THE NATIONAL ARCHIVIST

Draft 14.06.2004

Regulation pursuant to the Archives Act of 1 December 1999 No. 1566 concerning further technical and archival provisions regarding the handling of public records, Chapter VIII:

PROVISIONS CONCERNING ELECTRONIC RECORDS MATERIAL TO BE TRANSFERRED TO THE ARCHIVES

Laid down by the National Archivist on 11 May 2000 pursuant to Regulation of 11 December 1998 No. 1193 concerning public records (the Archives Act) § 5-8, as last amended on 14 June 2004.

1. General provisions

§ 1-1 Scope
These provisions concern central government bodies.

These provisions apply to electronic registry and recordkeeping systems as well as information of archival value stored in other electronic systems.

The National Archivist should, through particular, general or specific provisions, have decided in advance which systems are to produce archival copies for transfer or deposit, and what specific information should be stored from each particular system, cfr. §§ 3-20 and 3-21 of the Archival Regulation. See also the general provisions in "Felles bevarings- og kassasjonssystemer for statsforvaltningen" (common preservation and disposal systems for central government), specified in chapter III of the present regulation.

§ 1-2 Responsibilities
Upon *transfer* to the Archives, the National Archivist seizes ownership of records material as well as responsibility for all further maintenance, cfr. § 10 of the Archives Act.

When a copy of electronic records material is *deposited* in the Archives, cfr. §§ 3-17 and 5-2 of the Archival Regulation, the concerned body maintains ownership of as well as full responsibilities for the records material. The concerned body serves and maintains the material pending its subsequent transfer. The Archives preserve the deposited copy, unavailable for use, performing only technical maintenance.

Unless a need for review has been notified by the concerned body or the Archives, the status of a deposit automatically changes to transfer as soon as the time of transfer is reached.

§ 1-3 Choice of method for producing archival copy
An archival copy consisting of an extract of the data to be preserved from an electronic system should be produced and transferred to the Archives at specified intervals, cfr. also § 2-2. The National Archivist determines the method as well as the periodization to be used for the production of extracts from each particular system, cfr. Annex 1.

Separate provisions apply to registry and recordkeeping systems, cfr. Annex 2.

2. Transfer or deposit procedure

§ 2-1 Archival copy

Upon transfer or deposit, data should be exported from the system database (or similar) and converted into a specific, standardized format, cfr. § 4-1. This data extract should be deposited and preserved, independently of the original system, as the archival copy of that system.

Archival copies from all electronic systems, with the exception of registry and recordkeeping systems which are based on the Noark standard, should consist of the following:

a)   extracts from database tables (table extracts), cfr. §§ 4-1 – 4-4,
b)   a description of the table extract structure and contents, cfr. § 6-1,
c)   any associated electronic documents, cfr. §§ 5-1 – 5-3,
d)   documentation of the original system and the use of it, cfr. § 6-3,
e)   information concerning transfer or deposit, cfr. § 6-4.

Archival copies from Noark systems should conform to the specifications in Annex 2. In this case, neither a description of structure/content nor documentation of the original system is required, but information on the transfer or deposit should be enclosed, cfr. § 6-4.

§ 2-2 Time, frequency and mode of transfer

The National Archivist determines the time and frequency of transfer of archival copy from each system.

The National Archivist determines whether the status of the archival copy should be transfer or deposit, and may require depositing to take place prior to the time of transfer. Transfer would normally take place when the material is approximately 25-30 years old, but the National Archivist may also provide for transfer at an earlier or later date, cfr. § 5-2 of the Archival Regulation.

If the use of a system for administrative purposes is discontinued without the data being moved to another system, or if the operation of a body is discontinued, a complete *transfer* should be performed, cfr. § 5-3 of the Archival Regulation.

Systems ("registers") which are subject to licensing from the Data Inspectorate, should be *transferred* no later than the expiry date of the licence of the concerned body.

From electronic registry and recordkeeping systems, an archival copy should be *deposited* as soon as a records period (usually a five-year period) is completed, cfr. § 3-17 of the Archival Regulation. Any associated electronic documents must be enclosed on deposit. Associated documents as printouts must be enclosed as part of the normal transfer when the material is 25-30 years old, cfr. Annex 2.

§ 2-3 Implementation

Transfer or deposit of records material should be arranged with the Archives in advance, cfr. § 5-9 of the Archival regulation. For practical purposes, the Archives may request a sample of the archival copy to be delivered prior to transfer or deposit.

The costs of transfer and deposit should be borne by the concerned body, cfr. § 10 of the Archival Act.

A body which transfers electronic records material, must not delete the data on which the archival copy is based, until the transfer has been approved by the Archives.

§ 2-4 Approval

Upon transfer and deposit, the Archives should approve the archival copy and sign for the material received.

The Archives may refuse to approve of the archival copy if:

-       it does not contain the required components, cfr. § 2-1,

- the table extracts do not contain the required information,
- it does not fulfil the requirements with regard to archival medium, character set, formats or file and directory structures,
- errors or deficiencies in the data may be attributed to errors in the production of the extract itself, or
- the enclosed description of structure/content is erroneous or incomplete.
- If the material is not approved, the Archives may require the concerned body to produce a new archival copy. Alternatively, the Archives may demand reimbursement for the work necessary to make the material conform to the provisions, cfr. § 11 of the Archival Act.

§ 2-5 Lending back

Transferred or deposited electronic records material may not be lent back to the concerned body. Instead, a copy of such material may be supplied on the condition that the body covers all the costs incurred in the process, cfr. § 5-10. of the Archives Regulation

A copy of transferred material is supplied if this is not in breach of any provision adopted pursuant to a legal act which imposes time limits on the concerned body's access to the material.

3. Requirements with regard to archival medium

§ 3-1 Archival medium

Upon transfer and deposit, a physical storage medium should be used which has been approved by the National Archivist. The approved medium is CD-R. The use of any other medium requires special permission from the Archives.

The burning of the CD-R should be performed in one complete process, as "single-session". Each storage device is supplied in two identical copies. The storage devices should be of a quality that is readable on other equipment apart from that used for the burning process.

Each storage device is to be labelled with the name of the root directory (ARKIV1, ARKIV2, etc.). The "master" storage device is labelled A, whereas the "copy" is labelled B. CD-R discs should not have a label affixed, but be labelled using a non-bleed (felt) pen specially designed for CD-R.

Storage devices are to be stored in a closed plastic container with an affixed explanatory label.

§ 3-2 File names and file locations on archival medium

The files and directories on the storage device should be named and organized as specified in Annex 3.

When storing files, an approved storage standard for the medium in question should be used. The naming of files and directories on CD-R should comply with the ISO 9660: 1988, Level 1 standard. This means that only the following characters are accepted in file and directory names: the numbers 0-9, uppercase letters A-Z and the underscore character (_). File and directory names may have no more than eight characters. File names may have an extension consisting of up to three characters.

For electronic documents from Noark-4 systems, the extended ISO 9660 specification in Microsoft Joliet is approved with a view to enabling file names exceeding eight characters.

4. Format requirements for table extracts

§ 4-1 Table extraction from databases

Records from database tables should be extracted and transferred in sequential text files. Each table in a relational database should be extracted as a single file. The relationships between the tables/files in the extract should be described in the enclosed documentation, cfr. § 6-1 and Annex 6.

Extracts from older network databases and hierarchical databases should be normalized to relational form, as specified in Annex 5. If a body is not able to perform such normalization, the Archives may do this against reimbursement.

Table extracts from Noark-based systems should be produced in the format specified in the appropriate version of the Noark standard, cfr. Annex 2.

§ 4-2 Character sets

Table extracts and accompanying descriptions of structure and contents should be transferred as text-only in an unencrypted, unpacked and uncompressed form, and should use an approved character set.

The approved character set is ISO 8859-1, Latin 1. For Sami characters, ISO 8859-4: 1998, Latin-4 is used.

From IBM-compatible mainframe systems, the EBCDIC character set is also accepted, but only by arrangement with the Archives.

If a concerned body is not able to convert into an approved character set, the Archives may do this against reimbursement.

§ 4-3 Field lengths

The field lengths of table extracts may be up to 2047 characters. A field which is longer than 2047 characters, should be stored as an electronic document in a text-only format, cfr. § 5-3. The document should be associated with the table extract as an independent file in accordance with the provisions of § 5-1.

§ 4-4 Text-file formats

Text files containing table extracts should have one of the following formats:

a)   *"Fixed format"*, i.e., records and fields having a fixed length. The length and/or initial position of each field should be specified in the enclosed description of structure and contents, cfr. § 6-1 and Annex 6.

b)   *"Character-separated format"*, i.e., records and fields having variable length and being separated by fixed, unique delimiters. The record delimiter should be <CR><LF>. The field delimiters should be specified in the enclosed description of structure and contents, cfr. § 6-1 and Annex 6.

c)   *XML format.*

Specific provisions regarding formats of table extracts are provided for in Annex 4.

5. Format requirements for electronic documents

§ 5-1 Associating electronic documents with table extracts

Electronic documents should be associated with table extracts from a system for entry into records or another electronic system for registering documents, cfr. § 1-2 of "Bestemmelser om elektronisk arkivering av saksdokumenter" (provisions regarding electronic filing of documents) in chapter IX of this regulation.

The electronic documents should be stored as separate files and transferred together with the table extract. The table extract should contain a reference to the file name of each individual document. If a document consists of more than one file, the files should be grouped in a separate subdirectory. If file or directory names are modified during export, the changes should be reflected in the file references, cfr. <u>Annex 3</u>.

§ 5-2 The use of encryption and digital signatures

Electronic documents should be transferred in an unencrypted state, unless specific arrangements have been made for encryption. Such an arrangement must include rules for the use of decryption keys.

Electronic documents may have digital signatures applied to them based on public key cryptography, preferably in accordance with the pattern of Noark-4, chapter 10.2.

§ 5-3 Approved document formats for archival purposes

Electronic documents should be stored in document formats which have been approved for archival purposes. The following formats have been approved by the National Archivist:

-   Text-only — ISO 8859-1: 1998, Latin-1 and ISO 8859-4: 1998, Latin-4 for Sami characters
-   TIFF — Tag Image File Format, version 6
-   PNG — Portable Network Graphics, version 1.0
-   XML — Extensible Markup Language, version 1.0
-   PDF — Portable Document Format

The following formats have been approved for photographs and audio and video sequences:

-   For digital photographs as well as large digital images and maps: JPEG (ISO 10918-1)
-   For digital audio: MP3 (ISO 11172-3)
-   For digital video: MPEG2 (ISO 13818-2)

Specific provisions regarding the use of each individual format, provisions regarding the use of compression techniques and provisions regarding previously approved document formats which are no longer approved of for the purpose of transfer and deposit, are listed in <u>Annex 7</u>.

§ 5-4 Electronic documents from Noark systems

Electronic documents and electronic reports from Noark systems should, at the time of transfer or deposit, be stored as specified in <u>Annex 2</u>. See also § 5-1.

6. Requirements with regard to the documentation of data extracts

§ 6-1 Description of structure and contents

To enable the recreation and later use of electronic records material, a complete description of the structure and contents of the table extracts should be enclosed. The description should include the tables (i.e., files/record types) and fields in the extract transferred or deposited, *not* the table structure of the database from which the extract has been exported.

The description of structure and contents should be transferred electronically as specified in <u>Annex 6</u>. The description should be in a text-only format using an approved

character set. It should preferably use XML syntax. When the description has been designed in XML, it should be based on the syntax rules contained in a particular DTD (Document Type Definition) which the National Archivist has prepared and published specifically for this purpose: ADDMML — *Archives Data Description and Manipulation Markup Language.*

The National Archivist may on a case-by-case basis decide that the description *must* be XML-based; in such cases, the rules in the DTD named ADDMML should be followed.

A description of structure and contents is not required for extracts from Noark systems.

§ 6-2 Original searches in the data

In order to describe searches, data combinations and extraction methods which were used in the original system, the Archives may require that a number of SQL queries from the system be enclosed as documentation. These should conform to the SQL-92 standard (ISO 9075: 1992) and be adapted to the table extract. The SQL statements should be included in the description of structure and contents if the latter has been designed in XML.

§ 6-3 System and usage documentation

A summary should be provided with regard to the overall structure of the original system, including the relationships between the tables and the primary and secondary keys of the individual tables in the system database. This summary should take the form of an EAR diagram.

There should also be a summary description of the main system features. Alternatively, manuals may be enclosed which describe the system features. A manual with a description of screen captures and printouts *ought to* be enclosed if it exists.

The purpose of the system that has produced the archival copy, its use and its administrative context should be described. Any statutory authority for creating the system should be cited. If parts of the material are classified, subject to licensing or professional secrecy or withheld from public disclosure for some other reason, this should be made clear in the documentation.

The documentation should be in electronic form, in one of more of the formats specified in § 5-3. The file names and locations of the archival medium should be as specified in Annex 3.

Extracts from Noark systems do not require system or usage documentation.

§ 6-4 Information on transfer or deposit

The following information regarding transfer or deposit should be enclosed:
a) the name of the originating organization,
b) the name of the system from which the data extract of the archival copy originates,
c) the time span of the extract,
d) the number of records in the extract (for the major tables) and other quantification of the size,
e) the number of storage devices enclosed, both masters (A versions) and copies (B version),
f) the name of the person responsible for the transfer in the originating organization.

The information should be in electronic form, as text-only having an approved character set. The file names and locations on the archival medium should be as specified in Annex 3.

7. Annexes

Annex 1:
Extraction method for the production of archival copy for transfer or deposit (cfr. § 1-3)

Before transfer or deposit, it should be established how the data extraction from the concerned system for archival purposes should be performed. The extraction method must be based on the updating method employed in the concerned system. Alternatively, information may be updated through
- overwriting of existing data,
- overwriting combined with a history log (history information),
- appending new records (accumulation),
- a combination of these methods.

Based on the updating method used, the National Archivist determines whether the extract of the system archival copy should be produced as a *section* (snapshot or last status, usually as an annual version) or include accumulated information for specific time intervals.

The National Archivist determines on a case-by-case basis whether an archival copy's data extract should include:
1) all (or most) tables from the system,
2) one or more derived (consolidated) tables,
3) one or more reports stored as electronic documents in one of the document formats specified in Annex 7.

Provisions concerning extracts from Noark systems and other registry and recordkeeping systems (cfr. § 1-3)

1. Common provisions for Noark systems

According to § 3-17 of the Archival Regulation, extracts from electronic registries, including any associated electronic documents, should be *deposited* in the Archives as soon as a registry period has expired. That part of the case records which consists of printed documents, should be transferred to the Archives first as part of the normal *transfer* when the material is approximately 25-30 years old. Electronic documents may only accompany extracts from Noark-4 systems which have been approved for electronic recordkeeping ("O2"-approved systems).

The registry information should be exported from the database as table extracts spanning entire records periods, cfr. the provisions for periodization and remote storage in the Noark standard. The extract should include:

1) all the cases that were completed during the period as well as all the registry entries in these cases,
2) any electronic case documents associated with the registry entries.

Enclosed with the extract, there should follow two electronic reports:

3) a complete report, ordered by case, which contains all the cases that were *completed* during the period as well as any registry entries associated with these cases,
4) a complete, chronologically ordered report which contains all the registry entries as well as the record date within the period (in Noark-4: the concerned records section), i.e., all the registry entries that were *registered* during the period.

2. Systems based on Noark-3 (1994) and Koark (1995)

Extracts based on Noark-3 and Koark should include all the completed cases from a *registry period*. The extraction should be performed according to the specifications of Noark-3 (1994), chapter 15.5.1, *option 1*, with the following amendments:

- The medium should not be magnetic tape, but as specified in § 3-1.
- The filing plan of the registry system should be enclosed. The filing-plan table should be named ARKIV.DAT and should have the following format: a fixed 16-position file code and fixed group text (code text) of 255 positions. No record delimiter should be used.

The following reports in electronic form should be deposited together with the table extract:
1. *A final report ordered by case* (see Noark-3, ch. 14.3.4).
2. *A final, chronological registry report* (see Noark-3, ch. 14.3.1). In order to encompass all the registry entries that were registered during the period, this must be produced *before* completed cases are separated from the active database at the time of reorganization.

The reports should be transferred in document format, either as text-only, as PDF documents or as XML documents accompanied by a DTD conforming to the format requirements of Annex 7.

Two *electronic* reports should subsequently accompany the paper-based case records when they are *transferred*:

- *Transfer list*
- *Downgrading list*, in those cases where this feature is used.

These additional reports should be transferred in document format, either as text-only, as PDF documents or as XML documents accompanied by a DTD conforming to the format requirements of <u>Annex 7</u>.

For extracts based on Noark-3 and Koark, no other accompanying documentation is required than the files containing information on the transfer or deposit itself, cfr. § 6-4.

3.   Systems based on Noark-4 (1999)

Extracts based on Noark-4 should include one or more completed entire *records sections*. The extract should be produced in the transfer format specified in Noark version 4-1 (the National Archivist, 2002), including any subsequent supplementary specifications from the National Archivist.

The format of the attributes, their distribution among tables and the specific attribute and table names which are used in Noark-4, are mandatory during data export for transfer and deposit. Extracts from Noark-4 systems using any other internal structure or deviating table and attribute names must be converted into this format prior to transfer or deposit.

Extracts from the DOKVERSJON table should include references to file names for all enclosed electronic documents.

The electronic reports which should accompany table extracts from Noark-4-based systems, are:

1.   *Case and document summary* (see Noark 4.1, ch. 11.3.4).
2.   *Registry* (see Noark 4.1, ch. 11.3.3).

The format, etc., of these reports should subsequently be enclosed with the paper-based records on *transfer*:

- *Remote-storage and transfer lists*
- *Records summary* (only in cases where transfer or deposit includes data from several records sections)
- *Downgrading list* (only in cases where this feature is used)
- *Disposal list* (only in cases where this feature is used)

These additional reports should be transferred as table extracts in text-only format or as electronic documents in one of the formats specified in § 5-3.

For extracts based on Noark-4, there should be enclosed specific documentation from the system vendor which describes the parts of Noark-4 which are included in the system. No

further accompanying documentation is required apart from the files containing information on the transfer or deposit itself, cfr. § 6-4.

4.   Systems which have not been approved as Noark systems

Registry and recordkeeping systems which have not been approved as Noark systems, should produce extracts which as far as possible conform to the specifications for Noark systems. Extracts from systems which do not conform to the Noark standard, should be adequately documented as specified in §§ 6-1 – 6-4, unless the National Archivist has decided otherwise for an individual system.

Annex 3:
Provisions concerning file organization on archival media (cfr. §§ 3-1, 3-2, 5-1, 6-3 and 6-4)

If an archival copy consists of more than one storage device, a root directory bearing a unique name, e.g., ARKIV1, ARKIV2, etc., should be created on each unit.

At the *root level of the primary storage device*, there should be a file called INFO.TXT. This file should contain the transfer or deposit information specified in § 6-4.

A file called INNHOLD.TXT should exist on *level one under* the root directory on *each storage device*. This file should provide a summary of the directory and file structure of the concerned unit. If the extract includes electronic documents, the complete directory structure of the document files should be described. Furthermore, information should be provided as to the naming principle for document files.

The following information should be located on *level one under* the root directory on the *primary storage device*:

- The *description of structure and content* which accompanies the archival copy (cfr. § 6-1). This file should be named METADATA.XML if the description is XML-based (ADDMLL), or METADATA.TXT if it is not XML-based.
- *System and usage documentation* (cfr. § 6-3) if it exists in electronic form. A *directory* named SYSDOK should be created, and the files containing the documentation should be located under it (on level two under the root directory).

Files containing table extracts, any reports and electronic documents should be organized in the following manner:

- *Table extracts* should be located on *level one* under the root directory of each storage device. The file names should be unique, and should include the DAT extension when a "fixed format" or "character-separated format" is used, or the XML extension when the format is XML. A file may be split up if it is too large to fit onto one storage device. In such a case, the file extensions should be 001, 002, 003, etc., while the file name remains the same. This split must not result in the splitting up of a record into several parts.
- *Reports* (electronic printouts) from the computing system should be stored under a *directory* named RAPPORT on level one under the root directory of the concerned storage device. The report files should be grouped under this directory (on level two under the root directory).
- For *electronic documents* should be created a *directory* named DOKUMENT on level one under the root directory of the concerned storage device. The document files should be grouped under it (at, or starting from, level two under the root directory). The files may optionally be structured into several subdirectories. It is recommended that the number of documents in a directory does not exceed 5000.
- Documents which make up an electronic outgoing letters database, should be stored under a directory named KOPIBOK on level one under the root directory of the concerned storage device. The document files should be grouped under it (at, or starting from, level two under the root directory). If there are separate table extracts

referring to the documents of the outgoing letters database, this should be documented in INFO.TXT.

The file format for accompanying electronic documents and documents with system or usage documentation should be indicated by the file extension, e.g., TXT, XML, PDF or TIF. If a single document consists of more than one file, these should be grouped together in one directory having a unique name. File and directory names must conform to the pertinent references in the table extracts.

Annex 4:
Provisions concerning the formats of table extracts (cfr. § 4-4)

1. Exporting different data types
During export to text files in an approved character set, the various data types of the original database should have the following layout:
- When a "fixed format" is used, text strings should be left-adjusted in the fields. Subsequent positions should be filled with blanks. Integers and decimal numbers should be right-adjusted without leading zeros, and the prior positions filled with blanks.
- Negative numbers and decimal numbers should have a minus sign as the first digit. Positive numbers are not normally represented by a plus sign.
- Decimal numbers should be have a comma as a decimal separator.
- Dates should be laid out as YYYYMMDD (ISO 8601).

2. The use of packed and binary numerical fields
Packed numerical fields of the type "packed decimal" or "Comp" and binary numerical fields are not permitted in table extracts. If a body is not able to perform the unpacking of packed fields, the Archives may do this against reimbursement.

3. Table extracts in XML format
Table extracts in XML format should conform to version 1.0 of the XML standard.

Annex 5:
Provisions concerning the normalization of extracts from old databases (cfr. § 4-1)

In extracts from a database of an old type (network database or hierarchical database), where internal files may contain records of more than one type, data should be extracted in the following manner:

- Each record type should be exported to one single sequential file. The original physical pointers should be converted into logical keys (primary or foreign keys).
- Repeated groups of fields (i.e., one or more fields which are repeated within a record) should be separated from the original record and exported as separate records in a new file with associated keys).

Annex 6:
Provisions concerning the documentation of table extracts (cfr. § 6-1)

Structure and content description of table extracts

For all data files which contain table extracts, there should at the time of transfer or deposit be enclosed a complete structure and content description in order to permit later use of the material.

The description should be enclosed in electronic form. This should be text-only in an approved character set (cfr. § 4-2). The description should include the tables (i.e., files/record types) and fields of the extract being transferred or deposited, *not* the table structure of the database from which the extract has been exported.

The documentation should include:

- *File and record description*
  Each single table (file/record type) is to be specified. A short description of the contents is provided. Relationships between tables are indicated, and primary keys and foreign keys as well as cardinality are defined.
- *Field description*
  The field names of all tables (files/record types) are described. The data types of the fields are specified. When a "fixed format" is used, the initial position and length of each field is specified. The contents of each single field must be explained if it is not apparent from the field name.
- *Code explanations*
  In those cases where fields contain coded values, complete code explanations are required. If the system stores the codes in separate help tables, extracts from these should be included in the archival copy.

When the structure and content description uses XML syntax, it should be based on the rules in the DTD (Document Type Definition) which the National Archivist has prepared and published specifically for this purpose: ADDMML – *Archives Data Description and Manipulation Markup Language*.

By special arrangement with the Archives, the structure and content description may be based on another XML DTD than ADDMML. In such a case, the alternative DTD which is used, must accompany the description as documentation.

The structure and content description should be in a separate file, cfr. the provisions of Annex 3. When the description uses XML syntax, the file should be named METADATA.XML. When XML syntax is not used, the file should be named METADATA.TXT.

Provisions regarding document formats (cfr. § 5-3)

1. Summary of approved document formats

The National Archivist continually reviews the provisions regarding document formats for archival purposes.

Given the limitations and specifications which follow from point 3 below, electronic documents may as of March 2004 be stored in the following document formats when being transferred or deposited:

- Text only – ISO 8859-1: 1998, Latin-1 and ISO 8859-4: 1998, Latin-4 for Sami characters
- TIFF – *Tag Image Format*, version 6 (Aldus/Adobe, 1992)
- PDF – *Portable Document Format*
- XML – *Extensible Markup Language*, version 1.0 (1998)
- PNG – *Portable Network Graphics*, version 1.0 (1996)

For digital photographs and large, digital images and maps, the following has been approved:

- JPEG (ISO 10918-1)

For audio and video sequences, the following have been approved:

- MP3 (ISO 11172-3) for digital audio
- MPEG2 (ISO 13818-2) for digital video

2. General provisions concerning compression

For normal, scanned case documents, lossless compression is accepted in accordance with CCITT/TSS group 4 for documents in black-and-white and LZW *or* Deflate (LZ77) for colour/grey scale documents. Specific requirements have been laid down regarding the implementation method for compressing TIFF files, cfr. point 3 below.

For audio files in MP3, video files in MPEG2 and photographs and large image files in JPEG, even lossy compression is accepted. Requirements with regard to JPEG are specified below.

3. Specific provisions concerning individual formats

TIFF:

TIFF documents may be "single page" or "double page". If "single page", each page of the document must be stored in a separate file under a shared directory. TIFF documents should not contain user-defined tags.

When Deflate compression is used in TIFF documents, compression 8 should be used. The older compression 32946 (80b2 HEX) is not accepted.

When JPEG is used in TIFF (see below), compression 7 should be used.

PNG:

For PNG files, only Deflate/LZ77-based compression is used.

JPEG:

Compressed JPEG documents must not be modified and stored several times, as this could result in repeated compression and further reduced image quality. JPEG documents stored in TIFF are also accepted (as "multiple page"), provided that the provisions concerning JPEG compression in TIFF specified above are observed.

PDF:

Adobe PDF (version 1.0 – 1.5) are accepted *temporarily* pending the definition of a limited subset of this format called **PDF-A** as an ISO standard specifically for archival purposes. Adobe PDF may be used with the following restrictions:

- The documents may only use fonts in accordance with Type 0, Type 1 (including "multiple master fonts"), Type 3 and TrueType.
- All Type 0, Type 1 and TrueType fonts used in a PDF document should be embedded in the document.
- All security features which are possible in PDF documents, such as passwords and encryption, must be disabled at the time of transfer.
- The documents should not contain schema data in separate fields, comments (annotations and thumbnails), embedded files (including audio, video and other multimedia objects) or links to other documents, files or web sites.
- Scanned documents which have been converted into PDF, should be transferred in their original version, not in a version where the original is partially or completely replaced by OCR-scanned text.

XML:

For XML documents, an associated DTD (Document Type Definition) and any associated template files (e.g., stylesheet) should be enclosed as separate files in those cases where DTDs and template files have been used. As an alternative to a DTD, an associated XML schema may be enclosed. Graphics files which are referenced in XML documents (e.g., GIF or JPEG files), should for the purpose of transfer be converted into PNG. Alternatively, the entire document may be converted into PNG, TIFF or PDF.

HTML documents are *not* accepted. These should, before transfer, be reconverted into XML or converted into another approved document format.

4. Previously approved document formats

The following document formats, which were approved according to the transfer provisions laid down in May 2000, are no longer approved of as document formats for archival purposes:

- TIFF/IT (ISO 12639)
- SGML
- HTML

These formats will no longer be accepted for archival purposes after 1 January 2005. This also applies to Adobe PDF without the restrictions specified above.