CDATA sections <![CDATA]]> may not be used.

5.E. Digital documents

5.E.1 A digital document, (see, however, 5.F and 5.G), must be stored in one of the following formats:

– The graphical bitmap format TIFF, version 6.0 baseline.

– JPEG-2000 according to the ISO/IEC 15444-1:2004 standard. Information technology – JPEG 2000 image coding system – Part 1: Core coding system.

5.E.1.a Use of both formats is allowed within the same information package.

5.E.2 TIFF documents must be compressed according to the following compression rules:

5.E.2.a Black/white documents must be compressed using CCITT/TSS grp. 3, grp. 4, PackBit or LZW.

5.E.2.b Documents with grey tones or colours must be compressed using PackBit or LZW.

5.E.3 TIFF RGB documents may only use the following bit depths: 1, 2, 4, 8, 24 and 32.

5.E.3.a TIFF RGB documents may use a maximum of 3 colour channels with a maximum bit depth of 24 bits (8x8x8 bits), possibly supplemented with a maximum of an alpha channel (8 bits), so that the total bit depth of an image file cannot exceed 32 bits.

5.E.4 TIFF CMYK documents may only use the following bit depths: 1, 2, 4, 8, 32 and 40.

5.E.4.a TIFF CMYK documents may use a maximum of 4 colour channels with a maximum bit depth of 32 bits (8x8x8x8 bits), possibly supplemented with a maximum of one alpha channel (8 bits), so that the total bit depth of an image file cannot exceed 40 bits.

5.E.5 TIFF documents' use of XResolution and YResolution (TIFF Tag 282 and 283) must be such that the width-to-height ratio using these values (scaling) corresponds to the page dimensions of the original document.

5.F. Audio and video

5.F.1 Audio files must be stored in accordance with the MP3 DS/EN ISO/IEC 11172-3 standard.

5.F.2 The receiving archive may permit submission of audio files in the WAVE LPCM format as specified in *Multimedia Programming Interface and Data Specifications 1.0. IBM Corporation and Microsoft Corporation, August 1991*. However, bit depths must be multiples of 8.

5.F.3 Video files must be stored in accordance with one of the following standards:

– MPEG-2 DS/EN ISO/IEC 13818-2. Audio, if any, must be encoded as MP3 as specified in ISO/IEC 13818-3.

– MPEG-4 AVC DS/EN ISO/IEC 14496-10 (ITU-T H.264). Video must be encoded as specified in ISO/IEC 14496-10. Audio, if any, must be encoded as AAC as specified in ISO/IEC 14496-3. Video and audio must be packed in the MPEG-4 format as defined in ISO/IEC 14496-14.

5.G. Geodata

5.G.1 Data from geographical information systems and other geodata must be stored as GML files in accordance with specifications from OGC (Open Geospatial Consortium) in ISO 19136 GML version 3.1.1.

5.G.1.a Geodata files archived for cases in systems with documents can be converted to TIFF or JPEG-2000, cf. 5.E., and treated as a separate document in accordance with the rules in Schedule 4, 4.G. and Schedule 6, 6.C.4 - 6.C.6 in agreement with the receiving archive.

5.G.2 In the *localShared* subfolder of the *Schemas* folder, the schema collection from Open Geospatial Consortium to GML version 3.1.1 is placed in an unchanged form. The schemas can be downloaded as a joint zipped package at the National Archives' website.

5.G.3 GML files larger than 1 GB are subdivided into smaller units as specified by the receiving archive.

5.G.4 Each GML file, after possible subdivision, cf. 5.G.3, is treated as a separate document in accordance with the rules in 4.G.

5.G.4.a Each GML file is annotated in *docIndex.xml* and the element *gmlXsd* is annotated with a reference to the local XML schema that specifies the GML file's syntax rule, cf. Schedule 4, Section 4.C.6.a and 4.C.6.b.

5.G.4.b Columns in the information package's other table data containing information on *Dokumentidentifikation* and *Lagringsform*, as well as other relevant special information, must be identified via the *functionalDescription* element in *tableIndex.xml*, cf. Schedule 6, Section 6.C.4 - 6 .C.6.

5.G.5 Local XML schemas

5.G.5.a The required local XML schemas for documenting and validating the contents of the GML files must be included in the information package.

5.G.5.b Local XML schemas must comply with the syntax for schemas according to *W3C XML Schema Definition Language (XSD) 1.1 Part 1*.

5.G.5.c Local XML schemas must be encoded in the UTF-8 character set, which must be specified in the XML declaration. The use of characters is limited, cf. Schedule 5, Section 5.D.

5.G.5.d A local XML schema is placed in the same folder as the GML file that the schema validates, cf. the rules in 4.G.7. A local XML schema can also work globally for multiple GML files in an information package if the structure of the information package's GML files is contained in the schema's syntax rules, and in that case is placed in the *localShared* subfolder of the *Schemas* folder, cf. 4.G.7.a.

5.G.5.e In the root element of the local XML schema, the following namespaces must be marked: gml, xlink, xsi and the namespace of the submitting authority, cf. figure 5.3. The namespace »xxx« must be substituted by the URI of the submitting authority.

Figure 5.3 Namespace in the XML schema's root element

```
<schema
xmlns="http://www.w3.org/2001/XMLSchema"
xmlns:gml="http://www.opengis.net/gml"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xxx="http://www.xxx.dk/gml/xxx"
targetNamespace="http://www.xxx.dk/gml/xxx" elementFormDefault="qualified">
```

5.G.5.f The local XML schema must contain an import element for the namespace for GML version 3.1.1.

5.G.5.g The local XML schema must contain a general description of the content of the GML file documented using the XML documentation elements *annotation* and *documentation*. The scope and content of the documentation is agreed with the receiving archive.

5.G.5.h In a local XML schema, at least one local GML feature with associated GML geometry must always be annotated. The scope and content of features are agreed with the receiving archive.

5.G.5.i The GML features and geometries of the local XML schema are annotated as a sequence of features with the *sequence* element and are enclosed in the GML file by an XML collection element named »GEOMETRI«.

5.G.5.j The collection element *GEOMETRI* is defined in the local XML schema and inherits from the class *gml:_Feature* annotated using the XML attribute *substitutionGroup* and annotated with the elements *complexType* and *complexContent* and placed in the namespace of the submitting authority.

5.G.5.k The local XML schema must inherit the properties of the GML class *gml:AbstractFeature-Type*.

5.G.5.l The authority's local GML features and geometries must be described and documented using XML documentation elements *annotation* and *documentation*. The scope and content of the documentation are agreed with the receiving archive.

5.G.5.m The unique *gml:featureMember* in the GML file must establish a unique link to the information package's other table data, if one exists. The link between table data and the GML instance must be documented in the local XML schema and in *tableIndex.xml* for the local features and columns that make up the reference. Possible solutions for the establishment of references are agreed with the receiving archive.

5.G.6 GML files

5.G.6.a All GML files in an information package must comply with the applicable syntax rules for respectively XML version 1.0 and GML version 3.1.1.

5.G.6.b GML files must be encoded in the UTF-8 character set, which must be specified in the XML declaration. The use of characters is limited, cf. Schedule 5, Section 5.D.

5.G.6.c Geometries and features from the same map/table should as far as possible be kept together within the same GML file. Deviations from this are specifically agreed with the receiving archive.

5.G.6.d A *gml:FeatureCollection* element must be used as root element, cf. Figure 5.4.

5.G.6.e In the root element, the following namespace must be marked: gml, xlink, xsi and the namespace of the submitting authority, cf. Figure 5.4. The namespace »xxx« must be substituted by the *URI* of the submitting authority.

Figure 5.4 Namespace in the GML file's root element

```
<gml:FeatureCollection
xmlns:gml="http://www.opengis.net/gml"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xxx="http://www.xxx.dk/gml/xxx"
xsi:schemaLocation="http://www.xxx.dk/gml/xxx ./1.xsd">
```

5.G.6.f The root element must specify a local schema reference *xsi:schemaLocation* with a relative path for the location of the local schema, cf. Figures 5.4 and 5.G.5.d.

5.G.6.g All GML files must be marked with a coordinate reference system. »EPSG:25832« alias »ETRF89 / ETRS89 UTM zone 32N« shall be used as mark-up if the submitted geodata is located

in Jutland, Funen or Zealand and the waters around it, i.e. within the points specified in *lowerCorner* and *upperCorner*, cf. 5.G.6.i. If this is not the case, another coordinate reference system can be used upon agreement with the receiving archive.

5.G.6.h The coordinate reference system is specified using the GML element *gml:boundedBy* with the subelements *Envelope* and *lowerCorner* and *upperCorner*. This mark-up applies globally to the entire GML file and should not be set locally in the individual *gml:featureMember* elements.

5.G.6.i The geographies in the GML file must be located within the points *lowerCorner* (212481.60 6019669.40) and *upperCorner* (961440.75 6510422.51). Another or more precise indication of these extremes can be used upon agreement with the receiving archive.

5.G.6.j The dimension of the geography must be marked-up using the GML attribute *srsDimension* in the GML element "*Envelope*" together with the GML attribute *srsName*, which specifies the map's coordinate reference system.

5.G.6.k A GML file must contain at least one *gml:featureMember* element. The collection element *GEOMETRI* containing the local features and geometries of the GML file is placed in *gml:featureMember*. The element *GEOMETRI* must be included in the submitting authority's namespace.

5.H. Compression

5.H.1 An information package may not be compressed more than the compression provided for or following from the document formats required for use in the information package, cf. 5.E and 5.F.

5.I. Optimisation

5.I.1 In specific cases, the receiving archive may decide to reduce the space required for the documents, e.g. by instructing that optimal bit depth be used.

5.J. No deterioration

5.J.1 In connection with the creation of the information package, the quality of the documents, including audio and video, must not be reduced over and above what might result from the format required for the information package or the instructions otherwise issued by the receiving archive.

Metadata describing the information package

The metadata describing the information package consists of the following elements:
– Archive description file, cf. 6.A
– Context documentation, cf. 6.B
– Metadata describing the tables of the information package (table index), cf. 6.C
– Any SQL queries, cf. 6.D

6.A. Archive description file

6.A.1 All information packages must contain an archive description file including the information stated in Figure 6.1.

6.A.2 The archive description file must be named archiveIndex.xml and conform to the related schema, cf. Schedule 8.

6.A.3 The content of the archive description file is agreed following discussion between the authorities and the receiving archive.

Figure 6.1 Information in archiveIndex.xml

Element name	Designation	Description	Outcome space	Cardinality	Mandatory
archiveInformationPackageID	Information package ID	Unique ID allocated to the information package by the National Archives	Information package ID as defined in this Executive Order	1	Yes
archiveInformationPackageIDPrevious	Previous submission	Unique information package ID from previous submission from the same IT system or data source	Information package ID as defined in this Executive Order or previous executive orders	0-m	No
archivePeriodStart	Information package start date	Start date of the data submitted	Year, year-month or year-month-day	1	Yes
archivePeriodEnd	Information package end date	End date of the data submitted	Year, year-month or year-month-day	1	Yes

Figure 6.1 Information in archiveIndex.xml

Element name	Designation	Description	Outcome space	Car-dinal-ity	Manda-tory
documentPeriod-Start	Start date of documents	Start date of digital documents in infor-mation packages	Year, year-month or year-month-day	0-1	Yes, if the in-for-mation package contains digital docu-ments
documentPerio-dEnd	End date of documents	End date of digital documents in infor-mation packages	Year, year-month or year-month-day	0-1	Yes, if the in-for-mation package contains digital docu-ments
archiveInfor-mationPacket-Type	Final submis-sion	Specification of whether the infor-mation package is a final submission	Boolean value	1	Yes
creatorName	Archive creator	The authorities, insti-tutions, organisations or persons, which cre-ated the records. May also be organisational units within authori-ties or other contribu-tors to the records	Free text	1-m	Yes
creationPeriod-Start	Archive creator start date	For each archive crea-tor, the date when the archive creator started adding data	Year, year-month or year-month-day	1-m	Yes
creationPerio-dEnd	Archive creator end date	For each archive crea-tor, the date when the archive creator stopped adding data	Year, year-month or year-month-day	1-m	Yes
archiveType	Information package period type	Specification of whether it is a closing archive period/year	Boolean value	1	Yes

Figure 6.1 Information in archiveIndex.xml

Element name	Designation	Description	Outcome space	Cardinality	Mandatory
archiveType-ClosedFiles	Information package closed cases	Specification of whether it is a snapshot of metadata for IT systems with documents but only documents from closed cases	Boolean value	0-1	Yes, if the information package contains digital documents
systemName	System name	Official name of the IT system, writing out all abbreviations	Free text	1	Yes
alternativeName	Alternative system name	Alternative names of the IT system or data source	Free text	0-m	No
systemPurpose	System purpose	Description of the purpose of creating and operating the IT system or collected data	Free text	1	Yes
systemContent	System content	Description of the central population and central variables in the data	Free text	1	Yes
regionNum	Region numbers	Specification of whether region numbers are registered in the data in a systematic manner	Boolean value	1	Yes
komNum	Municipality numbers	Specification of whether municipality numbers are registered in the data in a systematic manner	Boolean value	1	Yes
cprNum	Civil registration numbers (CPR)	Specification of whether civil registration numbers are registered in the data in a systematic manner	Boolean value	1	Yes

Figure 6.1 Information in archiveIndex.xml

Element name	Designation	Description	Outcome space	Cardinality	Mandatory
cvrNum	CVR numbers	Specification of whether CVR numbers (number assigned by the Danish Central Business Register) are registered in the data in a systematic manner	Boolean value	1	Yes
matrikNum	Title numbers	Specification of whether title numbers are registered in the data in a systematic manner	Boolean value	1	Yes
bbrNum	BBR numbers	Specification of whether BBR numbers are registered in the data in a systematic manner	Boolean value	1	Yes
whoSygKod	WHO disease codes	Specification of whether WHO disease codes are registered in the Data in a systematic manner	Boolean value	1	Yes
sourceName	Data source	The official name of other data sources having delivered data to the data set (lookup, transfer, integration etc.)	Free text	0-m	No
userName	Data users	Other IT systems having used data from the IT system (lookup, transfer, integration etc.)	Free text	0-m	No
predecessor-Name	Predecessor systems	Systems previously handling the same function	Free text	0-m	No

Figure 6.1 Information in archiveIndex.xml

Element name	Designation	Description	Outcome space	Cardinality	Mandatory
formVersion	FORM version	Indication of the FORM version (the joint public business reference model) from which the categorisation below was retrieved	Valid FORM version numbers	0-1	Mandatory for IT systems from public authorities
formClass	FORM classification	Categorisation of the information package according to FORM	Any valid FORM reference according to the FORM version specified	0-m	Mandatory for IT systems from public authorities
formClassText	FORM classification plain text	FORM classification in plain text	Free text	0-m	Mandatory for IT systems from public authorities
containsDigitalDocuments	Contains digital documents	Specification of whether the information package contains digital documents in addition to the context documentation	Boolean value	1	Yes
containsGeodata	Contains GML files	Specification of whether the information package contains GML files	Boolean value	1	Yes
containsResearchData	Contains research data	Specification of whether the information package contains research data	Boolean value	1	Yes
researchSIP	Research data submitted according to Schedule 9	Specification of whether the data is submitted according to Schedule 9	Boolean value	1	Yes

Figure 6.1 Information in archiveIndex.xml

Element name	Designation	Description	Outcome space	Cardinality	Mandatory
documentsDisposal	Discarding	Specification of whether documents have been systematically discarded prior to or in connection with the production of the information package.	Boolean value	1	Yes
searchRelatedOt herRecords	Search facility for other cases or documents	Specification of whether the information package is a required search facility for paper cases/documents or cases/documents in another IT system	Boolean value	1	Yes
relatedRecordsName	Addressed records	Reference to the records for which the information package is a search facility	Free text	0-m	Yes, if the answer to the previous question was yes
systemFileConcept	Existence of a case expression in the IT system	Indication that the IT system has a case expression, i.e., a registered relationship between documents pertaining to the same case	Boolean value. Can only be *true* if the IT system contains digital documents	1	Yes
multipleDataCollection	SOA architecture	Specification of whether the IT system is composed of data and any documents from different IT systems in a service-oriented architecture	Boolean value	1	Yes

Figure 6.1 Information in archiveIndex.xml

Element name	Designation	Description	Outcome space	Car-dinal-ity	Manda-tory
personalDa-taRestrictedInfo	Personal data	Specification of whether the infor-mation package con-tains sensitive per-sonal data pursuant to the General Data Pro-tection Regulation	Boolean value	1	Yes
otherAc-cessTypeRe-strictions	Information ex-tending the time limit	Specification of whether the infor-mation package con-tains other infor-mation that may jus-tify a longer access period	Boolean value	1	Yes
archiveApproval	Approval ar-chive	Indication of which public archives are approving the infor-mation package	Identification of the relevant archives (2-4 characters)	1	Yes
archiveRestric-tions	Access limita-tions	Indication of any con-ditions governing ac-cess to the material. The element can be used as instructed by the receiving archive	Free text	0-1	No

6.B. Context documentation

6.B.1 All information packages must contain documents documenting the IT system's administrative function as well as its structure and functionality or the research data's source and structure.

6.B.2 After consulting the submitting authorities, the receiving archive determines which documents are to be submitted, including which points in Figure 6.2 are not relevant to the submission in ques-tion.

6.B.3.a The documents are placed in one or more of the categories shown in Figure 6.2.

6.B.3.b Metadata describing the categorisation is recorded in the index file contextDocumentationIndex, cf. 4.C.4.a.

6.B.4 The documents must be stored in one of the permitted document formats for the information package, cf. 5.E - 5.F.

Figure 6.2 Document categories in contextDocumentationIndex.xml

1. Documentation regarding the administrative use of the IT system		
	Element name	**Description**
1.a	systemPurpose	Purpose of the IT system
1.b	systemRegulations	Legal basis for the IT system
1.c	systemContent	Content, population and special expressions of the IT system
1.d	systemAdministrativeFunctions	Administrative functions in the IT system
1.e	systemPresentationStructure	Presentation structure of the IT system
1.f	systemDataProvision	Provision of data
1.g	systemDataTransfer	Transfer of data
1.h	systemPreviousSubsequentFunctions	Data and functions identical in the predecessor and successor systems
1.i	systemAgencyQualityControl	The authorities' own quality control
1.j	systemPublication	Publication of and on data
1.k	systemInformationOther	Other
1.l	systemTaxonomy	Registration systematics
1.m	systemInstruction	Instructions for using the system
2. Documentation regarding the technical design, operation and development of the IT system		
	Element name	**Description**
2.a	operationalSystemInformation	Structure of the operational version
2.b	operationalSystemConvertedInformation	Conversion by the authorities
2.c	operationalSystemSOA	Documentation of the composition of data and any documents from several different IT systems in a service-oriented architecture
2.d	operationalSystemInformationOther	Other
3. Documentation regarding the archive creator's data submission		
	Element name	**Description**
3.a	archivalProvisions	The archives' provisions, including submission provisions
3.b	archivalTransformationInformation	Documentation of the conversion from operational version to information package
3.c	archivalInformationOther	Other
4. Documentation regarding the archives' receipt of data (to be filled in by the receiving archive)		
	Element name	**Description**
4.a	archivistNotes	The archivist's notes
4.b	archivalTestNotes	Test notes
4.c	archivalInformationOther	Other

	Element name	**Description**
5. Documentation regarding the archives' preservation of the information package (to be filled in by the receiving archive)
	Element name	**Description**
5.a	archivalMigrationInformation	Conversion by the archives
5.b	archivalInformationOther	Other
6. Other documentation
	Element name	**Description**
6.a	informationOther	Other
7. Documentation of research data
	Element name	**Description**
7.a	researchProjectDescription	Project description applicable to the submitted data
7.b	researchQuestionnaire	Questionnaire, interview guide and/or registration form used to collect and analyse the submitted data
7.c	researchProtocol	Protocols and method reports
7.d	researchPublication	Publications on the basis of the submitted data
7.e	researchInformationOther	Other

6.C. Metadata describing the tables of the information package

6.C.1 An information package must contain documentation of the tables and relations of the information package (table index). The table index must contain the information shown in Figure 6.3 below.

Figure 6.3 Information in tableIndex.xml

	Element name	**Description**	**Cardinality per database**	**Mandatory**
1. Database information
1.a	version	Format version, always '1.0'	1	Yes
1.b	dbName	Short name of the database	0-1	Yes, if the information is available in the database
1.c	databaseProduct	Name and version of the database product from which the data were filed	0-1	Yes, if the information is available in the database
2. Metadata describing tables and views
	Element name	**Description**	**Cardinality per information package**	**Mandatory**
2.a	tables	List of tables in the database	1	Yes
2.b	views	List of views in the database	0-1	Yes, if views are included in the information package
3. Metadata describing the individual tables

	Element name	Description	Cardinality per table	Mandatory
3.a	name	Table name. Must not start with a number.	1	Yes
3.b	folder	Name of the folder containing the table and the related schema	1	Yes
3.c	description	Description of the content and meaning of the table	1	Yes
3.d	columns	List of columns in the table	1-m	Yes
3.e	primaryKey	The primary key of the table	1	Yes
3.f	foreignKey	Foreign key in the table	0-m	Yes, if there are foreign keys
3.g	rows	The number of rows in the table	1	Yes

4. Column information

	Element name	Description	Cardinality per column	Mandatory
4.a	name	Column name	1	Yes
4.b	columnID	Unique column identification. Starts with the letter 'c' followed by the number of the column in the table.	1	Yes
4.c	type	SQL:1999 data type	1	Yes
4.d	typeOriginal	Original data type	0-1	Yes, if the information is available in the database
4.e	defaultValue	Default value of fields in the column	0-1	Yes, if the information is available in the database
4.f	nullable	Specification of whether fields in the column can be 'NULL'. Boolean value.	1	Yes
4.g	description	Description of column content	1	Yes
4.h	functionalDescription	The column function according to Figure 6.4, Figure 6.5 and Figure 6.6	0-m	Yes, if the column has one of the functions specified

5. Primary key information

	Element name	Description	Cardinality per key in the table	Mandatory

5.a	name		Name of primary key. The name must be unique within the information package and be in accordance with the rules for *constraint names* in the SQL:1999 standard	1	Yes
5.b	column		Original name of columns in the primary key	1-m	Yes

6. Foreign key information

	Element name		Description	Cardinality per key/reference	Mandatory
6.a	name		Name of foreign key. The name must be unique within the information package and be in accordance with the rules for constraint names in the SQL:1999 standard	1	Yes
6.b	referencedTable		The table to which the foreign key refers	1	Yes
6.c	reference		Collection element that pairs the referring and referenced columns together. For compound foreign keys, the element is repeated.	1-m	Yes
6.d		column	Name of the column that is referenced from	1	Yes
6.e		referenced	Name of the column that is referenced to	1	Yes

7. Metadata describing views and queries

	Element name	Description	Cardinality per query or view	Mandatory
7.a	name	View name	1	Yes
7.b	queryOriginal	Original SQL query defining view or SQL query created specifically for the information package	1	Yes
7.c	description	Content and meaning of view or SQL query	0-1	Yes, if they can be extracted from the database or if it concerns queries created specifically for the information package

6.C.2 System views are not included.

6.C.3 The receiving archive may direct that the most important views must be described, cf. Figure 6.3, 7.c.

6.C.4 Special information must be marked up for IT systems with recording of information on documents.

6.C.5 For information packages of IT systems mentioned in 6.C.4, columns containing special information, cf. Figure 6.4 - Figure 6.6, must be identified using the element functionalDescription, cf. Figure 6.3, 4.h.

6.C.6 The special information specified in Figure 6.6 must be marked up to the extent that it has been recorded in the IT system. If the information in Figure 6.6 is not recorded in the IT system, any alternative mark-ups must be agreed in connection with submission to ensure identification of documents pertaining to the same case, cf. the applicable provisions.

Figure 6.4 Functional Descriptions in IT systems shared by several authorities

Mark-up of columns containing fixed information in information packages of IT systems with recording of documents shared by several authorities	
Information	**Description**
Authority identification	The column(s) in the information package containing information on the authorities which have recorded the case or the document

Figure 6.5 Functional Descriptions in IT systems with documents

Mark-up of columns containing fixed information in information packages of IT systems with documents		
Information	**Description**	**Outcome space in table data**
Document identification	Used to identify the column(s) in the information package describing the unique identifier of the documents	ID of up to 12 digits, which is found among the named document folders and is unique within the *Documents* folder, cf. 4.G.4
Storage form	Used to identify the columns(s) in the information package describing whether the document has been stored digitally, on paper or not associated with a file	Fully or partly digitally = 1, paper = 2, not relevant = 3
Submitted	Used for submission of snapshots, etc. to identify the column(s) in the information package describing whether the document has already been submitted in a previous information package	Previously submitted = 1, not previously submitted = 2

Figure 6.6 Functional Descriptions in IT systems with document registration

Mark-up of columns containing information in information packages of IT systems with document registration

Information	Description
Case identification	The column(s) in the information package describing the unique identifier of the cases
Case title	The column(s) in the information package containing the case titles
Document title	The column(s) in the information package containing the document titles/descriptions
Document date	The column(s) in the information package containing information on the dates of submission and receipt of the documents
Sender/recipient	The column(s) in the information package containing information on the sender or recipient of the documents
Digital signature	The column(s) in the information package containing information extracted from a digital signature
FORM	The column(s) in the information package containing a reference to FORM (the joint public business reference model)
Discarding	The column(s) in the information package containing information on preservation and discarding

6.D. SQL queries

6.D.1 The receiving archive may decide that a number of SQL queries must be defined for an information package for documentation of certain relations in the information package.

6.D.2 The SQL queries must be created in accordance with the SQL:1999 (core) standard.

6.D.3 The queries must be placed in »Metadata describing views and queries« in the table index, cf. Figure 6.3, 7, and be named according to the authorities' own choice; however, such that the name of the relevant queries starts with »AV«.

Submission medium

7.A.1 Information packages may be submitted on a CD-R, DVD-R, BD-R or USB medium.

7.A.2 The number of CD-Rs, DVD-Rs or BD-Rs in one submission may not exceed ten, unless otherwise agreed between the submitting authority and the receiving archive.

7.A.3 When submitting information packages over 4 GB on USB medium, use USB 3.0 (USB 3.1 Gen1) or faster USB standards.

7.A.4 The submitting authority and the receiving archive may make an agreement on submission on other media or other methods of transporting data.

7.A.5 Media with sensitive personal data are encrypted as instructed by the receiving archive.

Schemas

Complete schemas for use for creation of an information package are available on the website of the National Archives.

Information package for certain types of research data

9.A. Submission of research data

9.A.1 The rules set out in this Schedule apply only to data created in connection with research using scientific methods and that are created or processed in statistical programs or the like.

9.A.2 The submitted information package is converted to an information package according to the rules in Schedules 1-8, of the receiving archive.

9.B. Information package's folder structure

9.B.1 In the root of the file system on the submission medium, cf. Schedule 7, a folder with the name of the information package must be placed. The name of the information package consists of the prefix »FD.« and a unique serial number for the information package.

9.B.2 The serial number of the information package is provided by the National Archives.

9.B.3 The contents of the information package are distributed in folders as specified in Figure 9.2.

9.B.4 The folders must be named as specified in Figure 9.2.

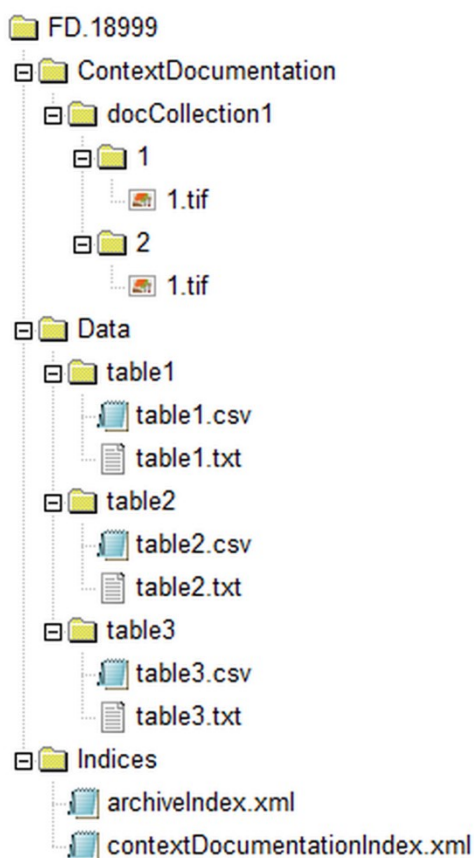**Figure 9.1 Graphical overview of elements and structure in an information package**

Figure 9.2 Information package folders

Folder name	Content
ContextDocumentation	Context documentation, cf. 9.D.
Data	Data file and metadata file, cf. 9.E.
Indices	Index files, cf. 9.C.

9.C. Indices folder

9.C.1 The *Indices* folder must contain the following index files with information about the submission package and its contents:

– archiveIndex.xml
– contextDocumentationIndex.xml

9.C.2 The index files must comply with their associated schema, cf. Schedule 8.

9.C.3 The archive description file archiveIndex.xml must comply with the rules in Schedule 6, Section 6.A.

9.C.4 The context documentation file contextDocumentationIndex.xml must comply with the rules in Schedule 4, Sections 4.C.4.a and 4.C.4.b.

9.D. ContextDocumentation folder

9.D.1 The *ContextDocumentation* folder must contain context documentation, cf. 4.E.

9.E. Data folder

9.E.1 An information package must contain one or more data sets. Each data set must consist of a data file and an associated metadata file.

9.E.2 The data file and metadata file are placed in the *Data* folder in a subfolder named according to Schedule 4, Sections 4.D.2.a and 4.D.2.b.

9.E.2.a The data file is named with the subfolder's name followed by the extension ».csv«.

9.E.2.b The metadata file is named with the subfolder's name followed by the extension ».txt«.

9.F. Text format

9.F.1 The character set for the data file and the metadata file shall be encoded as UTF-8, as specified in ISO/IEC 10646:2003 Annex D and as described in *The Unicode Standard 5.1*, Chapter 3, and shall in addition comply with the provisions of Schedule 5, Sections 5.D.1.b - 5.D.1.d.

9.G. Data file

9.G.1 The data file is submitted as a semicolon separated text file according to RCF 4180 Common Format and MIME Type for CSV Files that adheres to the syntax described as EBNF in Figure 9.12.

9.G.1.a The first line of the data file must always specify all variable names, listed in the same order as in the metadata file.

9.G.1.b If a semicolon is included in a value for a variable, the whole value must be enclosed in double quotation marks »"« (U+0022). If a double quotation mark is included in a value for a variable, the double quotation mark must be preceded by a double quotation mark and the entire value is enclosed in double quotation marks.

9.G.1.c One of the following line break methods must be used as a line separator: »CR+LF« (U+000D) and (U+000A) or »CR« (U+000D) or »LF« (U+000A). Content in variables in the data file must not contain line breaks.

9.G.2 One of the following three types can be a missing value: Missing value (empty) (cf. 9.G.2.a), special code for missing value (cf. 9.G.2.d) or user-defined code for missing value (cf. 9.I.6).

9.G.2.a Missing values (empty) in the data file must either be represented as no value »« or a space » « (U+0020).

9.G.2.b In a data file, only special codes or user-defined codes may be used for missing values.

9.G.2.c Special codes for missing values may only be used for categorical and numerical variables.

9.G.2.d Special codes for missing values may only be used for integers and decimal numbers and must be specified either as a value from A-Z or .a-.z.

9.G.3 The contents of the individual variables must be cleared of any prefixed and suffixed blank characters.

9.H. Data types

9.H.1 The six standardised data types to be used in the data file are shown in Figure 9.3.

9.H.2 Data format notations for used data formats must be specified in the metadata file, cf. Figure 9.3. The data format notations are case sensitive.

9.H.2.a Values for the letters »w« and »d« in Figure 9.3 must be indicated consistently. »w« indicates the total width of the data type. »d« indicates decimals or fractional second precision.

Figure 9.3 Permitted data types

Data type	Data format in data file	Data format notation in metadata file	
Text	UTF-8 character set, cf. Section 9.F.1.	xml	string
		Stata	%ws
		SAS	$w.
		SPSS	aw
Numeric integer	Representation of an integer with or without a sign according to DS/ISO 6093:1985 (NR1) standard, cf. syntax rule in Figure 9.6.	xml	int
		Stata	%w.0f
		SAS	fw.
		SPSS	fw
Decimal number	Representation of a decimal number with or without a sign according to DS/ISO 6093:1985 (NR2) standard, cf. syntax rule in Figure 9.7.	xml	decimal
		Stata	%w.df or %w.dg
		SAS	fw.d
		SPSS	fw.d
Date	Specification of calendar date according to DS/ISO 8601:1993 extended format: CCYY-MM-DD Alternatively, the following format can be used, cf. syntax rule in Figure 9.8:	xml	date
		Stata	%tdCCYY-NN-DD
		SAS	yymmdd10.
		SPSS	sdate10

	CCYY/MM/DD		
Time	Specification of time according to DS/ISO 8601:1993 extended format, cf. syntax rule in Figure 9.9: hh:mm:ss	xml	time
		Stata	%tcHH:MM:SS
		SAS	time. or time8.
		SPSS	time8
Time stamp	Specification of date and time according to DS/ISO 8601:1993 extended format: CCYY-MM-DDThh:mm:ss.sss or CCYY-MM-DD hh:mm:ss.sss Alternatively, the following format can be used, cf. syntax rule in Figure 9.10: dd-mmm-yyyy hh:mm:ss Fractional seconds are optional and are allowed with a precision of up to 6 digits. Time zone in time entries is not allowed, cf. syntax rule in Figure 9.10.	xml	datetime
		Stata	%tcCCYY-NN-DD!THH:MM:SS %tcCCYY-NN-DD!THH:MM:SS.sss
		SAS	e8601dt19. e8601dtw.d
		SPSS	datetime20 or ymdhms19 ymdhmsw.d

9.I. Metadata files

9.I.1 Metadata from a data file is submitted as a structured text file that complies with the syntax described as EBNF in Figure 9.11.

9.I.1.a The metadata file is designed as indicated in Figure 9.4, where eight tags divide the metadata into specific categories.

9.I.1.b Each tag must appear once in the metadata file. The prevalence of the content of the label is shown in the columns »Cardinality« and »Mandatory« in Figure 9.4.

9.I.1.c Tag names are reserved words and must not be used to name other contents of the metadata file.

Figure 9.4 The structure of the metadata file

Tag	Description of the tag contents	Outcome space	Cardinality	Mandatory
SYSTEMNAVN	Name of the program from which data is extracted or the original format of the data.	SPSS SAS Stata Excel or free text	1	Yes
DATAFILNAVN	The name of the data file as it is referred to in the context of use. Data file names must	ISO/IEC 9075:1999 -	1	Yes

	comply with the requirements of Section 9.I.2.	Database Language SQL (SQL-99)		
DATAFILBESKRIV-ELSE	Description of the contents of the data file.	Free text	1	Yes
NØGLEVARIABEL	The unique key variable of the data file is specified with the names of the variable(s) of the key variable.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	0-1	Yes, if a unique key variable exists
REFERENCE	References to other data files in the information package as specified in Section 9.I.3.	ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	0-m	Yes, if reference to other data file in the information package exists
VARIABEL	Variables in the data file. A variable is indicated as a set consisting of a variable name followed by a notation for the variable's data format, cf. Figure 9.3, and a possible code list reference, cf. 9.I.5.f. The variable must comply with the requirements under Section 9.I.4.	Naming of variables must comply with ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	1-m	Yes
VARIA-BELBESKRIVELSE	Descriptions of the contents of the variables.	Free text	1-m	Yes
KODELISTE	Code lists specified with the code list's name followed by several sets consisting of code and code description. Code lists must comply with the requirements in Section 9.I.5.	Naming of code lists must comply with ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	0-m	Yes, if code lists exist
BRUGERKODE	User-defined codes for missing values indicated by the name of the variable where the codes are used, followed by the user-defined codes. User-defined codes for missing values must comply with	Naming of variables must comply with ISO/IEC 9075:1999 - Database Language SQL (SQL-99)	0-m	Yes, if there are user-defined codes for missing values

	the requirements under Section 9.I.6.			

9.I.2 Data file names must be unique within the same information package.

9.I.3 Reference

9.I.3.a A reference to another data file in the information package is specified by the name of the data file referenced (foreign data file), followed by the variable name of the foreign data file's key variable (foreign variable) followed by the variable name of the variable (reference variable) in the data file, which refers to the foreign variable in the foreign data file.

9.I.3.b There must be complete consistency between the data type and length of the key variables included in the reference.

9.I.4 Variable names must be unique within the same metadata file.

9.I.5 Code list

9.I.5.a Only categorical and numerical variables may have references to a code list.

9.I.5.b Only code lists for variables containing data of the data type numeric integer, decimal number or text are allowed, cf. data types in Figure 9.3.

9.I.5.c All codes in the data must be defined in a code list and explained with a code description.

9.I.5.d If codes are defined as an interval where not all codes have code descriptions, this must be documented in the variable description.

9.I.5.e Codes must be unique within the same code list.

9.I.5.f The reference between a code list and the variable that refers to the code list is provided under the tag VARIABLE, cf. Figure 9.4, as described in 9.I.5.g and 9.I.5.h.

9.I.5.g If the data type is a numeric integer or a decimal number, the reference is specified with the selected code list name and a trailing full-stop ».« (U+002E).

9.I.5.h If the data type is text, the reference with the selected code list name is indicated by a leading dollar sign »$« (U + 0024) and a trailing full-stop ».« (U + 002E).

9.I.6 User code

9.I.6.a A user-defined code for a missing value is only allowed for categorical and numerical variables.

9.I.6.b A user-defined code for a missing value indicated in the metadata file must always be specified in the code list to which it belongs.

Figure 9.5 EBNF explanation of characters

::= defined as
() indicates a grouping that must be performed together
[] indicates the possibility of (0 or 1)
{} indicates a possible repeat (0 or more)
| indicates a choice (either or)
… indicates a continuous interval
!! indicates a descriptive explanation in plain English
"" encloses actual values to be written without interpretation

Figure 9.6 EBNF for integer type DS/ISO 6093:1985 (NR1) standard

Nonterminal	Terminal
INT ::=	NR1
NR1 ::=	[SIGN] DIGIT {DIGIT}

Nonterminal	Terminal
SIGN ::=	”+” | ”-” *!! It is optional to use signs for positive integers*
DIGIT ::=	”0” | ”1” | … | ”9”

Figure 9.7 EBNF for decimal number type DS/ISO 6093:1985 (NR2) standard

Nonterminal	Terminal
DECIMAL ::=	NR2
NR2 ::=	[SIGN] DIGIT {DIGIT} | [SIGN] DIGIT{DIGIT} DECIMAL-MARK DIGIT {DIGIT} | DECIMAL-MARK DIGIT{DIGIT}
SIGN ::=	”+” | ”-” *!! It is optional to use signs for positive decimal numbers !! The use of negative sign before the value 0 is not permitted*
DIGIT ::=	”0” | ”1” | … | ”9”
DECIMAL-MARK ::=	”.” | ”,”

Figure 9.8 EBNF for types of dates

Nonterminal	Terminal
DATE ::=	ISO-8601-DATE | ALTERNATIVE-DATE
ISO-8601-DATE ::=	CC YY HYPHEN-MINUS MM HYPHEN-MINUS DD
CC ::=	DIGIT DIGIT *!! millennium is indicated by two integers*
HYPHEN-MINUS ::=	”-” *!! hyphen-minus (U+002D)*
YY ::=	DIGIT DIGIT *!! year is indicated by two integers*
MM ::=	DIGIT DIGIT *!! month is indicated by two integers*
DD ::=	DIGIT DIGIT *!! day is indicated by two integers*
DIGIT ::=	”0” | ”1” | … | ”9”
ALTERNATIVE-DATE ::=	CC YY SLASH MM SLASH DD
SLASH ::=	”/” *!! slash U+002F*

Figure 9.9 EBNF for types of times

Nonterminal	Terminal
TIME ::=	HOUR COLON MINUTE COLON SECOND
HOUR ::=	DIGIT [DIGIT] *!! hour indicated by one or two integers !! outcome space is 0-23*
COLON ::=	”:” *!! colon (U+003A)*
MINUTE ::=	DIGIT DIGIT *!! minute indicated by two integers !! outcome space is 00-59*
SECOND ::=	DIGIT DIGIT *!! second indicated by two integers !! outcome space is 00-59*
DIGIT ::=	”0” | ”1” | … | ”9”

Figure 9.10 EBNF for datetime types

Nonterminal	Terminal
DATETIME ::=	ISO-8601-DATETIME | ALTERNATIVE-DATETIME | IBM-DATETIME
ISO-8601-DATETIME ::=	CC YY HYPHEN-MINUS MM HYPHEN-MINUS DD SEPARATOR HOUR COLUMN MINUTE COLON SECOND [FULL-STOP MILISEC-ONDS]
CC ::=	DIGIT DIGIT *!! millennium is indicated by two integers*
DIGIT ::=	”0” | ”1” | … | ”9”
HYPHEN-MINUS ::=	”-” *!! hyphen-minus (U+002D)*
YY ::=	DIGIT DIGIT *!! year is indicated by two integers*
MM ::=	DIGIT DIGIT !! month is indicated by two integers
DD ::=	DIGIT DIGIT !! day is indicated by two integers
SEPARATOR ::=	”T” | SPACE
SPACE ::=	” ” *!! space character (U+0020)*

Nonterminal	Terminal
HOUR ::=	DIGIT DIGIT *!! hour indicated by one or two integers !! outcome space is 0-23*
COLON ::=	":" *!! colon (U+003A)*
MINUTE ::=	DIGIT DIGIT *!! minute indicated by two integers !! outcome space is 00-59*
SECOND ::=	DIGIT DIGIT *!! second indicated by two integers !! outcome space is 00-59*
FULL-STOP ::=	"." *!! full-stop (U+002E)*
MILISECONDS ::=	DIGIT {DIGIT} *!! miliseconds indicated by up to maximum six digits*
IBM-DATETIME ::=	DD HYPHEN-MINUS MONTH HYPHEN-MINUS YEAR SPACE HOUR COLON MINUTE COLON SECOND
MONTH ::=	"Jan" | "Feb" | "Mar" | "Apr" | "May" | "Jun" | "Jul" | "Aug" | "Sep" | "Oct" | "Nov" | "Dec" *!! English abbreviations must be used for month names and they are not case sensitive.*
YEAR ::=	DIGIT DIGIT DIGIT DIGIT *!! year indicated by four integers*

Figure 9.11 EBNF Syntax rules for metadata file

Nonterminal	Terminal
METADATAFILE ::=	SYSTEMNAME DATAFILENAME DATAFILEDESCRIPTION KEYVARIABLE REFERENCE VARIABLE VARIABLEDESCRIPTION CODELISTS USERCODE
SYSTEMNAME::=	"SYSTEMNAVN" LINEBREAK "SAS" | "Stata" | "SPSS" | "Excel" | FREETEXT LINEBREAK LINEBREAK {LINEBREAK}
DATAFILENAME ::=	"DATAFILNAVN" LINEBREAK TITLE LINEBREAK LINEBREAK {LINEBREAK}
LINEBREAK ::=	(CR LF) | CR | LF
CR ::=	*!! carriage return (U+000D)*
LF ::=	*!! line feed (U+000A)*
TITLE ::=	(LETTER {LETTER | DIGIT}) | (QUOTATIONMARK LETTER {LETTER | DIGIT} QUOTATIONMARK) *!! A TITLE must never begin with a number, but can be a mix of letters and numbers of maximum 128 characters, and if the title is a reserved word in SQL: 1999, the title must be enclosed by quotation marks, cf. ISO/IEC 9075:1999 - Database Language SQL (SQL:1999)*
QUOTATIONMARK ::=	""" *!! quotation mark (U+0022)*
LETTER ::=	"A" | "B" | … | "Z" | "a" | "b" | … | "z" | "_" *!! As well as other national characters that are allowed in constraint names under the standard SQL:1999 !! underscore (U+005F)*
DIGIT ::=	"0" | "1" | … | "9"
DATAFILEDESCRIPTION ::=	"DATAFILBESKRIVELSE" LINEBREAK FREETEXT LINEBREAK LINEBREAK {LINEBREAK}
FREETEXT ::=	LETTER | DIGIT | OTHERCHARACTERS {LETTER | DIGIT | OTHERCHARACTERS} *!! FREETEXT can be a mix of letters, numbers and other characters as long as it makes semantic sense*
OTHERCHARACTERS ::=	*!! all permitted characters in UTF-8 which are not letters and numbers, cf. Section 9.F*
KEYVARIABLE ::=	"NØGLEVARIABEL" LINEBREAK {VARIABLENAME SPACE} LINEBREAK LINEBREAK {LINEBREAK}
VARIABLENAME ::=	TITLE
SPACE ::=	*!! space character (U+0020)*

Nonterminal	Terminal
REFERENCE ::=	”REFERENCE” LINEBREAK {FOREIGNDATAFILE SPACE FOR-EIGNVARIABLE SPACE REFERENCEVARIABLE LINEBREAK} LINEBREAK
FOREIGNDATAFILE ::=	TITLE *!! name of the data file referenced*
FOREIGNVARIABLE ::=	APOSTROPHE VARIABLENAME {SPACE VARIABLENAME} APOSTROPHE *!! name of the key variable in the referenced data file*
REFERENCEVARIABLE ::=	APOSTROPHE VARIABLENAME {SPACE VARIABLENAME} APOSTROPHE *!! name of the variable in the referenced data file, which refers to the key variable in the referenced data file*
APOSTROPHE ::=	”’” *!! apostrophe (U+0027)*
VARIABLE ::=	”VARIABEL” LINEBREAK VARIABLESET {VARIABLESET} LINE-BREAK {LINEBREAK}
VARIABLESET ::=	VARIABLENAME SPACE DATAFORMATNOTATION SPACE [CODELISTREFERENCE] LINEBREAK
CODELISTREFERENCE ::=	[DOLLAR] TITLE FULL-STOP
DATAFORMATNOTATION ::=	INTEGERNOTATION | DECIMALNOTATION | DATENOTATION | DATETIMENOTATION | TIMENOTATION | TEXTNOTATION
INTEGERNOTATION ::=	*!! see Figure 9.3 for data format notations for numeric integers*
DECIMALNOTATION ::=	*!! see Figure 9.3 for data format notations for decimal numbers*
DATENOTATION ::=	*!! see Figure 9.3 for data format notations for date types*
DATETIMENOTATION ::=	*!! see Figure 9.3 for data format notations for time stamp*
TIMENOTATION ::=	*!! see Figure 9.3 for data format notations for time*
TEXTNOTATION ::=	*!! see Figure 9.3 for data format notations for text*
DOLLAR ::=	”$” *!! dollar sign (U+0024)*
FULL-STOP ::=	”.” *!! full-stop (U+002E)*
VARIABLEDESCRIPTION ::=	” VARIABELBESKRIVELSE” LINEBREAK DESCRIPTION {DE-SCRIPTION} LINEBREAK {LINEBREAK}
DESCRIPTION ::=	VARIABLENAME SPACE APOSTROPHE FREETEXT APOSTROPHE LINEBREAK
CODELIST ::=	”KODELISTE” LINEBREAK {CODESET} LINEBREAK {LINE-BREAK}
CODESET ::=	CODELISTNAME LINEBREAK {CODE SPACE CODECESCRIPTION LINEBREAK}
CODELISTNAME ::=	TITLE
CODE ::=	APOSTROPHE INTEGER | DECIMAL | FREETEXT APOSTROPHE
CODEDESCRIPTION ::=	APOSTROPHE FREETEXT APOSTROPHE
INTEGER ::=	*!! see Figure 9.6 EBNF for integer type*
DECIMAL ::=	*!! see Figure 9.6 EBNF for decimal number type*
USERCODE ::=	”BRUGERKODE” LINEBREAK {VALUESET} LINEBREAK {LINE-BREAK}
VALUESET ::=	VARIABLENAME SPACE VALUE {SPACE VALUE} LINEBREAK
VALUE ::=	APOSTROPHE INTEGER | DECIMAL | FREETEXT APOSTROPHE

Figure 9.12 EBNF Syntax rules for data file

Nonterminal	Terminal
DATAFILE ::=	DATACONTENT
DATACONTENT ::=	HEADING LINEBREAK ROW {ROW}
HEADING ::=	VARIABLENAME {SEPARATORCHARACTER VARIABLENAME}*!! A separator character is not used after the last variable name in the header line*
VARIABLENAME ::=	TITLE
TITLE ::=	(LETTER {LETTER | DIGIT}) | (QUOTATIONMARK LETTER {LET-TER | DIGIT} QUOTATIONMARK) *!! A TITLE must never begin with a number, but can be a mixture of letters and numbers, of a maximum of 128 characters, and if the title is a reserved word under SQL:1999, the title*

	must be enclosed by quotation marks, cf. ISO/IEC 9075:1999 - Database Language SQL (SQL:1999)
LETTER ::=	"A" |"B" | … | "Z" | "a" | "b" | … | "z" | "_"*!! As well as other national characters that are allowed in constraint names under the standard SQL:1999 !! underscore (U+005F)*
DIGIT ::=	"0" | "1" | "…" | "9"
SEPARATORCHARACTER ::=	";" *!! semicolon (U+003B)*
LINEBREAK ::=	(CR LF) | CR | LF
CR ::=	*!! carriage return (U+000D)*
LF ::=	*!! line feed (U+000A)*
ROW ::=	VALUE | (QUOTATIONMARK VALUE QUOTATIONMARK) {SEPARATORCHARACTER VALUE | (QUOTATIONMARK VALUE QUOTATIONMARK)} LINEBREAK *!! A separator character is not used after the last value in a row. See 9.G.1.b for enclosing a value containing separator characters and/or quotation marks*
VALUE ::=	LETTER | DIGIT | OTHERCHARACTERS {LETTER | DIGIT | OTHERCHARACTERS}*!! VALUE can be a mixture of letters, numbers and other characters !!*
OTHERCHARACTERS ::=	*!! all permitted characters in UTF-8, which are not letters and numbers, cf. 9.F.*
QUOTATIONMARK ::=	"""" *!! quotation mark (U+0022)*