

5. MODULE FOR ELECTRONIC RECORDKEEPING

5.1 Purpose of module

The purpose of this module is to link the records management module to electronic documents which constitute case records. Even before Noark-4, it was possible to link electronic documents to registry entries, but the specifications of Noark-3 and Koark were on a general level. Besides, the archival copies were meant to be in hardcopy (paper). As from Noark-4, it is possible to make the transition from paper-based to electronic recordkeeping. The module for electronic recordkeeping should, however, be able to handle a situation where some of the case records are still in paper form.

The relationship between this module and the rest of the system is shown in figure 5-1.

The electronic records must contain the same types of documents as those which are currently stored in paper form: text documents, construction sketches, maps, pictures, etc.

Electronic case records should fulfil the same role as traditional paper-based records; they should, among other things:

- provide information to executive officers for use in the handling of cases
- document for posterity the processing that has been done

However, electronic recordkeeping provides new opportunities:

- The case documents are immediately made available to the workstations of the executive officers. This will replace traditional distribution and lending.
- The contents of the stored documents may be re-used in new text production.
- It is easier to give the public insight into ungraded documents.

However, electronic recordkeeping as opposed to traditional, paper-based recordkeeping, also entails some problems and challenges which this module need to address:

- The documents must be readable and available even in the future; this is made more difficult by a multiplicity of production formats which are frequently released in updated versions.

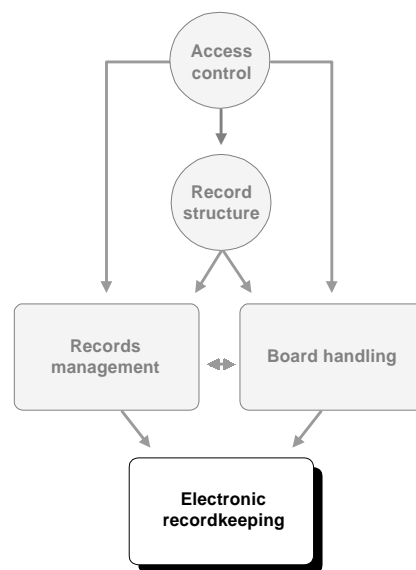


Figure 5-1: Module for electronic recordkeeping in Noark

- In connection with periodization and reorganization, it must be possible to export the documents for storage in an archival repository together with its associated registry information.
- It must be possible to guarantee that the documents are authentic (not forged).
- In a paper-based entity of records, corrections, notes, etc., on the drafts may provide valuable information on the processing of the case, and there should be mechanisms which preserve this kind of information.

5.2 Module design

It may be useful to start off by clarifying how this module relates to the records management module and take a closer look at the types of documents which may be eligible for storage as electronic records.

5.2.1 How it relates to the records management module

The documents in the electronic records should be linked to the registry entries of the records management module. This is the starting point for searching for registered electronic documents, and this is where the access rights for these documents are managed. Thus, the records management module is the gateway to the case records, be they paper-based or electronic.

K5.1	All stored documents in the electronic case records should be subject to Noark's records management. All access control is managed from here.	O2
------	---	----

In Noark-3 and Koark, it has been possible to associate a single document with a registry entry. Noark-4 should allow more than one document to be associated with the same registry entry. There may, for instance, be one main document and several attachments. It should, furthermore, be possible to associate a document with several registry entries, for instance as a main document in one registry entry and as attachments in others. This would prevent multiple storing of documents in the electronic records and ensure that the document appears as identical in all contexts as well as guaranteeing uniform access rules for the document in question.

K5.2	A registry entry may refer to more than one document. Only one of these may constitute the main document.	O2
K5.3	A document may be associated with several registry entries. However, it may constitute a main document only in one of these.	O2

5.2.2 Versions, variants and formats

In paper-based records, it is common to store drafts with their scribbled corrections, notes and authorizing signatures (initials) from the responsible executive officers at various steps and levels of the processing. This may provide valuable information on processing and evaluations that have taken place. In the context of electronic recordkeeping, it is often desirable to have several *versions* of a document (draft) in order to identify what has been

done and authorized by whom during the processing. Separate versions addresses the need for guarantee of integrity. The case handling system may be designed so as to preserve all or selected versions of a document.

In a *recordkeeping system*, one may choose to store only the final (finalized) document. However, Noark-4 should also provide for the storage of previous *versions* of a document when this is significant in order to document the processing. Common practice from paper-based recordkeeping may be perpetuated in Noark-4 in a modified form.

Any electronic document is from the beginning stored in a technical *format* determined by the tool used to produce the document. If many different tools are used, for instance one or more text editors, spreadsheet applications, graphics applications, etc., the result will be records consisting of an infinite variety of different formats. It must be possible to convert these formats into a few standardized *archival formats* suitable for long-term storage. This is discussed in more detail in paragraph 5.3. However, even when a document has been converted to such an archival format, it may be desirable to preserve the original *production format*, e.g., in order to reuse text. For this to be possible, a document must exist in both an archival format and the production format, possibly in both formats simultaneously.

In Noark-4, it should also be possible to store public versions of documents which are exempt from public access. Likewise, it should be possible to store documents to which digital signatures have been applied, as separate editions. In the latter case, there may for instance be solutions where both the document itself, the signature and the signature certificate are stored in the same file (see paragraph 10.2). Noark refers to such public and digitally signed copies of a document version as *variants* (in practice, however, it is often natural - and unproblematic - to refer to them as "versions"). A *variant* is subordinate to an ordinary (and usually final) version of a document, and is always stored together with this.

Figure 5-2: Versions, variants and formats

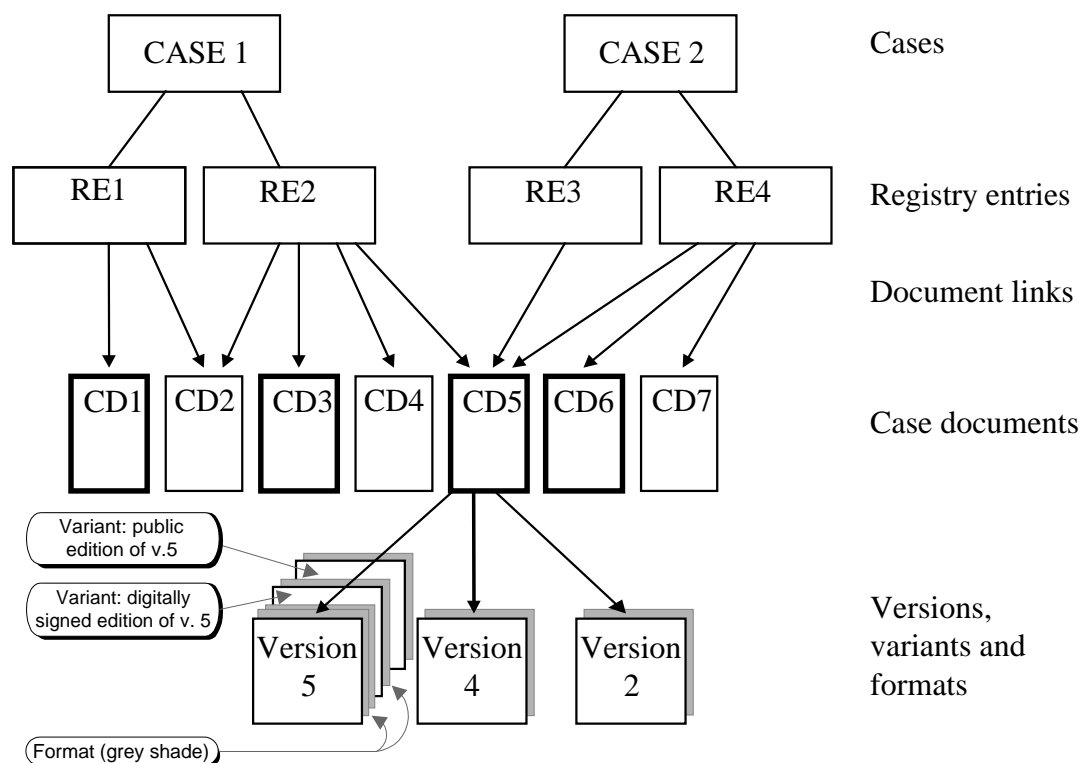


Figure 5-2 illustrates the relationship between cases, registry entries and formats. The figure shows an example of how case documents may be associated with one or more registry entries within the same case or belonging to several cases. Main documents are indicated by heavier outlines. This example also shows how a case document may exist in several versions, variants and formats.

- Case document CD1 is only associated with registry entry RE1 as a main document.
- Case document CD2 is associated with both registry entries RE1 and RE2 as an attachment.
- Case document CD3 is only associated with registry entry RE2 as a main document.
- Case document CD4 is only associated with registry entry RE2 as an attachment.
- Case document CD5 is associated with registry entry RE3 as a main document and with RE2 and RE4 as an attachment.
- Case document CD6 is only associated with registry entry RE4 as a main document.
- Case document CD7 is only associated with registry entry RE4 as an attachment.

Case document CD5 has three stored *versions*: 2, 4 and 5. The absence of versions 1 and 3 from the records would normally indicate that they are not considered to be of archival value and have thus been deleted, or that they have not been transferred from the case handling system. Versions 2 and 4 are drafts to be preserved, whereas version 5 is the final version of the document.

Version 5 also has two *variants*: one public and one digitally signed edition.

All versions and variants are stored in a *format* (shaded grey in the illustration). The figure does not distinguish between archival formats and production formats. It does, however, indicate that version 5 has been stored in two formats, i.e., both in archival format (e.g., PDF) and production format (e.g., Word97).

K5.4	It should be possible to store a document in several versions which reflect different stages of the development towards the final document.	O2
K5.5	It should be possible to store the same version of a document with several associated "variants", i.e., alternative editions of the version which have had digital signatures applied to them or been specially adapted for public use.	O2
K5.6	If a document is stored in several versions and/or variants, this should appear clearly from the relevant panels.	O2
K5.7	It should be possible to store the same <i>version</i> <u>both</u> in the production format and in an archival format (format for long-term storage). It should be possible to store a <i>variant</i> <u>either</u> in the production format or in an archival format.	O2
K5.8	If several formats exist for a document (or a version of a document), this should appear clearly from the relevant panels.	O2

5.2.3 Paper-based vs. electronic recordkeeping

Noark should be able to handle *both* paper-based and electronic storage of documents. Noark allows for combined storage of cases with paper documents and cases with electronic documents. Normally, however, all *main documents* within a case should be stored electronically in order for the case to be considered as electronically stored. If this condition is not satisfied, the entire case must be stored on paper. Any electronic documents which may exist within such a case, are regarded as work (copy) versions.

The *Case* table should contain an attribute called *Stored on paper*, which shows whether the case documents are stored in paper-based or electronic form. If this attribute has a value which says that the case is stored electronically, then the system should automatically check that all *main documents* within the case have been stored in electronic form.

In electronically stored cases, all attachments *ought to* be stored electronically as well. It is, however, permissible to store certain attachments, such as extensive reports, on paper. In such cases, the system must contain a clear reference to the location of the paper-based attachment. The attribute *Archival note* in the table *Version* (see paragraph 14.3.3) is used to refer to the paper document.

In addition to this, it should be possible to store a large *main document* in an electronic case on paper if the first page is scanned or a reference document inserted instead of the main document. The existence of a paper-based archival copy is identified through a front-page marker where the attribute *Association code* in the table *Document association* has the value "FH" (see paragraph 14.3.9).

Procedures relating to electronic recordkeeping are described in more detail in paragraph 5.5.2.

K5.9	The system must distinguish between cases where the official and valid documents are stored on paper, and cases where they are stored electronically.	O2
K5.10	If a case is stored electronically, the system should check that all <i>main documents</i> which belong to the case, are also stored electronically.	O2
K5.11	Even if a case is stored electronically, the system should allow for <i>attachments</i> to be stored on paper. The system must provide a unique reference to the physical location of the attachment.	O2
K5.12	The system should also provide for substituting a reference document which refers to an archival copy on paper, for a main document in an electronically stored case.	O2

5.2.4 Module design

The main tables of the module are *Document link*, *Document description* and *Version*. The following figure shows how the tables relate to each other. Tables residing in other modules are in grey shade.

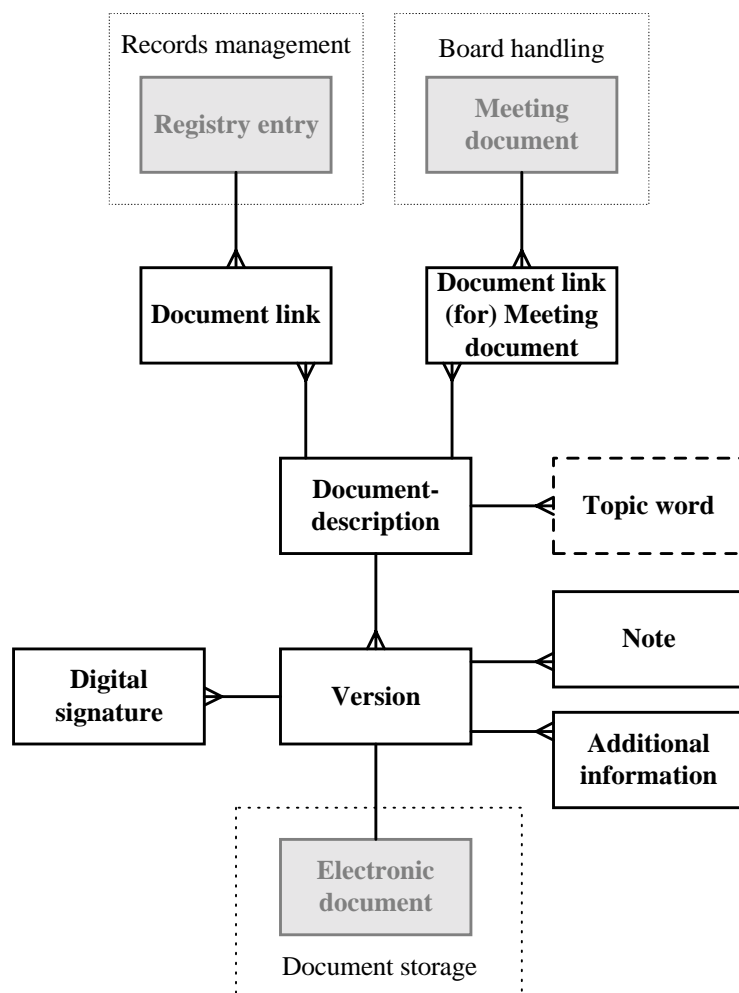


Figure 5-3: Electronic recordkeeping - simplified data model

Brief description of the contents of the above tables:

- *Document link* (cfr. paragraph 14.3.1): This table is introduced in order to dissolve the M:M relationship between registry entries and documents. A document may be associated with the registry entry as a main document, a dispatch letter, an attachment, etc. Thus, there must be an attribute in this table which shows what kind of association we are dealing with. One of the documents must always be the main document.

There is a similar association between the electronic records and the board-handling module, called *Document link meeting document* (cfr. paragraph 14.6.6). From the board-handling module, reference may be made to the board documents which are not entered into the records, and which are thus not referred to from the records management module. These may for instance be minutes, which constitute a separate

series within the records. Board documents are discussed in connection with the board-handling module in chapter 9.

- *Document description* (cfr. paragraph 14.3.2): All electronically stored documents should be linked to this table. According to SGK specifications, there should be an equivalent table in the case handling system. If the case handling system is integrated with Noark, the contents of several of the table's attributes will normally be registered from the case handling system, such as *Document title* and *Document category* (letter, memo, report, etc.). Such registration of information in Noark is primarily carried out in connection with attachments. The relationship between electronic records and SGK is described in more detail in paragraph 5.2.6 below.

The attribute *Document status* in the table *Document description* is a "flag" which shows if documents are being produced (status B) or if they have been processed by the executive officer (status F). The use of status values during document production is part of Noark's process management and is described in more detail in chapter 6.

It should be possible to assign an access code to an attachment that is different from that of the registry entry. For this reason, the table *Document description* must contain the attributes *Access code* and *Access group* - as well as the attributes *Authority to exempt from public access*, *Date of downgrading* and *Downgrading code*.

- *Version* (cfr. paragraph 14.3.3): It must be possible to store a document in several versions, and to store a version in several formats. Information on versions and formats are stored in the table *Version*. There is a 1:M relationship between *Document description* and *Version*. The *Version* table also constitutes the interface to the electronic document itself, and there is a 1:1 relationship between the two.

One of the attributes in the *Version* table is *Version number*. The last version of the document has the highest number. The version number must be combined with another attribute called *Variant*. This shows what edition or what format of the version the individual registry entry contains. Permissible values for *Variant* are production format (P), archival format (A), signed document (S) and public version (O). If a specific version consists of several variants, the same version number may occur more than once. The document format (Word 6.0, TIFF 6.0, PDF, etc.) for the version/variant should be registered (automatically) in the attribute *Storage format*. Based on this, the records entity should check that documents are stored in an archival format (A).

- *Note* (cfr. paragraph 14.2.30): It has been described how several versions of the same document may contribute towards documenting the processing. However, it is also necessary to document the processing by other means. It should be possible to register comments and remarks directly related to a document in the same way that we register notes on cases and registry entries. It should be possible to link notes to the tables *Document description* and *Version*, and there should be a 1:M relationship between these and the table *Note*. The text which constitutes the note, is registered in a separate attribute called *Note*, and this should be so designed as not to impose any length restriction on the registered information.

Primarily, all notes should be registered in the attribute *Note*, but it should also be possible to link electronic documents to the note. These may, for instance, be handwritten notes that have been scanned. There should, therefore, be a 1:1 relationship between *Note* and the table *Document description*.

K5.13	It should be possible to link one or more notes of arbitrary length to any document or document version.	O2
K5.14	It should also be possible to use scanned documents as notes.	A

Digital signature (cfr. paragraph 14.3.4): During exchange of e-mail, digital signatures may be used to verify the authenticity of documents (ensuring the addressee of the identity of the sender) as well as their integrity (that the contents have not been modified). Digital signatures may also be used for guaranteeing the authenticity and integrity of documents which are filed in the electronic records. The signature must then be applied to the document after this has been converted into an archival format.

In Noark, it should also be possible to file documents with digital signatures and their associated certificates as a separate variant S. Such documents will then still be verifiable. It should, furthermore, be possible to store signatures and certificates in a separate table: *Digital signature*. In this table, it should also be possible to store information such as who verified the signature and when. This preserves traces of signatures and verification even in cases where opportunities for further verification are lost. There is a 1:M relationship between *Version* and *Digital signature*.

Digital signatures and e-mail is described in more detail in chapter 10, cfr., in particular, the requirements formulated in K10.36 - K10.53.

- *Additional information* (cfr. paragraph 14.2.31): The system should automatically log certain information in connection with the electronic storage of a document. This applies to the filing time and who carried out the filing. Likewise, information should be logged with regard to when the document was converted into an archival format and who converted it. Such log information is stored in *Additional information*. Both *Document description* and *Version* are linked to this table in a 1:M relationship. Even *Case* and *Registry entry* are linked to the table *Additional information* (see part II, Technical specifications, for a closer description of how the log function may be implemented).

K5.15	The system should log the filing time and the person who carried out the filing.	O2
K5.16	The system should log the time when a document was converted from its production format into an archival format, as well as the person who carried out the operation.	O2

5.2.5 The electronic document storage

Noark-4 does not specify how documents are to be stored and organized in the electronic document storage. The documents may be stored in tables in a database system, or as individual files in a file system. It is up to vendors to choose an appropriate solution. The chosen solution must, however, support the export of documents to individual files as described in paragraph 5.4.

Vendors must also decide if documents are to be stored in the same system as the registry information (Noark), or if they are to be stored in a separate system. If Noark is integrated with a case handling system, it seems natural that the two systems share a common document storage. However, irrespective of the solution selected, the individual electronic document should be linked to the table *Version* in a 1:1 relationship.

5.2.6 How it relates to SGK

Electronic document storage is thoroughly discussed in the report *Elektronisk saksbehandling. Statens generelle kravspesifikasjon* [Electronic case handling. General specification of requirements for the state administration] (Statskonsult 1997). This report presupposes that all documents which executive officers need in order to perform their task, should be available from a common document storage. Many - but by no means all - of these documents are registered documents (case documents). The storage also contains previous versions (drafts) of finished and registered documents, and there may be notes, reports and any kind of collected information which executive officers must have access to, but which are not to be entered into records.

SGK, like Noark, presupposes that all documents are linked to a document description. If the case handling system is integrated with Noark, this document description may be a table which is shared by the two systems. However, if a solution with separate tables is chosen, there must be mechanisms which automatically copy information from one table to another.

In a case handling system, one may choose to register a topic (subject) word and selected keywords when documents are produced. It should also be possible to search for these keywords from the records management system.

Electronic document storage provides for free-text searching in the documents, and in the case handling system it should be possible to retrieve documents by searching in the document text itself. It should also be possible to perform free-text searches from Noark.

K5.17	It should be possible to search for registered keywords and topic words in the document description of individual electronic documents.	A
K5.18	It should be possible to do free-text searches in the documents of electronic records.	A

5.3 Document formats

5.3.1 Production formats

Documents may be stored in a number of different formats. By document format is understood the way in which characters, structures and layout are coded and organized. The original format of the document depends on the tool used to produce it. In Noark-4, such an original format is referred to as the *productin format*.

These formats may be divided into different categories, the most important of which are:

- character formats where only the characters (letters and numbers) are stored
- text formats which also preserve the structure and layout of the document
- graphics formats for storing pictures (often subdivided into raster graphics and vector graphics)
- video formats for storing "moving pictures"
- audio formats
- multimedia formats which combine text, layout, graphics, video and audio

The two most important formats for Noark-4 are text formats (produced by text editors) and raster graphics (scanned paper documents). Vector graphics may also occur (e.g., illustrations produced using CAD/CAM tools). Graphics elements are frequently pasted into text documents, e.g., as letterheads and logos. There may also be formats which are difficult to fit into the above categories, such as spreadsheets and digital maps.

Many of these production formats are so-called proprietary, i.e., they are specific to an individual vendor. In many cases, they are not openly documented. To read such a format, we must often use the same tool that was used to produce it. In many cases, we must also have the same version of the tool to be able to read the document. New versions are typically released frequently, which means that what is apparently one and the same format, is continuously changing. There are, admittedly, several "viewers" on the market which can read (but not edit) a number of formats, but there are hardly any viewers that can read all pertinent formats. Furthermore, viewers must be constantly upgraded in order to be able to read the latest versions.

The result of this development of continuous change will be that many organizations will not be able to read their own text files which are more than 5-7 years old. If availability is to be maintained, these must be constantly converted into newer versions. And if we still store many different formats, this converting job will grow out of proportion and quality control will be very difficult.

One of the problems is the limited durability of electronic storage media such as tapes and CD's. However, the development of formats and software is a much more critical factor than the period that data may be stored intact on a certain medium. The need to convert into new formats will occur far earlier than the need to copy to a new medium due to limited durability.

When electronic documents are transferred to state or local depot institutions, these institutions take over responsibility for maintaining the availability of the documents. If they were to receive a multiplicity of document formats, the situation would be impossible to handle. It goes without saying that they cannot be expected to possess all the software (in the correct versions) that was used to produce the documents transferred to them. Long-term storage in a proprietary format is out of the question for the depot institutions. The solution must therefore be to find one or a few stable and universal formats.

5.3.2 Archival formats

A precondition when storing case documents electronically - without, at the same time, storing the paper printouts of those documents - is that the availability of the documents is

maintained for a long time. This is a fundamental requirement. It affects something as fundamental as the possibility of maintaining Norwegian administrative practice. The opposite would be a situation where the public administration wipes out its own traces, an administration where documentation and the general public's right to information is limited to a few years.

It would be possible to maintain availability for some time if the documents could be converted to a *standardized format*. Such a format must be openly documented, and preferably approved as an ISO standard. Standardized formats may be read across computer platforms and operating systems, and many of them have been developed for use as exchange formats. Such standardized formats are hereafter referred to as *archival formats*.

Not even an approved and well-established standard is guaranteed to last forever. Even standards have their generation gaps. Even if we select a certain standard as archival format, we must be prepared for later conversions, but such conversions will occur less frequently, and they will presumably be less risky and more complete than conversions between arbitrary production formats or between versions of the same format.

A Noark system with electronic records should have functions for converting between production formats and archival formats. This function should be easy to use. It should provide for converting individual files as well as for "mass conversion" of major parts of the document storage as batch jobs.

The conversion process must maintain the integrity of the documents. The basic requirement is that the contents (the "text") are rendered exactly as in the original document. It is also desirable to maintain the look and layout of the original document. Conversion into certain formats will result in losses of parts of the visual integrity, and conversion into plain text will lose any formatting that existed in the document.

It should not be possible to edit a document that has been stored in an archival format. This is a security feature that must be built into the system. It is described in more detail in chapter 6. However, even if a document is locked and cannot be edited, it will still be possible to reuse text (from a number of formats) by cutting and pasting.

It should, if desired, be possible to store the same version of a document in both the production format and the archival format. The reason may be a wish to have access to all the edit functions of the production format. For long-term storage, it usually suffices to keep the archival format. During transfer to archival depot, it is usually the archival format that is transferred.

K5.19	It should be possible to convert documents from the production format to a standardized and openly documented archival format	O2
K5.20	The document conversion function must be easy to use.	O2
K5.21	It should be possible to convert documents individually and in batches.	O2
K5.22	The conversion should maintain the integrity of the document contents and preferably also the visual integrity of the documents.	O2
K5.23	A document which has been stored in an archival format, should be	O2

	locked in order to impede any further editing.	
--	--	--

5.3.3 Archival formats approved in Noark-4

Noark-4 poses the following requirements for a suitable archival format:

- The format should be openly documented.
- It should preferably be an ISO standard
- It must be supported by complete and well-established products in the market.
- Converting documents to the archival format should be easy.
- It should be possible to convert to the archival format in question from most commonly used production formats, even from graphics formats (graphics elements should be included with the text).
- It must be possible to convert the archival format into new formats later on.

The development of standardized document formats has not yet reached a "mature" stage, and probably will not do so in the foreseeable future. As far as richly formatted documents are concerned, we lack well-established and broadly useable ISO standards. Ambitions must therefore be lowered when selecting a format. It will be necessary to reconvert archival documents later on, possibly on a regular basis. This makes it all the more important to reach a complete decision as far as formats are concerned. Later conversions will then not have to start from an impenetrable multiplicity, but may be carried out in manageable batches.

Which archival formats may be used in public administration is decided on the basis of the Archives Act, and the authority is vested with the National Archivist. Special provisions with regard to formats are expected to be prepared on the basis of the Archives Act, which entered into force on 1.1.1999. For the time being, there is reason to believe that the following formats will be accepted:

K5.24	It should be possible to use four archival formats for long-term storage in state or local originating organizations and for transfer to archival repository: <ul style="list-style-type: none">• Text only: ISO Latin-1 8859-1:1987• SGML - ISO 8879:1986, <i>including subset formats HTML and XML</i>• TIFF, version 6• PDF	O2
-------	---	----

5.3.4 Text only - ISO Latin-1 8859-1:1987

ISO Latin-1 is not strictly a document format, but rather a character set. However, it is simple to handle, and it will most likely be available in the foreseeable future. ISO Latin-1 can represent 256 symbols. The first 128 characters are identical to what is normally called ASCII (American Standard Code for Information Interchange). This includes upper- and lowercase letters from A to Z, as well as the numbers (digits) from 0 to 9 and some special characters such as period, comma, colon and semicolon. A few formatting codes (of which

line feed and carriage return are the most important) are also defined. The letters Æ, Ø and Å as well as letters with accents are defined in the second half of the character set. In this second half, ISO Latin-1 differs significantly from what we normally know as ASCII.

Formatted documents which are converted into ISO Latin-1 will lose their structure and layout, and graphics cannot be represented in this format. However, the document contents - or text - are maintained. Thus, ISO Latin-1 is best suited to e-mail which is plain text, as well as notes and simple memos which are considered to be of archival value.

It is worth noting that ISO Latin-1 is also contained in other types of formats, such as SGML.

5.3.5 SGML (Standard Generalized Markup Language) - ISO 8879:1986

SGML was originally conceived as a standard for the printing industry. This standard defines the structure of a document and maintains all the structural editability of the document across platforms. Standard SGML does not, however, handle layout.

SGML is strictly speaking not a format but a syntax (in the form of "markups" or "tags") for defining a specific application. The export format for tables and attributes in Noark is based on the same syntax. The setup and meaning of the syntax used in an SGML document must be further defined in a separate document template - a DTD (Document Type Definition). The document layout may be described in a DSSSL (Document Style Semantics and Specification Language).

K5.25	During export/transfer of SGML documents, an associated DTD must be included, as well as any associated DSSSL.	O2
-------	--	----

The character set used in SGML should be ISO Latin-1. Graphics cannot be converted into SGML, and graphics elements cannot be built into the format. SGML may, however, contain references to separate files containing graphics, or to other external files. Noark should also be able to handle such complex documents. However, the remaining three archival formats selected for Noark may for the time entail considerable restrictions in terms of filing of complex documents.

HTML (Hyper Text Markup Language) may be regarded as an SGML application (a complete defined and fixed SGML code page). Thus, an HTML document needs no DTD. XML (Extensible Markup Language) is a subset of SGML ("SGML light"). Like SGML, it is a meta-language. Unlike HTML, it can be extended, and it allows the user to define a syntax. Even XML documents need a DTD. It may be an integral part of the document, in which case it constitutes the first part of it. It may also be included as a separate file, in which case it is handled and stored as for SGML documents.

For complex HTML and XML documents, Noark's requirements in terms of archival formats may impose the same restrictions as for SGML documents.

5.3.6 TIFF (Tagged Image File Format), version 6

TIFF is the most widespread and standardized raster-graphics format. It is the most widely used format for scanning paper documents. The format is portable across technology platforms and openly documented. It has recently been approved as an ISO standard (ISO 12639: 1997).

Raster graphics renders the document photo-identical, but it cannot be edited. There is, however, software which can convert from graphics to text format, so-called OCR software.

K5.26	TIFF files may be either "multiple page" or "single page". If <i>single page</i> is used, each single page of the document must be stored in a separate file in the same directory.	O2
-------	---	----

With graphics formats, the biggest drawback is the vast amount of space taken up by the documents - up to 10 times that of a text format. Graphics formats are therefore not suitable for e-mail exchange. On the other hand, it is quite common to compress TIFF files. Compressing in Noark should be "lossless", i.e., it should not result in a permanent qualitative deterioration of the document.

K5.27	For compressing TIFF files, only the following standards are approved: <ul style="list-style-type: none">• CCITT group 4 (for documents in black and white)• LZW(for documents with colours/shades of gray).	O2
-------	---	----

5.3.7 PDF (Portable Document Format)

PDF is a non-editable format (printing format) which handles both text and graphics. Everything that can be printed on a printer - including all the other archival formats - can be converted and stored as PDF files. Scanned documents (e.g., in TIFF format) may thus be converted into PDF, and the size is considerably reduced. The visual integrity of the documents is maintained after conversion. The format cannot be edited, but it is possible to re-use text by cutting and pasting, because pure text is used internally in the format.

PDF is portable across platforms and openly documented, but it is not an ISO standard, as it is controlled by the vendor, Adobe. Still, PDF appears to be the most versatile and useful archival format existing today. None of the other archival formats can compare to PDF when it comes to commercially available solutions and products.

K5.28	PDF files with text should be stored in a character-based form, not in binary form, in cases where it is possible to choose between the two options.	O2
K5.29	When PDF files are compressed, only "lossless" compression should be used: CCITT group 4 or LZW.	O2
K5.30	It is recommended that PDF documents be stored so that the fonts that are used, are embedded in the documents. The documents will then take up (sometimes considerably) more space, but this will ensure that the fonts that are used, are kept in typographically identical form across	A

	platforms and time.	
--	---------------------	--

5.3.8 Exchange formats

Documents which arrive by e-mail, are electrical from the start. The e-mail itself is transferred in a text-only format, and might enter straight into the document storage. In addition, other file types may be sent as attachments to an e-mail. It is presumed that electronic case documents are normally sent as such attachments to e-mail. These attachments may be in any format. In most cases, they will be word processor files. The format that is used in attachments to e-mail, is referred to here as *exchange format*.

There is also a need for standardized formats for exchange of electrical documents within public administration. It cannot be taken for granted that the addressee is able to read the format used by the sender. Noark does not have rules for regulating this, but the sender should always make sure the user is able to read the document he/she sends (cfr. ch. 10).

It should, nevertheless, be noted that there are several advantages if the exchange format is identical to the archival format of the addressee:

- The addressee will always be able to read the document and have it displayed in a version identical to the archival copy of the sender.
- The addressee may file the document as it was received, without having to convert it.
- A received document with a digital signature may be filed by the addressee with its signature intact. If, on the other hand, the addressee must convert the document to an archival format, the signature will be “broken” and no longer verifiable.

5.4 Export and transfer of documents

There must be functions for exporting documents from the electronic records to a system-independent format. Such export will make it easier to convert (move) the documents when a new system is implemented. Before reorganization takes place, all documents are to be removed from the database, exported in an *archival format*.

If a document consists of more than one file, all the files should be in the same directory. The system itself must handle references between the internal files in a complex document.

The standard medium for *transfer* to state or local archival repository is currently 74-minute CD-R discs (Compact Disc Recordable). The *file and directory structure* should conform to the ISO 9660 standard for maximum portability between technological platforms.

Document export must be carried out simultaneously with other export of selected attributes and tables from the system. The *Version* table represents the connection between the electronic documents and the rest of the system. Export from this table should contain the file name of the associated export documents.

The export format is described in more detail in the technical specifications in paragraph 15.3.4.

K5.31	It should be possible to export documents in archival format in the electronic records to individual files.	O2
K5.32	If a document consists of several files, all the files should be in the same directory.	O2
K5.33	DTD's and shared template files should be registered as documents in the tables <i>Document description</i> and <i>Version</i> .	O2
K5.34	The standard medium for <i>transfer</i> to state or local archival repository is currently CD-R disc (74 minutes). The file and directory structure should conform to the ISO 9660 Level 1. File and directory names should thus contain up to 8 characters. Valid characters are limited to A-Z, 0-9 and Underscore ("_"), all in upper case (capital) letters. File names may have a 3-character extension. The directory structure may consist of up to 8 levels. Alternatively, the file and directory structure may conform to the following: <ul style="list-style-type: none">• ISO 9660 Level 3, which have no length restrictions for file and directory names, character sets or the number of levels in the directory structure• Joliet (Microsoft's extended ISO 9660 specification)	O2
K5.35	It must be possible to link the exported documents to other data exported from the system.	O2

Special comment on K5.34:

The restrictions of ISO 9660 Level 1 with regard to character set and the length of directory and file names (maximum 8 characters) are in practice bound to force through machine-generated names containing combinations of numbers. ISO 9660 Level 3 may be used if this is problematical or undesirable. However, ISO 9660 Level 3 is rarely used. To escape the restrictions of Level 1, people often resort to platform-specific extensions of ISO 9660 such as "Rock Ridge", "CD-XA" or Apple's "HFS". These are not compatible with ISO 9660 and are therefore not authorized as logical file structures for Noark transfer to a depot institution.

However, Joliet, Microsoft's extension to ISO 9660 Level 1, is of a different character, since it uses two parallel filenames for each file and directory: an 8-character name conforming to the specifications of ISO 9660 Level 1, and a Joliet-specific name with up to 64 characters based on a Unicode character set (ISO 10646). Only Windows 95/98 and Windows NT 4.0/5.0 systems currently have access to "long" Joliet filenames, but all other ISO 9660-compatible systems will see Joliet's "short" filenames - and only these. Other systems will simply ignore Joliet's file system no. 2.

Pending the new ISO standard which will replace ISO 9660, Joliet may be accepted as a file-structure format for Noark transfers.

5.5 Procedure requirements

The introduction of electronic document storage instead of paper-based records will require major changes in daily routines for both registry personnel and executive officers. Many procedures must be changed, and careful planning is required before the new system is implemented. Many tasks which currently require a lot of time, will disappear. This applies, among other things, to all work relating to physical filing, retrieval, shelving, distribution and lending-out of cases and documents to executive officers and managers. In addition, new tasks seem to belong naturally to the registry. This means different qualifications are required from registry personnel.

The registry must be responsible for controlling the quality of the electronic records, for instance by checking that the correct document is linked to the correct version in Noark. The registry should be responsible for handling the official e-mailbox of the organization. All scanning of incoming documents should also be assigned to the registry. This also applies to converting into archival format, even if it is possible to let executive officers do this directly.

Even IT personnel will feel the effects strongly. They will be responsible for making sure the electronic records are in order at any time, and for the security in terms of daily backup, etc.

5.5.1 *What archive formats should be selected?*

The organizations must make a choice in terms of which of the four formats should be used for long-term storage of electronic documents. It is possible to use a mixture of formats. Incoming scanned documents may, for instance, be stored in TIFF format, some e-mail may be stored in ISO Latin-1, outgoing documents may be produced directly in SGML or converted from a word processor format into PDF. Alternatively, one may choose to store everything in one format only. In that case, the options left will be PDF and TIFF, since graphics (scanned documents) cannot be converted into the other formats. Standardization in terms of using only one format has its advantages. Converting batches of documents into other formats will, for instance, be much simpler.

5.5.2 *Gradual implementation of electronic recordkeeping*

It is possible to implement electronic recordkeeping gradually. For many, the easiest thing is to start by storing outgoing documents electronically. Many already do this today, using systems based on Noark-3.

A situation with a mixture of electronic and paper-based documents normally requires that all main documents within a case are stored either electronically or on paper. Unless all main documents within a case are electronic (e.g., only the outgoing documents are), the electronic documents of the case are merely to be regarded as work copies - not archival copies. In such a situation, all documents which have been created electronically, must be printed on paper, and these printouts must be stored in traditional records together with the original paper documents. During transfer to archival repository, the paper-based records

should be transferred. The electronic documents may be transferred in addition if this is desirable.

Noark permits *attachments* to be stored on paper even for electronically stored cases (see requirement K5.11). It may, for instance, be deemed inappropriate to scan such attachments as thick reports and comprehensive technical sketches; these may be stored in paper form. According to Noark, a case may still be regarded as electronic if the front page of the *main document* is scanned or a reference document is used as main document and the entire document is stored on paper (see paragraph 5.2.3 and requirement K5.12).

The requirement that all main documents be stored electronically for the case as such to be regarded as electronic (possibly by scanning parts of the main document) means, in practice, that consistently electronic filing requires extensive *scanning* of incoming paper documents, at least of incoming main documents.

If there is a mixture of electronic and paper-based cases, they belong naturally to separate *records sections*.

5.5.3 Procedures regarding filing of documents

- *Incoming paper documents*: Most documents received by the organization will for many years still be paper documents sent as mail. Incoming paper documents must first be scanned in order to be stored electronically. During scanning, an electronic snapshot of the document is taken, and this snapshot is saved in a raster graphics format. One may decide what format to use. If TIFF is selected as archival format, the natural thing would be to transfer the document directly to this format. During scanning, one may also select picture quality - e.g., resolution. The higher the resolution, the more space is required for the document. Quality must thus be weighted against space requirements.

The organization must decide who is responsible for the scanning of paper documents. The scanned documents are to be linked to registry entries and cases in the Noark system. It therefore seems natural that the registry should be responsible for the task of scanning documents.

- *Incoming e-mail*: Electronic document exchange is more and more used, and public organizations have already started exchanging official documents using e-mail. Procedures regarding e-mail are described in more detail in chapter 10, which also describes how information from the Noark system of the sender may be exchanged as a separate attachment ("Noark head"). This information may go straight into the system of the addressee while at the same time the transferred document is associated with the pertinent registry entry and case.

The organization should maintain an official e-mailbox, and the task of handling this should be assigned to the registry. If e-mail is sent straight to the executive officers, they should follow the procedures described in chapter 6.

Incoming e-mail may enter the electronic records directly as soon as they have been associated with the pertinent case. If an exchange format has been used that is identical to the archival format of the organization, no conversion is necessary.

- *Internally produced documents*: The vast majority of outgoing documents will be produced by executive officers using text editors. The same applies to internal

documents. Executive officers will normally make sure the documents are associated with the pertinent case and registry entry. However, it is necessary that the registry carry out subsequent quality control. The document production should be as closely integrated with the registry system as possible, so that, for instance, information from the case and registry entry can be easily merged straight into the document - or vice versa. What executive officers are permitted to do at any stage of the process, is described in chapter 6. It is possible to formulate procedures where the executive officers themselves convert documents into archival format, but even so, the registry needs to carry out quality control.

5.5.4 Converting into archival format

The conversion into archival format may be carried out at different times. It is recommended to do it as soon as possible after the document has been received (incoming documents) or produced (outgoing and internal documents). However, it should be possible to convert several documents in one operation (batch job) at a later stage.

When a document is converted into archival format, the production format may be deleted, unless regulations require that both formats be preserved.

5.6 Essential tables in the module

Only essential tables are included here. For a complete overview of the tables in this module and their attributes, see part II, Technical specifications, chapter 14.

Table name	Text
Document link	Establishes a relationship between the registry entry and the case document (document description). Also contains information on the type of document (main document, attachment, etc.) and the sequence of the documents within the registry entry. See also paragraph 14.3.
Document link meeting document	Establishes relationships between a meeting document and the individual documents which are included here. Also contains information on the type of document that is involved (draft, minutes, etc.) and the sequence of the documents within the meeting document. See also paragraph 14.6.
Document description	Contains information on the case document, such as category (letter, report, circular, etc.) and title (heading). Also refers to the physical location if the document is stored on paper. The attribute <i>Document status</i> indicates how far the document has come in its life cycle. The tables also contain attributes for managing the access to the documents (<i>Access code</i> , <i>Access group</i> , <i>Authority</i> , <i>Date of downgrading</i> , <i>Downgrading code</i>). See also paragraph 14.3.
Version	Contains information on the specific version of the document, as well as what versions and formats the document has been stored in ("variant" is used technically to embrace both these categories). The

Table name	Text
	most common formats are production format and archival format. The pertinent version variants are: a publically available version of the document and a digitally signed version, for instance in PEM form (Personal Enhanced Message). See also paragraph 14.3.
Topic word	Contains topic (subect) words and keywords for the specific document. See also paragraph 14.3.
Note	Contains notes on a case, a registry entry or a version of a case document. This table is also described in connection with the records management module. See also paragraph 14.2.
Additional information	Contains various kinds of additional information on a case, a registry entry or a version of a case document. This table is also described in connection with the records management module. See also paragraph 14.2.
Digital signature	Contains one or more digital signatures (stored in binary form) and is linked with versions of documents. Digital signatures may be used to guarantee the authenticity and integrity of e-mail and filed case documents. Information on the verification itself is also stored here. See also paragraph 14.3.

5.7 Changes from Noark-3 and Koark

In Noark-3 and Koark, electronic recordkeeping is for all practical purposes treated at a general level. Most of the specifications in this chapter must therefore be regarded as new. Still, the following differences in detailed specifications should be noted:

- In Noark-3 and Koark, the relationship between registry entry and electronic document is 1:1. In Noark-4, it is M:M.
- Noark-3 and Koark require that the serial number should refer to electronic documents. In Noark-4, the structure of the document reference is left to the system developer.