Guidelines on best practices for using electronic information

How to deal with machine-readable data and electronic documents

Updated and enlarged edition

DLM-FORUM
Electronic Records

These multidisciplinary guidelines were initially conceived by Jean-Michel Cornu, consultant, in close cooperation with the Historical archives of the European Commission and experts from the Member States. A draft version was edited and distributed to the participants of the DLM-Forum on electronic records (Brussels, 18 to 20 December 1996): ISBN 92-827-9129-7 (EN, FR, DE).
It has been updated by Sylis (Lille, France) incorporating comments and annotations received during the forum and forwarded shortly afterwards to the Historical archives of the European Commission.

This updated and enlarged edition of the guidelines is one of the concrete results of the DLM-Forum. It illustrates the continuous cooperative efforts made by different disciplines (public administration, archives, industry and research) involved in the issue of electronic data and information management in the rapidly evolving information society.

A great deal of additional information on the European Union is available on the Internet. It can be accessed through the Europa server (http://europa.eu.int).

Cataloguing data can be found at the end of this publication.

Luxembourg: Office for Official Publications of the European Communities, 1997

ISBN 92-828-2285-0

Printed in Italy

Table of contents

PREFACE

The DLM Forum, organised jointly by the Member States of the European Union and the European Commission in Brussels in December 1996, brought together experts from industry, research, administration and archives to discuss a topic of ever increasing importance: the memory of the information society. Just a few years ago hardly anyone could have imagined that the archival activities would be so revolutionised by technological innovations that, lacking timely and appropriate measures, the authenticity and long-term conservation of vital information would be seriously endangered in the near future.

Archives services are an essential component of the information society. Since the report *Europe and the global information society* (the Bangemann report) was published in 1994, they have become increasingly prominent in the Commission's action and support programmes. Archives services will moreover play an increasingly important role in modern information management. Against this background the DLM Forum gave a clear signal and set the ball rolling.

The *Guidelines on best practices for using electronic information* are one of the principal outcomes of the DLM Forum. A preliminary draft of the guidelines was distributed to the participants as a basis for discussion. Following the more than 300 generally substantial proposals for amendments and improvements which were made to the European Commission's services by national experts during and after the forum, this new, much revised and expanded edition of the guidelines was produced.

I am convinced that these multidisciplinary and jointly developed guidelines will help to define the short and medium-term strategies needed to solve the pressing problems of managing and conserving electronically stored data. Decision-makers in administration and industry are called upon to agree the necessary measures with experts on electronic archiving and develop possible practical solutions, working together on a pan-Europe scale. This would go a long way towards increasing the confidence of companies and private individuals in the reliability of data stored on the new media.

The structure of the guidelines makes them a rich source of information for experts and laypersons alike. They represent an important step forward both in ensuring the lasting conservation of the memory of the information society and in increasing the transparency of the activities and decision-making processes in the administrations of the Member States and the institutions of the European Union for all its citizens.

Dr Martin Bangemann
Member of the European Commission, responsible for
industrial affairs, information and telecommunications
technologies

1. Introduction

1.1. About the multidisciplinary guidelines

The great majority of documents nowadays are still in paper form. But many of them – whether electronic messages, memos, or minutes of the meetings which punctuate our working lives – are produced using computer equipment. The rapid spread of electronic mail has brought large increases in the number of electronic documents in both government and industry. As a result the digital form has even begun to replace paper. There is, therefore, a growing need to consider the impact of this phenomenon on current practice and rules for document use and archiving.

Documents and databases
increasing use
of electronic media

Many organisations have also developed their own databases, often containing information that has to be kept for some time after it has served its immediate purpose, either for legal reasons or for future research. So, thought also needs to be given to the long-term storage and accessibility of information and its potential to produce new information.

These multidisciplinary guidelines cannot claim to give comprehensive or definitive answers to every question. There will have to be further discussion between different disciplines concerned, i.e. public administration, archives services, industry and the research community. However, the guidelines give examples of best practice and offer advice to help individual organisations define their own strategy for electronic information. Their purpose is not to present a single, Europe-wide approach to the subject, but to pool the experience acquired by national, regional and European organisations for the benefit of all.

Examples
of best practice
to help define a strategy
for electronic information

A draft version of these multidisciplinary guidelines was distributed by the European Commission to the participants of the DLM-Forum on electronic records, which took place in Brussels from 18 to 20 December 1996. As a follow-up, written comments and annotations were received from national experts and further discussions held for the preparation of this updated and enlarged edition.

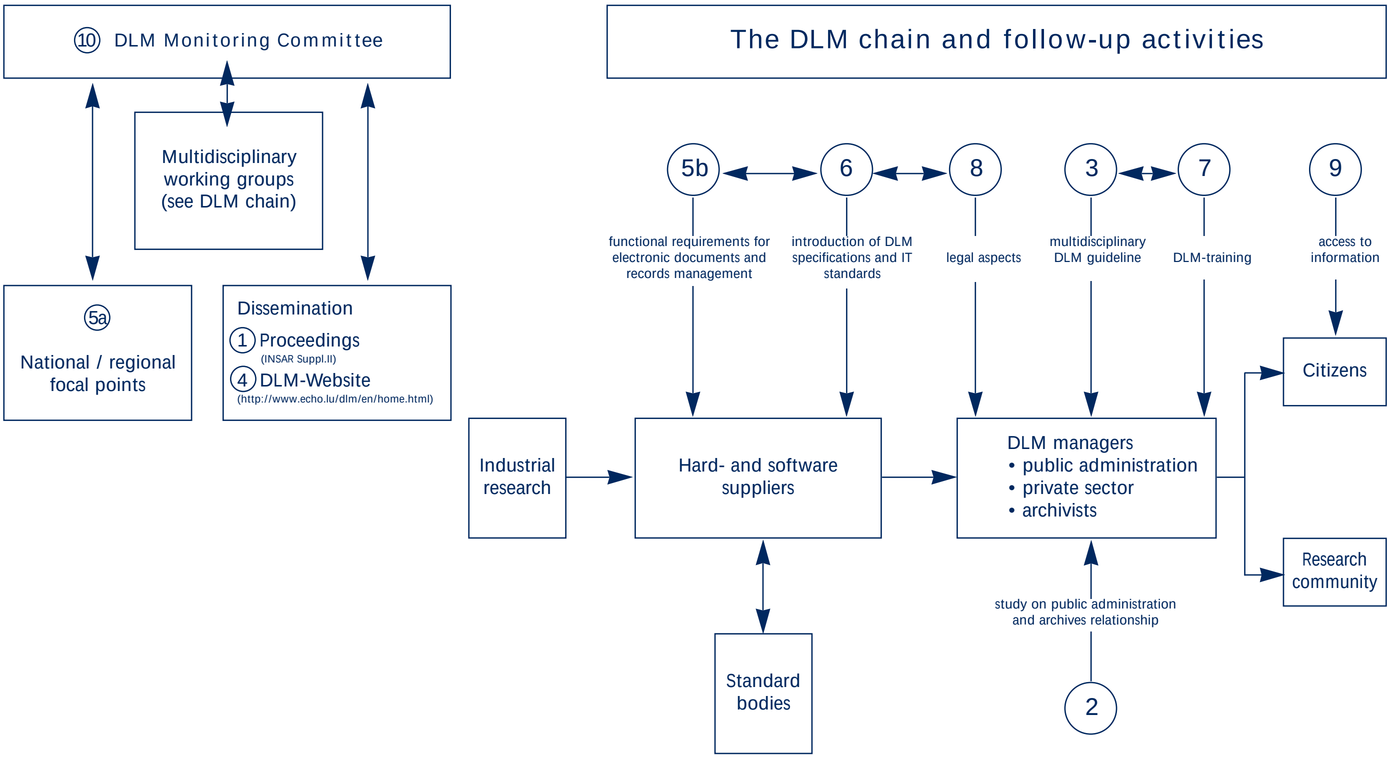
They can also be used in conjunction with the International Council on Archives' (ICA) guide on electronic records which deals with electronic records from the point of view of a single discipline.

The follow-up actions of the DLM-Forum which were fixed as the so-called 'Ten Points'([1]) (see circles in the following charts) are related to the whole chain of production and maintenance of electronic records. They are monitored by the DLM-Monitoring Committee which was set up as one of the follow-up measures. These actions include:

- support to the users with training and these multidisciplinary guidelines;
- a comprehensive study on relationships between public administration and archives;
- specific indications on functional requirements on DLM specifications, on IT standards to hard- and software suppliers and standard bodies, on legal aspects to DLM managers and on access to information for the citizen and the research community.

The DLM chain
includes several types
of actions:
- specific indications
- studies
- support to users

([1]) 'Proceedings of the DLM-Forum on electronic records, Brussels 18 to 20 December 1996', *INSAR — European archives news*, Supplement II, EUR-OP, Luxembourg, 1997, p. 353 and on Internet: http://www.echo.lu/dlm/en/home.html


10 DLM Monitoring Committee
Multidisciplinary working groups (see DLM chain)
5a National / regional focal points
Dissemination
1 Proceedings (INSAR Suppl.II)
4 DLM-Website (http://www.echo.lu/dlm/en/home.html)
The DLM chain and follow-up activities
5b
6
8
3
7
9
functional requirements for electronic documents and records management
introduction of DLM specifications and IT standards
legal aspects
multidisciplinary DLM guideline
DLM-training
access to information
Citizens
Industrial research
Hard- and software suppliers
DLM managers
• public administration
• private sector
• archivists
Research community
Standard bodies
study on public administration and archives relationship
2

A multidisciplinary approach taking into account the needs of all parties concerned

These multidisciplinary guidelines are aimed at people having only a basic knowledge in electronic records management as well as those with a more advanced level, working in a wide range of fields, especially those in:

- public administration;
- archives services;
- industry (software and hardware suppliers);
- the research community;

in the Member States and institutions of the European Union.

1.2. Using the multidisciplinary guidelines

The main body of the text is aimed at the general reader.

To make it easier for readers to find the right information for their needs and level of knowledge, further material is set out in four different kinds of colour-coded boxes.

CD-ROM

Basic concepts

(Boxes describing basic concepts)

Basic concepts
These boxes briefly review basic concepts used in different fields

Graphics files

Advanced topics

(Boxes describing advanced topics)

Advanced topics
These boxes give more technical explanations for those who want to go into detail

Germany

Example

(Boxes giving examples)

Example
Examples give some idea of the kinds of solutions adopted

Scanning a document

Options

(Boxes giving advice to help you make choices)

Options
These boxes offer advice or decision trees to help make the right choice

Standards
Not all standards have equal force

Not all the standards cited in this guideline have equal force. The traffic lights provide a quick and easy way to identify how far you can rely on them.

○ ○ ● ‘Green’ — stable, recognised standards;

○ ● ○ ‘Amber’ — standards pending or used by only a few suppliers;

● ○ ○ ‘Red’ — proprietary standards which cannot be guaranteed to last.

1.3. From individual production to general use

A document or database
created by an individual at work is one component of an organisation’s overall information system

Why is it so important to have a long-term strategy when creating, updating or supplying information?

Even though documents or databases are created by one or more individuals, they will often be of interest to many more people than might first be supposed.

- They may be used or updated by someone who played no part in their creation.
- They will often be used in conjunction with many other records, some of which were not even known to the original author.
- The information they contain may be used long after they cease to be in current use (for example, for legal or historical purposes).

A document or database should therefore be conceived as one component of a wider information system. It is vital that preservation, access to information and the protection of privacy be taken into account from the outset.

1.4. Three stages in the information life cycle

Three stages:
- design
- creation
- maintenance

There are three main stages in the life of electronic information.

Design
This is when an overall strategy is laid down.

Creation
This is when the data are actually created, normally by a limited number of people.

Maintenance
This includes the use and preservation of the data.

However, the three stages cannot all be approached in the same way. There are differences as regards:

- updating data
- the number of times they are accessed
- responsibility
- etc.

2. From data to structured electronic information

2.1. What is information?

A piece of information is an indication or an event brought to the knowledge of a person or a group. Information may be created, maintained, preserved or transmitted.

Information is the base of the organisation of business processes. This concept becomes so important that one now talks about the information society as a new step after the industrial society.

Information
is an indication or an event brought to the knowledge of a person or a group

2.2. What are data?

Data are basic units of information.

In a document, for example, many items of data are assembled to present an argument or describe an activity. Until recently, most data were preserved and transmitted on paper (or, in more ancient times, on other media such as stone).

Data
can be assembled to form a document or a list

Sometimes we find data in the form of lists, as in a telephone directory, where the aim is not to present an argument but to provide the raw material for a future action (finding Mr Brown's telephone number, for example). In this instance it is crucial that the data be classified in such a way as to make searching easy (the names are listed in alphabetical order for each town or area).

It is possible to store data on media other than paper. For example, information may be stored electronically for easier processing.

The speed of technological progress is making it more and more difficult to ensure lasting solutions when preserving data. File formats and electronic media are evolving rapidly and have a much shorter life expectancy than paper.

Data management
is an increasingly complex business

Also the volume of information being produced has grown enormously, especially in government and public administration. As a result, it is becoming an increasingly complex matter to classify and structure data so that they can be accessed long after their production.

2.3. What makes electronic information so different?

When data are stored on an electronic medium, they can no longer be read without using a specific tool, i.e. a machine (generally a computer).

The medium is not the message as it is with paper documents. In these multidisciplinary guidelines, electronic information is understood to mean data stored in a format which allows them to be processed automatically, generally speaking, on some kind of electronic medium.

The medium is not the message
You need a tool in order to read electronic information

Electronic information brings benefits and constraints for data management

There are several benefits to be gained from storing data on a medium which can be read directly by a machine.

- It is much easier to process the data by machine, and changing just a few items does not involve re-entering the data all over again.
- Electronic media generally allow more data to be stored in a smaller space.
- It is easier to copy a complete record.
- It is easier and quicker to transfer information from one place to another.
- It allows more elaborate use by using an electronic processed structure.

However, the use of electronic media also involves new constraints.

- Humans need a tool to read the data.
- Electronic media generally have a shorter lifespan than paper or microfilm.
- It is easier to duplicate or alter an original (which raises problems of proof and authentication).
- The rapid pace of change in technology and on the information market makes it difficult to find stable and long-lasting formats to use.

2.4. Creating electronic records

Data are stored on a medium. In the case of electronic information the data can be processed, communicated and interpreted by computer. A set of data may sometimes constitute a record.

What is a record ? International Council on Archives definition

The International Council on Archives (ICA) Committee on electronic records defines a record as 'a specific piece of recorded information generated, collected or received in the initiation, conduct or completion of an activity and that comprises sufficient content, context and structure to provide proof or evidence of that activity'.

In public administration a record has legal value as information or proof. The term 'record' is understood here in its administrative and archival sense, not in the technical sense usually used in computing.

The fact that an electronic record is distinct from the medium on which it is stored has a number of consequences. For example, it is easy to copy an electronic record from one medium to another. This makes it easier to copy and disseminate information, but it also makes it more difficult — though no less crucial — to define what is meant by an original (see in particular Annex 8.2 — Issues list).

Data stored in an electronic record must form a coherent and consistent set of information. Drawing up a classification system is one of the most important tasks when defining records. There are often several possible ways of grouping data in a record depending on the level of detail you want. With a database, for example, either the entire base or a coherent subset of it may form a single record.

Electronic records

Advanced topics

1. An electronic record consists of four main elements. The first three must be preserved.

2. The content of the record can include several types of data:

   - text (pages, paragraphs, words);
   - numbers (integers, floating-point numbers);
   - tables (complete tables or cells);
   - pictures, graphics, sound and video;
   - hypertext links.

3. The logical structure of the record may be incorporated into the document or database itself or it may be separate, in which case the same structure can be used for several records. The logical structure may be very different from the physical structure of the record.

4. The context is described in an associated document. This can include:

- the technical metadata (hardware and software environment — including version numbers, file structure, a description of the data and a history of links with other records);
- a description of the administrative context involved.

The context described in the documentation may be extremely complex if the record is integrated in a network architecture.

5. Presentation (in particular for documents) is increasingly treated separately from the record itself, so that the information is independent of how it is to be presented. Dissemination of information on different media (CD-ROM, on-line access, paper, etc.) is called cross-media management. A tool which will allow us to display the data recorded today in a few years time has yet to be invented.

Electronic records consist of four elements:

- content
- structure
- context
- presentation

One may not preserve the last of these, since it depends largely on the medium used to display it

2.5. Two ways of structuring data

To be able to find a specific item of information, it has to be structured. Depending on the information's purpose there are two main ways of structuring data.

- Databases: data are placed in a 'pool' of information from which they can be retrieved and updated.
- Documents: this structure is used when data are arranged in an ordered fashion to present an argument or describe an activity. A document may often serve as evidence of a particular activity (e.g. a legal instrument). In this case it needs to be captured as a record.

Databases and documents

Two types of data structure which are increasingly mixed to form a compound set of information

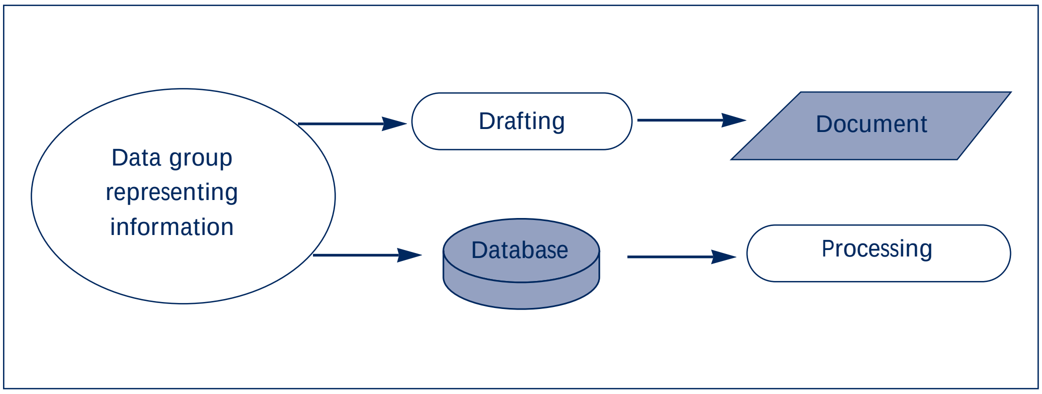


Figure 1 — Document or database?

Of course there are many different ways of structuring data. As information comes to play an increasingly important role in organisations, the structure of both documents and databases is becoming more and more complex (e.g. a page on-line or on CD-ROM that is automatically filled with data retrieved from a database).

In addition, there is often a close link between a database and a number of documents. Updating the data in a base, for example, may entail the production of a document defined by the administrative procedure. This kind of approach is increasingly common with 'workflow' tools, which regulate the flow of process within a firm.

There are a number of differences in approach depending on whether documents or database are involved. For example:

- the preservation of each type (databases and documents) involves different requirements;
- a document must be consistent and remain the same (updating produces a wholly new document);
- a database, by contrast, can be updated regularly;
- given these two contrasting approaches, composite information (a combination of database and document) raises completely new questions.

2.5.1. Documents

There are several stages in the transition from an office based on paper documents to an office based on electronic documents:

Paper and electronic documents coexist. The particular features of each type must be taken into account

1. the traditional office based on paper documents;
2. a mixed office with documents on paper and in electronic form;
3. the conversion of paper documents (by scanning);
4. an entirely electronic office where all documents are produced, received and distributed electronically.

These guidelines deal mainly with stages 2 and 3 as they represent the current situation of offices. They cover the relationship between electronic documents and traditional documents (mainly on paper and microfiche). Stage 4 is the direction in which we may be moving in the future rather than a stage in itself. Paper documents and electronic documents will probably continue to coexist for many years.

Documents come in many different forms — letters, notes, memos, forms, reports, etc. — each requiring special treatment.

Documents may be grouped together in files to form a coherent unit of information. The classification of documents is an important factor for finding information easily.

Classification of documents

Basic concepts

There are several ways to classify documents, the two main ones being:

- chronologically (each document is given a serial number when it is recorded);
- by subject matter (each document is assigned a number, for instance, in accordance with a particular classification plan).

The second of these is often the most effective for finding a record easily. But an efficient classification plan has to be defined first.

Another solution is to find information with the help of specific keywords or to search in the text itself.

The relation between the classification plan and the keyword system is similar to that between a table of contents and an index in a book. Both are efficient search tools, but they are not mutually exclusive.

Modern information technologies may add new ways of accessibility and classification.

Classification plan and keywords
Two complementary tools to help you search for information

2.5.2. Databases

Databases can pose a problem when it comes to accessing data a long time after the normal lifetime of the database (e.g. for legal or research purposes). There are very few standard database formats at present. In many cases there are only two options:

- copying the database to a lower-level format (e.g. as plain text or in indexed sequential access method (ISAM)) format;
- keeping the application program which generated the database including the documentation (database management system, accounting application, etc.).

Preserving a database is complicated by the lack of standard formats to guarantee long-term accessibility

The first solution may mean that some structural elements of the base are lost. The second will often mean keeping not just the application but also a computer system that it can run on, as well as maintaining a working knowledge of the software and hardware (not easy after several years).

In many cases the database is fully integrated in a proprietary application. This is often the case with management programs.

2.5.3. Procurement of hardware and software for electronic information

Procurement of hardware, software and services is an important aspect of the usage and archiving of electronic data. Many hardware and software products and services exist today, based on a variety of technologies, as well as standards and specifications. However, with the rapid advancement of information and communications technologies, there is a real concern that the long-term survivability of electronic data is unduly dependent on the format in which the data are kept, which may not be supported by future information technology (IT) products.

EPHOS
is the European Union program which aims to help purchasers in the public sector

In the European Union, public sector procurement is governed by the Council of Ministers' Decision 87/95/EEC, which makes it mandatory for all public purchasers to reference *de jure* standards (i.e. standards approved by the formal standardisation bodies) in their tendering procedures. EPHOS (European procurement handbook for open systems) is a European Union program which aims to help purchasers in the public sector to implement Council Decision 87/95/EEC, by providing a series of handbooks which contain strategic procurement advice on open systems technology. In particular, EPHOS references international standardised profiles of international standards where applicable. Current EPHOS modules deal exclusively with *de jure* standards, although some modules which are under preparation, address the use of publicly available specifications.

Procurement with ***de jure*** standards
poses practical problems due to unavailability or cost of products and competing *de facto* standards

However, experience has shown that in some of the information and communications areas, especially areas in the information technology sector, the provisions of Council Decision 87/95/EEC pose practical problems. This is due to:

- unavailability of products which support *de jure* standards;
- high cost of products which support these standards;
- the danger of these standards becoming obsolete in future and therefore not supported by future products;
- competing *de facto* standards (which may be publicly available specifications) which are widely implemented in products, with the products themselves widely deployed by users (e.g. Internet specifications).

It should, however, be noted that the problem of future proofing applies equally — and some experts would argue more particularly — to the deployment of products based on *de facto* standards and to products based on *de jure* standards.

Procurement of products
based on stable and open standards is a major criterion for evaluating products

The above analysis suggests, therefore, that public procurement should be based on products, rather than standards. Users who produce or maintain machine readable data must ensure that they have all the hardware, software and documentation they need to recover the data and documents produced by their applications in the long term. Nevertheless, it is recognised that it is not always a practical or realistic option to preserve all the hardware, software and documentation over a long period of time. Clearly, the procurement of products which are based on stable and open standards is a major criterion for evaluating products.

Specifically, users need to have a clear view of the standards that the IT products support, including an assessment of the stability and openness of these standards. This points to the requirement of a long-term procurement policy. It may be that a set of common procurement guidelines should be developed for the community of interests involved in the preservation of machine-readable data.

Documentation on formats
It is a good idea to make suppliers provide you with everything that is needed to recover data generated by their applications in order to facilitate the long-term preservation and use of the data

Procurement clause

Options

In the interim, it is suggested that the following standard clause be included in calls for tenders.

Standard clause for inclusion in calls for tender

'To guarantee preservation of the awarding organisation's data and long-term access, the supplier undertakes to provide all the hardware, software and documentation necessary to recover data generated by his applications and to export them to other environments and other formats.'

3. Information life cycle and allocation of responsibilities

3.1. Overview of the life cycle

Electronic information goes through several stages during the life cycle. Various tasks can be defined for the different stages. It is important to lay down rules and standards to ensure coordination between everyone involved in these tasks.

Life cycle
A set of stages which must be coordinated as far as possible

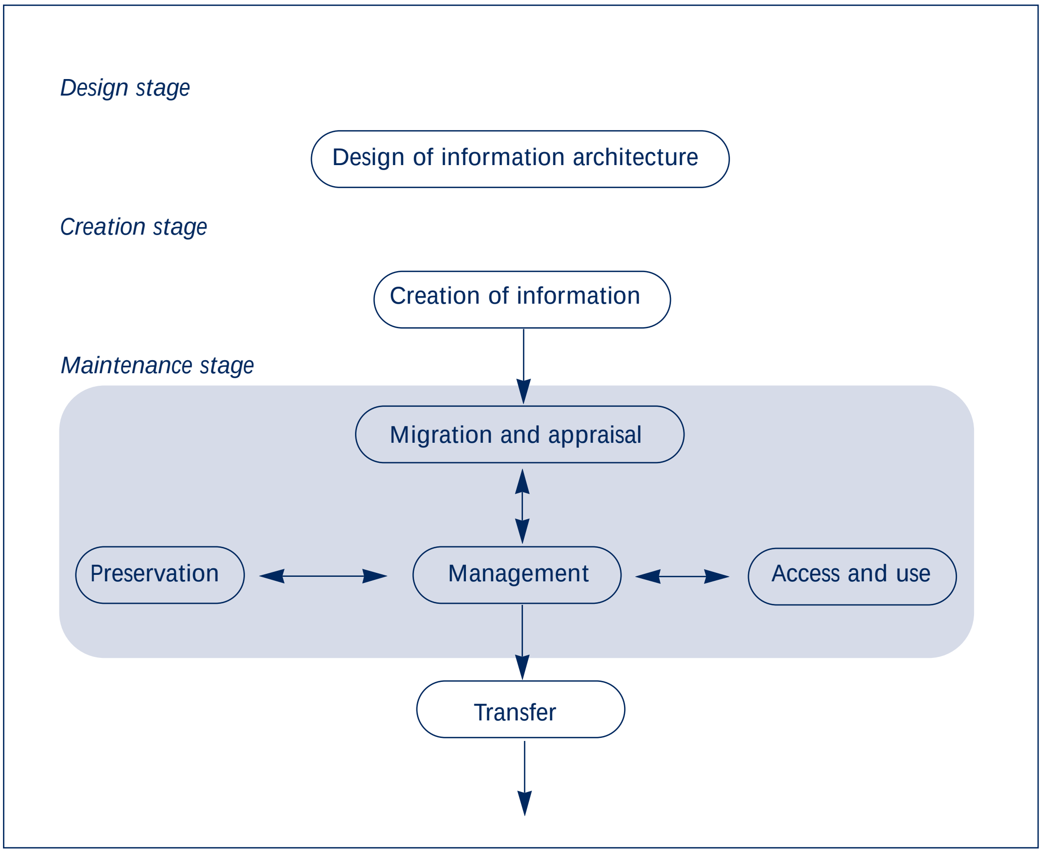


Figure 2 — Life cycle of electronic information

There may be several maintenance stages where responsibility for electronic information is transferred from one organisation to another.

Chapter 4 describes the tasks involved in creating and managing electronic information and gives some suggestions for defining a coherent strategy.

Chapter 5 contains some helpful suggestions for the preservation of electronic information (in particular, using different types of media and file formats).

Chapter 6 discusses several different ways of organising access to and use of electronic information, taking into account organisational and security aspects and data exchange standards.

3.2. Defining responsibilities for each stage

Defining responsibilities
Whatever choices an organisation makes, the responsibilities for each stage must be clearly defined

Documents and databases are vitally important for a public administration, as for any organisation. Information is the fuel on which organisations run. One of the major factors determining their success is their ability to define short, medium and long-term strategies for processing, keeping and accessing information.

The way in which responsibilities are assigned differs from one country to another and from one organisation to another. Whatever options are chosen, it is vital for responsibilities to be clearly defined as part of an overall strategy, rather than by chance or as a result of the traditional links between departments.

Allocation of responsibilities

Example

National and Community administrations have made different choices regarding the way in which responsibilities are allocated.

- Unesco has recommended that one should ‘use the expertise of the archivist in his or her ability to appraise information but use the source of information as the physical custodian of the record’ (Unesco/RAMP study).
- In the USA, records are kept by the National Archives and Records Administration (NARA) in more than 20 facilities, located throughout the United States, as well as ‘affiliated archives’ authorised by NARA. These records are still under the responsibilities of NARA, while the affiliated archives are responsible for their preservation, management, and access.
- In the European Commission the archives services/registries are responsible for providing distribution slips for the mail service in some Directorates-General, while other Directorates-General organise their affairs differently.

Allocating responsibility efficiently means taking into account the nature of each task as well as the culture and know-how of the organisation. The departments concerned and the archivists should be involved from the outset. Market information and research should also be taken on board.

When electronic information is produced and maintained by several different organisations, responsibilities should be identified through a dialogue between the organisations and their archives services.

What is the role of archivists?
They have valuable record appraisal skills and can offer useful input from the very beginning of the record life cycle

The archivist is responsible for record-keeping. He also has valuable skills to offer in assessing the value of a record. The archivist’s role is changing from the passive reception of records that have reached the end of the active phase of the life cycle to active involvement from the very start.

There is a need for a closer cooperation between archivists and persons involved in electronic information in public administration and in the private sector.

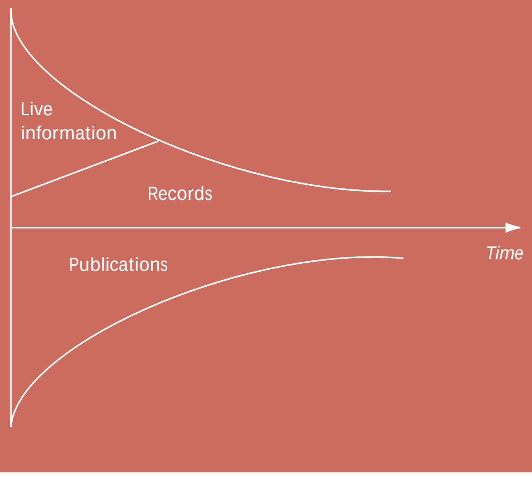
Three types of information

Advanced topics

There are three types of information.

- The live information is produced in the organisation as non-registered information. It can be ephemeral or can be registered to become a record
- The records: archivists are the experts to help in managing records during the entire life cycle. Information can be produced as a record, or live information can be registered. The archivists say ‘to capture a record’. In this case, the information is ‘frozen’ and the context becomes a key issue to facilitate access and use of the records.
- The publications: librarians are the experts to deal with this kind of information.

While the present multidisciplinary guidelines present electronic information and records from the point of view of the business process involving different disciplines, the ICA guide on electronic records gives more details (see Annex 8.9) on record-keeping from the point of view of archivists.



Various kinds of information need various expertise:

- the live information is often managed by the author of the document or the database manager
- the archivists have an expertise in records management
- the librarians have an expertise in publications management

4. Design, creation and maintenance of electronic information

Each of the tasks involved in the life cycle has it own characteristics. This chapter deals with the design, creation and maintenance of electronic information.

4.1. The design stage

Strategic team ▶ The best results are obtained by setting up a multidisciplinary team at the outset to define and monitor electronic information strategy

The best way to optimise the management of electronic information is to define a coherent global strategy at the outset, ensuring that everyone concerned is involved. One solution is to set up a multidisciplinary team to define and monitor the strategy. Although not very widespread, this approach can probably be regarded as best practice.

Defining a strategy

Options

The responsibilities of the multidisciplinary electronic information strategy team could include the following:

- taking account of users' requirements (including legal aspects);
- identifying important records;
- defining rules for an efficient classification plan;
- defining standards and specifications to ensure the independence of data from media and to guarantee their durability;
- defining an appraisal scheme;
- identifying those responsible for each task in the life cycle;
- defining a training and awareness policy in the organisation(s) concerned ([1]);
- monitoring the implementation of new systems.

(See also in Annex 8.4 the checklist for electronic information strategy)

Legal aspects ▶ are one of the key issues to be borne in mind in any electronic information strategy

In the case of a public authority, the legal aspects are among the most important issues to be borne in mind. But some questions regarding electronic information are still relatively unresearched and so great care needs to be taken (see the list of outstanding issues in Annex 8.2).

4.2. Creating electronic information

It is important to consider each of the tasks involved in the life cycle as early as possible. Common rules should therefore be applied when creating electronic information. This will facilitate communication between those responsible for processing it later on.

[1] Training for archivists and public service officials should include coverage of electronic information and electronic records.

Creating electronic information

Options

The following rules apply when creating electronic information.

- Each document or database must be clearly identified by the organisation responsible for electronic information management.
- Background documentation should be provided for all electronic information. It is then kept by the person responsible for electronic information management at each stage in the life cycle.
- Procedures should be established to process electronic information of uncertain origin.
- No data should be destroyed or changed without an approval procedure. This ensures that the context is preserved. (Destruction or updating includes any operation whereby the ability to combine, recognise, retrieve or identify data is lost.)

Rules for creating electronic information Observing certain rules will make matters easier at later stages in the life cycle

Information may be produced on paper initially and then digitised or it may be created in digital form straightaway (word processor, database program, electronic mail, etc.). In all these instances, it should be produced in a standard format or converted to one (unless authenticity is endangered). Chapter 5.2 gives details of the most suitable formats.

Electronic information or paper documents?

Options

- Usually information on paper is preserved in paper form. Electronic search tools may help for management and retrieval.
- Electronic information is kept on electronic media.

Paper documents or some of the contextual information they contain can be scanned to obtain electronic records in order to simplify searching and consultation.

4.3. Integration, conversion and appraisal of electronic information

4.3.1. Integration of information

Reorganising, selecting or aggregating data to create a more compact data bundle is a delicate task. The right to privacy must be preserved (see Section 6.3 on anonymity).

Grouping data can be useful for later searches, but it can also be a delicate business

Sometimes, however, it may be necessary to consolidate data from various organisations: future researchers who make use of public data are particularly interested in searching by subject as well as by organisation of origin.

What should be included?
Documentation, structure and a coherent set of texts and data

There is always more than one solution for grouping information. Should a report be grouped with subsequent amended versions? Which documents should be kept with a given database? Whatever the solution chosen, in addition to the document or the database, information should include its own documentation (including metadata) and structure (either as a separate file or incorporated into the main file).

Documents created from databases

Example

When a document is created dynamically from a database, there are various solutions for its long-term storage:

- 'freezing' the database with a specific query to produce a traditional document, in which case some possible combinations are lost;
- if the application has a full audit trail, the full database shall be exported when the system is out of operation;
- preserving the database and the application which generates the document dynamically and even the hardware which runs the application.

The solutions are not always obvious when dealing with composite documents which are more complex than simple documents.

The size of the unit of information can vary depending on the choices made. On one hand, the huge number of references and hypertext links between documents nowadays would mean saving all the data on the planet to form a single record! On the other hand, information units must be of reasonable size to be usable.

4.3.2. Conversion of information

Conversion from paper to electronic form:
- simple scan
- scan and encode

There are two main solutions for converting a document from paper to digital format:

- simply scanning the document to obtain an image of it;
- scanning the document and then encoding it in electronic form (e.g. using optical character recognition or graphics vectorisation as described in section 4.3.4).

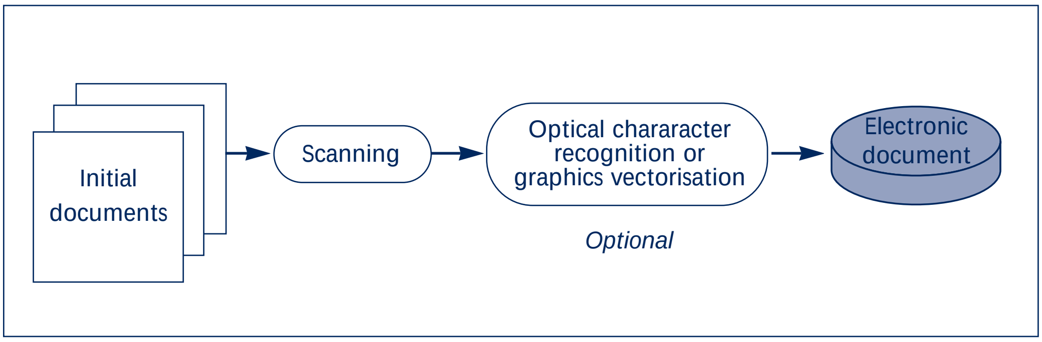


Figure 3 — Conversion from a paper document to an electronic document

To encode or not to encode?

Options

Simply scanning a text is the more straightforward solution and does not require any further processing, unlike optical character recognition (OCR).

However, the resulting file will be larger (more than 50 kilobytes per page, as against several kilobytes in the case of OCR). In addition, it is much easier to edit a text composed of characters rather than an image of a text. Indexing for cross references is also much simpler.

Another option involves a combination of these two approaches, saving the ‘raw’ images and then using optical character recognition for the parts comprising text. Details about formats suitable for short or long-term storage are given in Chapter 5 (Short and long-term preservation of electronic information), while the question of accessing data is dealt with in Chapter 6 (Accessing and disseminating information).

A third type of conversion involves changing one digital format into another. This should only be done to convert an existing record in a given format into a more standard and more lasting format or to convert into a standard which will allow more possibilities (e.g. adding a structure to a flat file). It can also be used as a third stage in the conversion of a paper document in order to obtain an electronic record with a more highly structured format (e.g. a document with an explicit structure or a database).

Digital conversion
Conversion to a different digital format can be done with all types of records

Although converting paper records to digital formats mainly involves documents, it may sometimes be possible to give a database structure to the electronic record obtained using this third stage (see section 4.3.3).

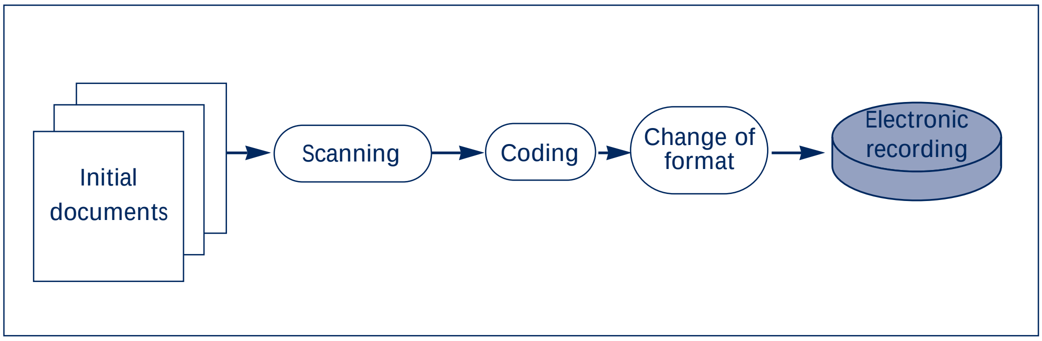


Figure 4 — Conversion of a digital format

4.3.3. From paper or microfilm to scanned image

Scanner and fax

Basic concepts

Scanning a document enables you to save the content of paper pages as a computer file containing an image of the initial document at a given resolution.

The resolution of a scanner is measured in dots per inch (dpi). Currently, scanners can easily achieve a resolution of 300 or 600 dpi in colour.

A fax machine consists of a scanner, a system for sending data via the telephone line and a printer at the destination. The scanner resolution of a fax is much lower (between 100 and 200 dpi).

Scanning ▸
A number of rules
must be observed
to obtain a usable result

A number of basic rules need to be observed when scanning documents. The quality of the result largely depends on this.

Text should not be stored as an image file unless it is short or where optical character recognition is impossible (e.g. signatures). In other cases it is preferable to encode text.

Scanning a document

Options

- All the items needed for the initial record must be transferred onto the same medium.
- Links between the record and the rest of the archive system must be preserved (e.g. other reference records).
- The quality of the original document (contrast, size of characters) must be good enough to ensure the best possible printout after conversion to a digital format.
- Before scanning documents, a sample should be tested.
- Even if scanning is subcontracted out, the organisation having the work done should check the digitised documents for quality and completeness.
- Forms should be designed with their possible conversion to electronic form in mind (font sizes, position of fields).

4.3.4. From scanned image to encoded format

Optical character ▸
recognition
OCR makes it possible to
generate text files from
images of text

Optical character recognition

Basic concepts

Optical character recognition enables the computer to ‘read’ a text.

OCR software works from a file containing an image of the text to be ‘read’ (e.g. a file generated by a scanner). It analyses the shapes of the characters and creates a file in text form which can then be edited by any word-processing program.

Recognition is not perfect. When the computer is unable to identify a character, it marks it for the human operator to identify it. More rarely, the computer may ‘read’ a character incorrectly. Automatic correction tools can help to optimise correction, but a human operator must always make a check after optical character recognition. This check may also involve using software tools.

The recognition rate for a clean image of a typewritten text is about 95% (i.e. two to three errors per line of text). This rate may differ according to the language/script of the source.

After scanning a document containing text and images, the text can be extracted using OCR. This makes it possible to edit the text, to use parts of it, or to index it for easier consultation.

Faxes
Character recognition from faxes is poor

Files from faxes

Advanced topics

It is often difficult to use optical character recognition with faxes on account of their poor resolution and the poor quality of the printout on paper.

One solution is to save faxes in their original digital format (ITU-T Group III or ITU-T Group IV). Although this does not allow the text to be reworked, it does provide a compressed version of images and text which uses up less memory space.

Vectorisation
works with charts, graphs, etc., and saves storage space

Charts and graphs in the original document can be vectorised to save space.

In some cases — with technical drawings, for instance — vectorisation has added advantages. However, vectorisation only works with charts, graphs and other images made up of outlines.

Vector graphics

Basic concepts

Vector graphics are made up of simple elements (straight lines, curves, rectangles, etc.). Instead of preserving an image made up of dots, as for a photograph, vectorisation identifies the basic elements that go to make up the graphic.

Like OCR, vectorisation has two advantages:

- a vector graphic uses less memory than a bitmapped graphic (only the type of element and its beginning and end coordinates are stored);
- it is easier to edit the graphic or to use a subset of it.

Encoding
A few simple rules ensure good quality results

When encoding ‘raw’ data (using OCR or graphics vectorisation) certain basic rules must be followed, as suggested in the boxes in this chapter. This makes it easier to process the electronic information at all the later stages in its life cycle.

Encoding text and graphics

Options

Parts of a document which are not encoded (e.g. photographs or pictures) should be saved separately in some other suitable format.

4.3.5. From one digital format to another

There are two instances where it may be useful to copy information from one digital format to another:

- migrating from a format to a more durable format to make long-term readability and consultation easier;
- adding a structure to a ‘flat file’ (plain text) to obtain a structured document or database which is easier to consult.

Conversion to a standard format

The main advantage of converting the format of electronic information is the guarantee of accessibility and readability

When electronic information is converted from one format to another, care has to be taken to avoid accidental loss of data. The features catered for by proprietary and standard formats do not always correspond. The two examples below illustrate this problem.

Standard formats for preserving and accessing electronic information are dealt with in Chapters 5 and 6.

Loss of information

Example

Example: loss of information from a document

Let us suppose a document contains the sentence:

> ‘It is recommended that the project be terminated.[1]’

and the footnote at the bottom of the page reads:

> ‘[1]Unless the funding initially proposed is finally allocated.’

The original document is in a proprietary word-processing format and is then saved in a new standard format. If no precautions are taken and the conversion process ignores footnotes, the purport of the recommendation would be substantially altered!

Example: loss of information from a database

Let us imagine that a database created under an accounting program is stored in a proprietary format specific to that program. Since the program will only run under a particular operating system on a particular computer, it is decided to extract the data to save them in an independent format. If no precautions are taken, existing links between the accounts before and after their amendment could be lost.

The second instance where conversion from one digital format to another may be useful is to add a structure to unstructured information (commonly known as a ‘flat file’). This is often the case with a scanned text where the addition of structural features (contents, index, etc.) can make consultation easier. Programs now exist that can help to redefine the structure of a document, for example using different fonts for different levels of heading.

Structuring information
A structure can be added to a ‘flat file’ to obtain a structured document or database

When restructuring scanned documents, it is important to recombine elements which are split by page or column breaks or by the insertion of an illustration, table, etc.

The same applies to databases formatted as flat files (a directory or forms that have been scanned, old databases preserved in a low-level format). Here the structure is often implicit, being indicated by separators (e.g. tabs or semi-colons) or by the position of fields on the scanned page (column, position of a field on a paper form, etc.).

This last example is especially worth noting since there is no *de jure* standard for structured data formats and data often have to be preserved in a lightly structured low-level format (see especially section 5.2.4 on data formats).

4.3.6. Appraisal of electronic information

Preserving information serves no purpose unless it can be consulted when required. When information has to be captured as a record to show evidence of an activity, appraisal is a specifically important point. Authenticity, integrity, value have to be evaluated carefully.

Checking records
Most problems arise during conversion or transmission of information

Disposal of records
which have no further use or value is an important task of appraisal

Disposal

Advanced topics

One major task of appraisal is identification of information or records which should be destroyed. It is neither possible nor useful to preserve this information. Records should be destroyed as soon as they have no further use or value. This aids access to the remaining valuable records.

This can apply to data items within databases etc. where items are derived or of little long-term value. Removing items and preserving the balance makes for speedier access and processing.

The archivists’ skills should be used to select records to be destroyed.

Best practices for disposal
Definition of best practices for disposal by national/regional archives

Best practices for disposal

Example

The Information Strategy Unit (ISU) of the department of Premier and Cabinet, in association with the Archives Office of Tasmania (Australia), has produced guidelines which provide some useful best practices on disposal:

- identification of records having continuing or long-term value;
- temporary records should have retention periods;

- implementation of ‘continuing disposal authorities’ which provide an ongoing legal authorisation for the deletion of records after a specified time.

The disposal schedule No 1 (DS1) manages the disposal of administrative records that are common to most agencies.

Section 11 of disposal schedule No 1 authorises the disposal of duplicate, ephemeral and routine documentation which is not intended or required as a formal record. This includes:

- notes and calculations which are used solely in the preparation of other records such as correspondence, reports, budget documents, programs and tabulations;
- draft documentation not intended for retention as part of the agency’s record, the content of which is incorporated into the record-keeping system or registered as a corporate document. This will include rough drafts and preliminary drafts. However, to meet evidential requirements or to document the development of policy, it may be necessary to retain working group drafts when comments on such drafts have resulted in substantial changes. Minor editorial changes are not regarded as significant;
- copies made from corporate records or records which are held elsewhere in the agency.

Checking information

Options

Various checks can be carried out on electronic information:

- compliance with established standards;
- write protection;
- readability of media;
- comparison with data included in the documentation (e.g. test by printing out the first few records);
- completeness of content.

As a rule, however, tests are not exhaustive. Sometimes a problem is only identified when a record is accessed in the usual way (during the access and utilisation stage in its life cycle). If the information was supplied by another organisation or department, they can be asked to try to reconstruct the initial information if possible. Most problems occur with conversion (described in this section) or transmission to another organisation (see section 4.5) rather than with the preservation of information.

Proprietary formats
What to do when an application generates data with proprietary formats

When the application used generates data in a proprietary format it may be necessary to preserve the entire system in order to access the information. This includes the application itself, the IT platform, the documentation and possibly even the staff with the know-how to use the application and the platform. The cost of such a solution should be carefully weighed against the intrinsic value of the information.

Alternatively, the information could be transferred in a low-level format (a flat file for a document or a flat sequential file for a database). In this case some information may be lost, in particular some features of the structure.

A third possibility is to convert the information to a more standard format. To develop or purchase a conversion tool will involve some cost. If the data need to be saved in such a way as to preserve the entire structure as well, this solution might be worth considering. Of course, the target format to which the data are to be converted must be as standard as possible to ensure greater durability.

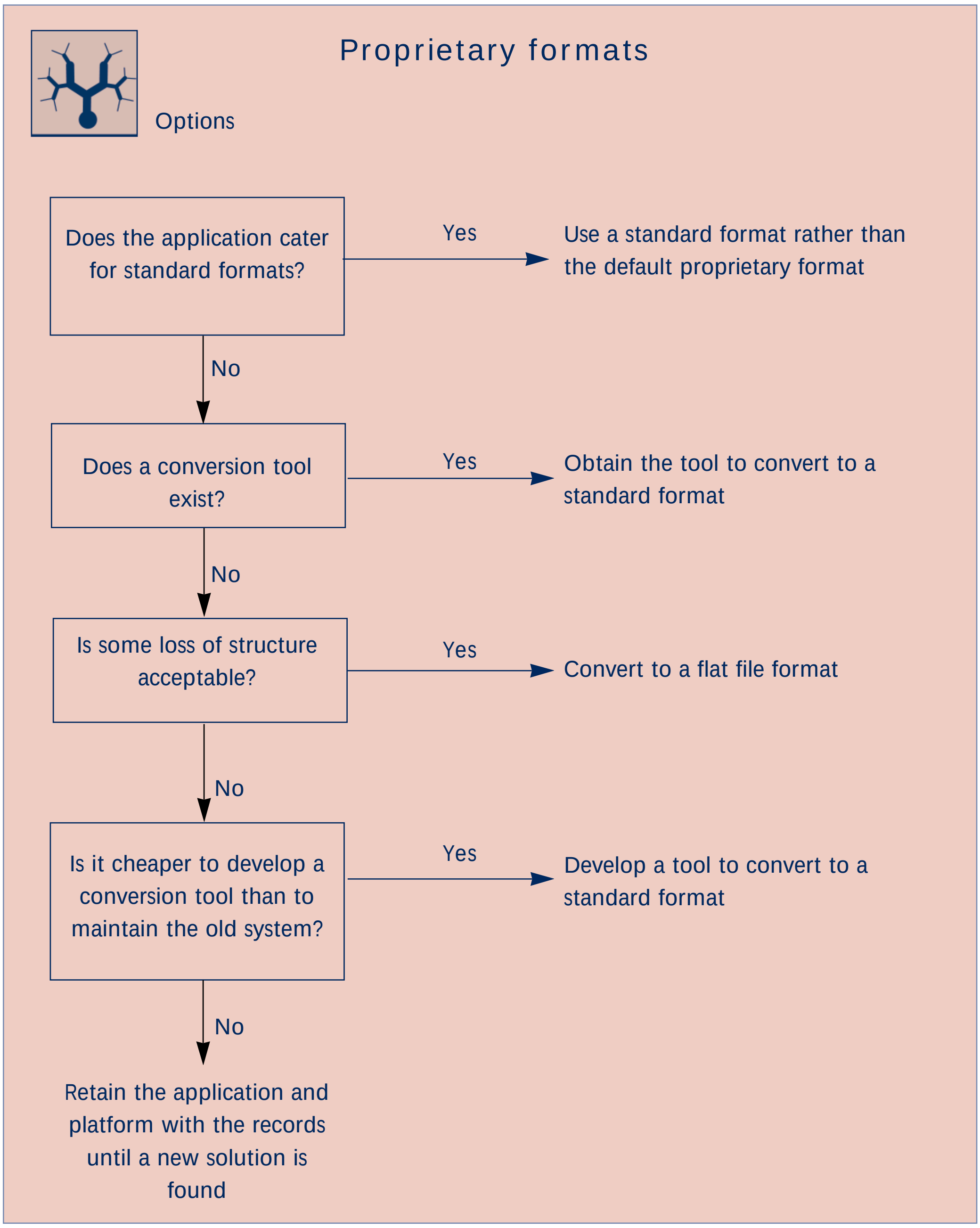

Proprietary formats
Options
Does the application cater for standard formats?
Yes
Use a standard format rather than the default proprietary format
No
Does a conversion tool exist?
Yes
Obtain the tool to convert to a standard format
No
Is some loss of structure acceptable?
Yes
Convert to a flat file format
No
Is it cheaper to develop a conversion tool than to maintain the old system?
Yes
Develop a tool to convert to a standard format
No
Retain the application and platform with the records until a new solution is found

Figure 5 — How to deal with proprietary formats?

Options

There are several solutions for long-term storage of data. Which is best depends on several criteria

It is worth noting that although the cost of converting electronic information to a new format may be high, not doing so could be even more costly.

4.4. Management and classification of electronic information

Managing electronic information:
- register new electronic information
- appoint a person in charge
- coordinate processing
- classify electronic information
- decide on transfers

The aim in managing electronic information is to preserve the reliability, authenticity, integrity and testability of information over time. This requires that the context of information be well defined. When the content, context and structure are sufficient to constitute proof of an activity, then the information becomes a record.

Managing the life cycle of electronic information also involves other tasks. Responsibility for managing a particular set of electronic information may be transferred to another organisation or department (see section 4.5 on transfers).

Management of electronic information includes the following tasks:

- Registering electronic information accessions:
  this requires updating the audit trail of the data in question;
- Assignment:
  when several organisations or departments are involved, the electronic information must be assigned to the right organisation or department and an electronic information manager should be appointed;
- Follow-up:
  this involves coordinating the various stages in processing (reception of electronic information, conversion, preservation, utilisation, transfer to another organisation);
- Classification:
  this is to make it easier to retrieve specific electronic information (see below);
- Decision to transfer to another organisation or department.

Management of electronic information involves more and more groups of persons. Workflow tools may be of great help to manage exchange between a cooperative group.

Classification of information

Classifying information is one of the most important management tasks

One of the most important tasks, and perhaps the most complex, is classifying information. The coding system must be clear enough to be understood by other organisations (especially if responsibility for electronic information is transferred to another department or organisation).

The structure enables us to locate a particular item of information in a document or database, while its classification enables us to locate it among all information stored.

The general classification plan should be used whatever the support of information is. Additional criteria for indexing information can be added to take into account the specificity of electronic information.

European Commission

Example

Some departments in the European Commission use a basic classification plan based on the way the institution operates, its administration, personnel and budget.

Various criteria can be used for classifying and indexing electronic information. These criteria are especially useful in the case of electronic messages.

Some examples of criteria follow, which can be used for classification and indexation:

- type of document;
- dates (production, expiry);
- author(s) (individuals, departments);
- signatory;
- destination (individuals, departments);
- copies (individuals, departments);
- electronic information number, version;
- assignment (date, file, department);
- subject;
- project or activity;
- key words;
- language;
- number of pages;
- status (official, unofficial), confidentiality;
- documents attached, links with other documents;
- other categories defined by the user.

Classification scheme
Various criteria are used to classify electronic information

Archival description of a record

Options

The International Council on Archives has produced international standards for archival description (ISAD/G) which may be taken into consideration when defining a classification strategy.

Identification of the archival unit of the record is called ‘identity statement area’. It is recommended not to modify it during the life cycle of the record.

Electronic mail

Advanced topics

Electronic mail may be managed in different ways depending on the content of the mail.

- It should be processed as an ‘electronic’ letter each time the content needs to be archived and may be viewed as an ‘official’ statement.
- It should be processed as an informal discussion (like, for example, a telephone call) each time the mail will not have any future utility. This is the case, for example, for an automatic reply stating that someone is ‘out and will read his mail when he comes back’.

Because electronic mail allows both formal electronic letters and informal discussions, it is sometimes difficult to decide how to manage it. The appraisal skill of archivists will help greatly to define new rules on e-mail management.

The Australian archives give some examples of ephemeral and routine documents which do not need to be recorded (see the example frame in section 4.3.6).

Electronic mail
must be managed as
- ‘electronic’ letters; or
- informal discussion, depending on the content

4.5. Transfer

At the end of the active part of the life cycle, records can be transmitted to the archives. Not all records have an archival value. A disposal of records which have no further use or value must be done with the help of archivists (see frame on disposal in section 4.3.6).

Two types of transmission:
- physical transmission
- transfer of responsibility

There are two main types of transmission:

- physical transmission of electronic information
- transfer of responsibility.

Information is usually transferred in the form of a record (after having been captured as a record). It can be transmitted to another organisation/department or transferred to archival services.

Responsibility for electronic information (especially management) and its physical custody need not necessarily go hand in hand. Electronic information can be kept by the originating department or a specialised agency. The ease with which electronic copies or transfers can nowadays be made creates the possibility of separating the two roles.

Nevertheless, checks must always be carried out before transferring electronic information.

Successful physical transfer

Options

The key to the successful physical transfer of electronic information lies in observing a few simple rules.

- Information must be checked for completeness (including contextual information).
- Responsibilities must be clearly defined in the sending and the receiving organisations.
- Whenever electronic information is transferred, both organisations (sender and receiver) must make sure that the information is not altered without prior approval.

United Kingdom

Example

The Public Record Office in the United Kingdom is currently proposing a new strategy for transferring electronic records.

Transfers are made only via a secure electronic network.
This means that records are independent of the medium used by the provider.

The agency receiving records for preservation (in this case the Public Record Office) will select a single type of medium for storing the records. It will retain control of that single technology, so ensuring durability.

The bundle transferred includes various elements which are stored together:

- the electronic record itself;
- an audit;
- metadata;
- a viewer that runs on a PC;
- a browser that also runs on a PC.

PCs are considered to have sufficient upward compatibility to ensure that records can continue to be read in the future. Each record, then, is autonomous and free of any formatting problem.

Checks made before the transfer will help to ensure that records are properly readable.

Checking records

Checks will ensure good quality transfers

Checks before transfer

Options

The following list is an example of the steps to be taken by the transmitting organisation when checking information before its transfer. This should be done in close cooperation with the receiving organisation (archives service or other).

1. Make two copies of the data.
2. Compare the data with the documentation supplied using a statistical program.
3. Identify and document any errors in the documentation.
4. Other cross-checks can be made, such as inter-record and inter-variable checks to verify the consistency of data.
5. Consult the supplier of the electronic information in the event of difficulty in identifying codes or if there are errors and inconsistencies in the data.
6. Document the physical files, indicating any difficulties encountered.

5. Short and long-term preservation of electronic information

Media and file format

Two key factors in long-term storage

Obviously the physical medium on which electronic information is stored should have as long a lifespan as possible. But so must the technology, since there is no point in physically preserving records if the hardware and software are no longer capable of processing the data they contain.

We have already noted that medium and content are not the same thing when it comes to electronic records. We will now look in turn at standard types of media and file formats, focusing on the lifespan of media and the maturity and durability of standards.

5.1. Data storage media

Three types of medium

Besides paper:

- microfilm
- magnetic media
- optical media

Many different types of medium can be used. Some are better suited for short-term storage, while others are better for long-term storage.

Apart from paper, storage media can be divided into three major families — microfilm, magnetic media and optical media — with a large number of subtypes within each group. Other, less common types of medium (such as paper tape) are not dealt with here.

Storage media

Basic concepts

- **Microfilm:** This form of storage is widely used in archives but microfilms do not easily lend themselves to reprocessing or searching within records. There are established standards and the lifespan is excellent.
- **Magnetic media:** The technology is fairly old, exploiting the polarisation of magnetic particles in one direction or the other to store each bit (0 or 1). This form of storage generally uses tapes which allow sequential access to data.
- **Optical media:** This is the most recent type of storage medium. Compact-disc technology uses the deflection of a ray of light by minute depressions in the surface of the medium to indicate changes in the value of bits. This type of storage normally uses disks, giving direct access to information (faster than tape) and offering a high storage density.

At present magnetic tape and microfilm are widely used for long-term storage, but optical media are becoming increasingly common as they are particularly suited for long-term storage.

For shorter-term storage a wider variety of media can be used, as durability and lifespan are less critical.

Magnetic media are usually in the form of tapes or cartridges

Magnetic media

Options

The various types of magnetic media are fairly standardised, with varying lifespans. The main types are described below.

○ ○ ● **Diskette:** $3^{1}/_{2}$ inch floppy disks are highly standardised and can be used with a great many systems (PCs, Macintosh and Unix). The amount of information they can hold is quite small (usually 1.44 MB) and since their lifespan is limited, they are used only for very short-term storage and file interchanges.

○ ○ ● **Magnetic cartridge**: cartridges are widely used for storing data on medium-sized systems. There are quarter-inch cartridges and the half-inch cartridges made by IBM.

○ ○ ● **Magnetic tape**: 1 600-bpi tape is readable on practically any tape drive and has been recognised by X-Open as an interchange format. 6 250-bpi tape with a capacity of 112.5 MB is widely used in older archives. Tapes have to be re-spooled every two years and rewritten every 10 or 15 years onto a new tape (of the same or a different type).

○ ● ○ **8 mm video cassette**: although standardised, there is only one main supplier. Typical capacity at present is 2.3 GB. Cassettes have to be rewritten every two years.

○ ● ○ **DAT (digital audio tape)**: DAT is standardised and although mainly used for audio recording, some use is also made of it in information technology. Typical capacity at present is 2.3 GB. Tapes have to be rewritten every two years.

- There are several formats which, for the time being, are fully proprietary. This includes high-density floppy disks and removable hard disks. In addition, credit card memory, while having a limited capacity, may become an important storage media in the future.

Optical media are developing rapidly and should eventually replace other types of media for long-term storage. At present, however, magnetic disks offer faster access times, giving them the advantage for purposes of mass storage.

Optical disks provide large capacity and (unlike tapes) direct access

Optical media

Options

The main types of optical media are described below.

○ ○ ● **CD-ROM**: CD-ROMs are standardised, which makes them a good choice for new archives. There is a difference between pressed CD-ROMs (more suitable for making large numbers of copies, but more reliable) and individually cut CD-ROMs (cheaper for small numbers of copies, but less reliable unless they are given a layer of varnish after cutting). CD-ROMs have to be rewritten every 10 or 20 years. While it is not always mandatory to store them in a controlled environment as with magnetic tapes, this can help in having a longer duration. In addition, more expensive CD-ROMs on glass have a much longer lifespan.

○ ○ ● **DVD (digital versatile disk)**: This new medium could well become an excellent medium for archives. DVDs have a large capacity (4.7 to 18 GB) and should soon become quite widespread. DVD players are also able to read CD-ROM (only more recent DVD players are able to read engraved CD-ROMs).

● ○ ○ **WORM disk (write once read many)**: There is no international standard and WORM disks are not very widespread. They have to be rewritten every 10 or 20 years.

○ ● ○ **Erasable optical disk**: Most of these disks use magneto-optical technology, which combines optical and magnetic storage technologies to give speed, high density and the possibility of rewriting data several times. However, there are very few standards at the moment. These disks have to be rewritten every 10 or 20 years.

Proprietary software is often used to access stored data. It is important to carefully check that file formats and tree structures can be accessed easily.

Given the rapid advances in technology and the limited lifespan of electronic media, it is advisable to rewrite digital archives periodically. Although this adds to cost, it overcomes many of the problems posed by non-standard formats and changing technology. However, most magnetic and optical media use error detection and correction units which facilitate automatic recovering of data errors.

Every time an archive is rewritten the user has to decide between:

- keeping the old files as they are; or
- converting to a more up-to-date medium and/or format (see also section 4.3.5 on converting from one digital format to another).

The factors to be borne in mind include not only the financial aspects but also accessibility, readability, durability and the preservation of authenticity.

Preserving electronic records
It is important to observe certain rules with regard to the storage environment

Long-term storage

Example

There are a lot of discussions within ISO, ANSI and ICA [1] about the best practices for storage of electronic records. The following figures illustrate some examples of practices:

- average temperature: +18°C/-5°C
- relative humidity: +40%/-5%
- interval between rewriting: 10 years.

[1] ISO: International Standards Organisation;
ANSI: American National Standards Institute;
ICA: International Council on Archives.

5.2. File formats

There is a wide range of standards, depending on the type of data to be saved. Best practice is to decide on a common set of standards from the outset to make it easier to circulate information. Preferably the same formats should be used for both short-term and long-term preservation.

Several types of format:

- bitmapped graphics
- vector graphics
- text
- data
- programs
- video and audio
- others

The different types of format

Basic concepts

The different formats that exist can be grouped together in several major families, according to their content.

- **Bitmapped graphics** (raster graphics): These consist of a set of dots. Scanning a document, for example, produces this type of graphic. Bitmapped graphics take up quite a lot of space and are usually compressed before being stored. They can be used as a source for subsequent encoding (for retrieving text or vector graphics) or saved 'as is' (a photograph, for example). The type of compression used dictates the type of graphics file format. Fax formats are a special case of compressed images.

- **Vector graphics**: When dealing with charts and graphs or other graphics consisting solely of outlines, a great deal of space can be saved by using a vector format, where only the coordinates of the vectors making up the graph (line segments, curves, arcs, etc.) are stored. This type of format is not suitable for photographs.

- **Text**: In general, texts involve three separate aspects:
  — the plain text itself, consisting of a set of coded characters;
  — the text structure (e.g. title, chapters, terms to be emphasised, lists, etc.);
  — the layout (terms emphasised will be in bold or red, etc.).

- **Data**: The possibilities to maintain data with the original functionality can be very different. For spreadsheets there is no economically realistic way to maintain the calculation capability today. For applications based on information and process models, the possibilities for preservation are increasing. Database extraction made in accordance with standards together with the models expressed in standards is the way to proceed.

- **Programs**: Programs are less independent of the IT platform than data and IT platforms are becoming obsolete more and more rapidly.

- **Audio, video** and all the other objects that may be included in a record.

The way ahead

Advanced topics

In the future, documents and database files will increasingly become composite documents or even object-oriented documents. In other words, documents will consist of several separate, linked elements (text, images, audio and video).

In the meantime, the standards currently being developed (Microsoft's OLE, IBM and Apple's OpenDoc or even Sun's Java language) are not yet stable enough to allow the use of electronic records that incorporate their own processing instructions (reading, navigation, etc.).

5.2.1. Bitmapped graphics

Bitmapped graphics compression can be:
- destructive
- non-destructive

Two types of compression are used for bitmapped graphics.

No loss compression: After being compressed and then decompressed, an image is exactly the same as the original. This requires low compression ratios of approximately 2:1.

Loss compression: In this case the less useful information in an image is not saved. In practice, the human eye distinguishes certain parts of an image much less clearly than others. Higher compression rates can be achieved with this type of format depending on how much deterioration is acceptable.

There are many different formats for bitmapped graphics

It is essential to choose a mature and lasting standard

Bitmapped graphics formats

Options

The following descriptions cover the main graphics file formats (including fax and video formats).

○ ● ○ **TIFF (tag image file format):** This format is often used for scanned files. There are several options, depending on the number of colours chosen. It is a no loss format but only a limited level of compression can be achieved. The latest version of TIFF (6.0) supports multi-page options. No other type of compression (using algorithms such as Packbits 32 733 by ITU-T, LZW or JPEG) should be used on top of the TIFF format to ensure that it remains properly portable.

○ ● ○ **GIF (graphics interchange format):** This format originated with CompuServe and is quite common, especially on the Internet. There are two specifications, GIF 87A and GIF 89A. Browsers can often read both.

○ ○ ● **JPEG (joint photographic experts group):** This international standard is becoming more and more widespread (including on the Internet). It is a loss compression format that offers high compression. It is definitely a good choice in terms of storage space and durability.

○ ○ ● **Faxes:** There are two fax file formats depending on whether a telephone line or an ISDN line is used.

○ ○ ● **Video:** There are two video formats:

MPEG-1, for computers and multimedia applications;

MPEG-2, a more recent version, for digital television (including sound).

● ○ ○ **Other graphics formats:** It may be dangerous to use other proprietary formats for graphics (such as BMP or PCX), as there is no guarantee as to their durability. Kodak's photo CD format is not often used for documents.

5.2.2. Vector graphics files

Vector graphics formats

Options

○ ○ ● **CGM (computer graphics metafile)**: This is a standardised format for vector graphics that offers a reliable guarantee for their durability.

○ ● ○ **Formats specific to particular types of application**: Some applications require a special approach with standards of their own, for example, geographic information systems (GIS) or computer-aided design (CAD).

● ○ ○ **Other graphics formats**: It is best to avoid using other proprietary formats (such as PICT on Macintosh, Microsoft's Windows Metafiles, or the many other specific application formats), as there is no guarantee as to their long-term durability.

Vector graphics formats

make it possible to save the structure of a graphic

5.2.3 Text files

There are several types of text file, depending on whether the structure and/or the layout are preserved.

- **A plain text file** is a low-level file containing just the text as a sequence of characters. It is difficult to navigate around inside it because of the loss of structure.
- **A structured plain text file** is very useful for navigation and provides a file which is independent of the hardware.
- **A fully formatted text file** contains the characters, structure and layout and is not independent of the hardware used to read it (colour or black and white; text designed for screen or printer, etc.).

Text

A text file, including the structure but not the layout, is independent of the hardware

Character sets

Advanced topics

There are three main families of codes for representing characters:

- **ISO 646** is (almost) equivalent to the ASCII character set. This 7-bit standard does not handle special European characters (such as accented characters).
- **8-bit character sets** are a superset of ISO 646. There are two standards of interest for the European Union: ISO/IEC 8859-1 for western Europe and ISO/IEC 8859-7 for Greece. There are other character sets for Arabic, Hebrew and Cyrillic.
- The **universal character sets (UCS)** standard (ISO/IEC 10646) is designed to allow most of the characters and symbols in use throughout the world to be coded, mainly using two bytes (UCS-2, also known as Unicode) or four bytes (UCS-4). Another code (known as UTF) uses a variable number of bytes and is designed to make it easier to transfer multi-byte characters between different machines. It does this by avoiding the use of 8-bit control characters in the multi-byte character sequences. (For example, the control code to indicate the end of a character sequence or string is often 00. In UTF no other character includes the value 00 among its bytes, whereas in UCS-2 the code for the letter 'A' is 00 64, for instance.)

Standard character sets

There are character sets to cater for different languages and various other user requirements

Layout

It is better to preserve just the structure rather than the style associated with each element

Many features of the layout of a text are dependent on the platform being used. For instance, it is no use specifying a blinking character for a paper printout. However, with technology evolving so rapidly, some means (as yet unimagined) may well be devised in the future to read the texts stored today.

Preserving the layout (character fonts, for instance, or text macros) for long-term storage poses problems when it comes to using the text. The best solution is to preserve only the structure of the text (e.g. this is a first-level heading) rather than the style associated with it (e.g. first-level headings are in 16-point Arial/Helvetica). It is left to the program used to view the text to choose a style for the different parts.

Structured text

There are some standards and a great many proprietary formats

Structured text

Options

There are several standards for storing structured text.

○○● **SGML (standard generalised markup language)**: Strictly speaking this is not a structure but a markup language. It is an international standard and is becoming more and more widespread. It can be used to save a text and its structure, but without the layout. There are several additional standards too, such as:

— DSSSL: document style semantics and specification language;
— SPDL: standard page description language;
— SDIF: SGML document interchange formats;
— font information interchange;
— DTD: document type definition (there are several types, depending on the type of document).

○○● **HyTime**: This is an extension of SGML which can be used to incorporate multimedia material into composite documents.

○●○ **HTML (hypertext markup language)**: This is a simplified implementation of SGML and is particularly common on Internet Web sites. It is still not very stable, however, nor is it very suitable for long documents. SGML is preferable for long-term document storage.

○●○ **ODA (office document architecture)**: This is an international standard which can be used to incorporate the text, the structure and the layout in a single document. ODA was developed for use in electronic office applications. It is not independent of the platform used.

●○○ **RTF (rich text format)**: This format is mainly used by all Microsoft Office software. There is no guarantee of stability and durability. It is preferable to use export filters or conversion software to save files in standard formats.

●○○ **PostScript**: This page description standard from Adobe is widely used for sending or printing texts with their layout. It now needs to be replaced by open standards.

●○○ **PDF (portable document format)**: This allows documents only to be displayed on different platforms.There are also many other proprietary formats.

Specific applications

Advanced topics

Text structure can be specific, with extra fields to facilitate indexing. This applies particularly to electronic messages.

Standards have emerged to make it easier to exchange structured information, mainly for business purposes. These are **EDI** (electronic data interchange) and especially the **Edifact** standard. In addition to the generic EDI definitions, specific EDI standards have been established in a variety of sectors such as banking and the automobile industry. Although the EDI approach is extremely attractive for formal exchanges between organisations, EDI standardisation is slow and does not always take account of the rapid developments in technology (as with graphics, audio and video).

EDI and electronic messaging

Two specific types of application using specific structures

5.2.4. Data and programs

At the moment there is no high-level standardised format for the data files used by spreadsheet and database programs. In order to be sure of being able to read data after a long period, users must have a tool that can read the old format or must keep the old software itself.

Data and programs

pose a difficult problem because there is no widely accepted standard format

It is important not to confuse the interface between the program and data (e.g. SQL in the case of a database) with the format of the file where the data are stored.

The problem is that programs are not as hardware-independent as data.

Where there is no standard format for a particular type of data the best answer is to:

- use a widespread proprietary format which can be reread by many programs; or
- establish a conversion strategy (or a strategy for keeping the software with the data). Section 4.3 gives some suggestions on devising such a strategy.

There is a similar problem with keeping programs, as the user then has to keep the source program or maintain a hardware system on which the program can be run.

Different types of data involve different types of files:

- spreadsheets
- databases
- accounting/business
- forms
- digital signatures

Data files

Options

Data can be saved in many types of file. The main types are:

● ○ ○ **Spreadsheets**: The EXCEL data format can be regarded as a *de facto* proprietary standard. The inclusion of tables in SGML may be a first step towards an open standard.

● ○ ○ **Database files**: Although more and more databases use SQL for retrieval, there is no high-level format for saving databases. The best solution is to use a widespread database program or to save the data as plain text with field separators (so that a new program can recreate the database). In this case the structure of the database has to be saved as well.

○ ○ ● **Edifact (electronic data interchange for administration, commerce and transport)**: Several types of message allow exchange of accounting files and invoices.

○ ○ ● **STEP (standard for the exchange of product data)**: This set of standards facilitates exchange of engineering product data. It includes the Express language for product data representation and exchange. These standards are widely used.

○ ○ ● **IDEF (function and information modelling)**: IDEF0 is the methodology used to scope each application protocol in the STEP standard. It is also widely used in business process re-engineering (BPR). IDEF3 focuses on business process improvement and gives some standard graphical formats for workflow.

○ ● ○ **Forms**: In this case only the information entered needs to be saved and a single copy kept of a blank form. Efforts to establish a standard for forms are under way (FIMS, HTML 3.0, etc.).

○ ○ ● **Digital signatures**: There are two formats for encrypted signatures, DES and RSA. They can be used to check document authenticity. Encryption techniques are described in section 6.3.1.

6. Accessing and disseminating information

6.1. Towards the information society

The European Commission is committed to the concept of the information society, a society in which electronic information will play a key role.

The information society puts electronic information centre-stage

The information society

Advanced topics

In January 1994, Jacques Delors, then President of the European Commission, brought out a White Paper on ‘growth, competitiveness and employment’.

At the same time, he asked Martin Bangemann, Member of the European Commission responsible for industry, to prepare a report with the help of a high-level group including prominent figures from the media, telecommunications and information technology industries. That report, ‘Europe and the global information society’, covers not just infrastructures but also universal services, applications and the societal aspects.

In February 1995, the Group of Seven held their first meeting on the information society and proposed setting up 11 pilot projects, including an ‘on-line government’ scheme to bring government closer to citizens using the new technologies.

The Commission continues to establish the building blocks of the information society in Europe and has already contributed to raising awareness of both the public and the decision-makers. Therefore the Commission has started up different initiatives, among which is the ‘Rolling action plan’ which presents a list of all important actions, in particular legal measures, required to further implement the Information Society in Europe, by fulfilling four major functions:

- to serve as a navigation tool for the Member States and other European institutions (clarity and transparency);
- to provide detailed information about the development of the regulatory framework;
- to provide open information to all interested parties, in particular the public, about EU policy , allowing constructive dialogue between the Commission and all relevant segments of society; and
- to provide a useful internal management tool for the Commission.

It is becoming vital for electronic information not only to be preserved but also to be made as easily accessible as possible. Information is consequently circulating more efficiently inside organisations. Not just official bodies but also ordinary citizens now enjoy easier access to the information to which they are entitled. Even when electronic information is no longer in current use, it may be used for statistical, scientific or historical research.

Archivists and the information society

Archivists are retaining the collective memory of the information society

The archival profession as well as other information-related professions are essential and active parts of the modern information society. They are called upon — the DLM-Forum on electronic records (Brussels, 18 to 20 December 1996) was a first interdisciplinary approach towards this end — to retain the collective memory of the information society.

Privacy and openness

Conflicting interests covered by national and Community legislation

One of the difficulties when disseminating electronic information is to identify the right information to preserve or group together. Information can be used in ways never anticipated by the originator. For example, the items in a database may be used for demographic research based on specific selection criteria.

There is a constant conflict between the interests of privacy and openness. The dividing line is set by law. It is important to check the rules laid down by both European directives and national legislation.

Visually impaired persons

Advanced topics

Persons with a reading impairment use two basically different methods for electronic information reading.

1. Persons who are blind or have a low vision can handle electronic information by using a combination of ‘standard’ computer programs (word processors, spreadsheets, databases, etc.) and an extra program, running in the background, providing screen enlargement for persons with low vision and providing extra outputs (Braille, synthetic voice) to blind persons. These ‘screen readers’, enlarging the textual contents of a screen or transforming it into synthetic speech or temporary Braille (i.e. a device with a row of tangible dots that can show Braille characters under computer control), exist nowadays for all computer platforms. Reading through large unstructured documents remains cumbersome as persons with a visual impairment lack the facility to ‘overview’ a page or a document.

2. Much faster reading for both groups is possible through the use of structured documents and special reading programs (= 3 D browsers) which send audio, speech or Braille information directly to the output devices (synthesiser or Braille reading line). Furthermore, the text on the screen is reformatted and linewrapped to be easily readable for persons with low vision. These browsers for SGML and HTML are available.

6.2. Data access standards

Disseminating information

Standards have to be chosen to allow easy access to data

Some standards are better suited than others to the task of making information accessible to as many people as possible. Some of these have already been discussed in Chapter 5 which dealt with storage standards. Here we list a selection of the standards best suited for disseminating information. Where appropriate, different formats can be used for storage and dissemination.

Dissemination standards

Options

Media for off-line distribution:

- 3 $^{1}/_{2}$ inch 1.44-MB high-density disk;
- standard CD-ROM;
- in the longer term, DVD (digital video disc) may be a more convenient medium.

Internet protocols for interchange and offering of documents on a network

- TCP/IP (transfer control protocol, Internet protocol): These two protocols enable exchange over a network. They are widely used over the world with Internet and Intranet networks;
- HTTP (hypertext transport protocol) for Web servers;
- FTP (file transfer protocol) for file servers;
- SMTP (simple mail transfer protocols) for electronic messaging;
- MIME (multipurpose Internet multimedia extensions) to allow several formats in the same message (e.g. text in various character sets, images).

Document formats:

- SGML for portable documents. The HTML format is better for short documents such as the home pages on the World Wide Web. Long documents can be put straight into SGML format on Web servers;
- Proprietary word-processing formats (Microsoft Word, WordPerfect). Although these are not recommended for long-term storage, they are *de facto* widely used standards in the PC Windows world.

Database access interface

- SQL 2 (structured query language) for relational databases;
- ISAM (indexed sequential access method) for indexed sequential files (low-level interface).

6.3. Security

6.3.1. Access rights

Users can be given different grades of access to electronic information:

- access to the cover page
- access to all or part of a record
- access to view and print or to print only
- others.

Making a file anonymous is less straightforward than it might seem

Access rights need to be considered carefully to preserve every individual's right to privacy and anonymity. There are two ways of making a file anonymous for release into the public domain:

- deleting certain fields (e.g. names)
- pooling information to produce statistics.

Anonymous or not?

Example

Some precautions should be taken to ensure that the procedure used to make a file anonymous will not allow a selective search in order to retrieve a small number of records.

Suppose, for instance, that people's names are deleted from a database. Nevertheless, a selective search could allow people with a given occupation in a particular village to be found. Further cross-searches could then be made to identify a particular individual and uncover further details about him or her (such as his/her salary bracket). It is unlikely, for instance, that there would be very many heart specialists in a village of 1 000 people.

6.3.2. Encryption and authentication

Encrypting data:
- to preserve confidentiality
- to authenticate a record

There are two very different reasons for encrypting data:

- to prevent anyone except selected individuals from reading a record;
- to guarantee that a record has actually been produced by a particular person and has not been altered by anyone else.

When encrypting data it is important to bear specific national features in mind. Legislation on encryption, for instance, varies from one country to another. There may be restrictions on exporting encryption software (as in the USA) or on its use (as in France).

There is much more to this issue than just accessing information or guaranteeing a certain level of confidentiality. It is also a question of ensuring that the correct information is accessed. Record authentication is an important factor in obtaining high-grade information.

Two types of encryption algorithm:
- single-key, DES type;
- double-key, RSA type

Encryption algorithms

Advanced topics

There are two major families of **encryption algorithms**.

- **Single-key (symmetrical) algorithms**
  DES (data encryption system) type (NIST FIPS 46-1, close to ISO 8227-DEA)
  In this case the same encryption key is used for coding and decoding data. The sender and receiver of a coded message have the same key.
- **Double-key (asymmetrical) algorithms**
  RSA type (Rivest, Schamir and Adelman — the names of its inventors).
  In this case there are two keys for each person, one which is kept secret (private key) and another which is made generally available (public key). A message coded with one of the keys can only be decoded with the other.

For a confidential message, the sender sends the message coded with the receiver's public key. Only the receiver can read it using his or her private key.

For an authenticated message, the sender sends a message coded with his or her own private key. Anyone with access to the public key can read the message and is guaranteed that it really was sent by that particular individual and that no one has altered it.

The **PGP** (pretty good privacy) program by Philip Zimmermann is a public-domain RSA-type encryption program which is available on the Internet (despite the US ban on the export of algorithms using sufficiently long keys to allow good security).

Authentication servers exist. They allow you to obtain a person's public key via a procedure that guarantees its authenticity.

6.4. Access to data

Several ways of allowing users access to data

There are several ways of giving people access to electronic information.

- the electronic reading room, i.e. premises open to the public with computer equipment on which electronic records can be read;
- copying electronic information unchanged for use at home;
- making and maintaining a fresh 'consultation copy' of electronic information in a different, more user-friendly format (access might be by consultation on the spot, on-line, or by providing the user with a copy on a given medium);
- using a general access model based on metadata to automatically generate a view of the electronic information that is of use to the user (access might be by consultation on the spot, online, or by providing the user with a copy on a given medium).

The last two solutions are appropriate where the policy is to disseminate data on the Internet.

Germany

Example

With databases in Germany there is a difference between 'research copies' which are made available in a format suitable for consultation by the broad mass of researchers, and 'archive copies' which are stored in flat-file format to avoid problems with formatting standards

Disseminating information:
- making information available (passive dissemination)
- targeted distribution of information (active dissemination)

Disseminating information calls for an information and awareness policy with regard to potential users. There are two potentially complementary strategies:

- making the information available and leaving it to the user to find it using on-line navigation tools (passive dissemination);

- distributing information to a targeted group of users (active dissemination).

These two strategies can usefully be combined by distributing to a target group details of where information has been made available.

It is essential to establish a dissemination policy in order to allow potential users to gain access to the information.

7. Conclusions

These guidelines have endeavoured to give some examples of the ways currently being used to handle electronic information and offer suggestions to help in defining a strategy.

No single approach can be valid for all countries and all situations. So it is up to you to work out your own policy for electronic information in collaboration with everyone else involved.

If these guidelines help you to achieve that, they will have fulfilled their purpose.

8. Annexes

8.1. Terminology

The following definitions are those used for the purpose of these guidelines. National legislation in various countries also contains definitions which have to be taken into account.

Information
An item of knowledge that can be transmitted.

Data item
Representation of a basic piece of information in a format that allows it to be processed.

Machine-readable data (MRD)
Data in a format that is suitable for retrieval, processing and communication by a digital computer.

Document
A consistent and coherent set of data structured to present a line of reasoning or to report on an activity.

Database
A set of data structured to facilitate retrieval and further processing.

Record
A consistent set of data recorded on a storage medium.
'A specific piece of recorded information generated, collected or received in the initiation, conduct or completion of an institutional or individual activity and that comprises sufficient content, context and structure to provide proof or evidence of that activity.' (ICA)

Electronic record
'A record where the information is recorded in a form that is suitable for retrieval, processing and communication by a digital computer.' (ICA)

Medium
The physical material on which records can be recorded, stored and retrieved.

8.2. Issues list

1. The legal value of electronic information
The concept of an 'original' as applied to paper documents is more problematic when it comes to electronic information. It is becoming increasingly easy to duplicate records in such a way that the original can no longer be distinguished from the copy. Furthermore, an electronic record may simply consist of links to other electronic records.

One way of tackling this problem is to use encryption with public and private keys (see section 6.3.2 on encrypting data). The author encrypts a record with his private key. Anyone can then retrieve the public key on an authentication server to read the record and can be sure that it has not been altered since it was encrypted.

2. The meaning of terms may vary from one country to another
Translation of the various terms is not enough by itself, as the problem of vocabulary is much more complex. One solution would be to create a list of concepts with the appropriate terms in each language.

8.3. Understanding standardisation

To allow free competition and ensure data portability regardless of who makes a product, it is important for specifications to include only the interfaces required and not specific products. This allows different suppliers to offer compatible products thus helping to ensure greater durability of data and applications.

Three levels of technological maturity:
- *de facto* standards
- PAS
- *de jure* standards

The level of maturity of a technology can vary, ranging from proprietary products to open standards.

- ***De facto*** standards. Where a product is firmly established in the market, compatibility of other applications and data is measured by reference to that product and changes in compatibility depend on the maker (e.g. Microsoft Word in word processing).
- Publicly available specifications (PAS). Sometimes several leading firms on the market join together in a consortium to define an interface standard. Defining the interfaces makes it possible to develop mutually compatible products (e.g. the X/Open or IETF specifications).
- ***De jure*** standards. Official bodies can guarantee consensus on a specification that then becomes an official standard (e.g. the ISO standard character sets).

As a technology matures, *de facto* standards emerge first, then PAS and finally official *de jure* standards.

There are several international standards organisations, such as:

International standards organisations in the field of data processing:
- ISO/IEC
- ITU-T

- ISO: the International Organisation for Standardisation which is active across a wide range of fields;
- IEC: the International Electrotechnical Committee.

These two organisations have set up a joint committee to deal with standards in the field of information processing.

- ITU: the International Telecommunication Union. The ITU-T committee (the new name for the CCITT) deals specifically with telecommunications standards.

New standards pass through several stages before becoming final. Sometimes a standard may start coming into use once it has reached the penultimate stage of draft international standard (DIS).

There are also several standards organisations for Europe, such as:

European standards organisations in the field of data processing:
- CEN/Cenelec
- ETSI

- CEN: European Standardisation Committee (*Comité européen de normalisation*);
- Cenelec: European Electrotechnical Standardisation Committee (*Comité européen de normalisation électrotechnique*);
- ETSI: European Telecommunications Standards Institute.

The first two deal jointly with information technology. They establish European standards (Euronorm — EN) and European pre-standards (ENV). These often mirror international standards. There is, however, a difference in legal terms: under Decision 87/95/EEC of the Council of the European Communities of 22 December 1986, reference to European standards in public procurement orders is compulsory, whereas compliance with international standards is voluntary.

These organisations develop and establish not only standards but also profiles (sets of standards with a choice of options to allow easy interoperability). The international profiles produced by the ISO/IEC are known as international standardised profiles (ISP).

Several other organisations produce specifications, such as:

Several organisations draw up specifications:
- Open Group
- IETF
- NIST
- other

- Open Group publishes its specifications in the *X/Open Portability Guide* (XPG);
- IETF (Internet Engineering Task Force) produces Internet specifications after issuing a Request for Comments (RFC);
- NIST: the US National Institute for Standards and Technologies draws up profiles known as Federal Information Processing Standards (FIPS);
- many other bodies also work on the many aspects of information processing standards.

The table below gives the references of all the standards cited in these guidelines.

Name	International standard or profile	European standard or profile	Other specification	Remarks
Storage media (section 5.1 and 6.2)				
3 $^1/_2$ inch floppy disk	ISO/IEC 9529-1 ISO/IEC 9529-2	EN 29529-1 EN 29529-2		
$^1/_2$ inch cartridge	ISO 8462-1 ISO 8462-1			
$^1/_2$ inch cartridge				
1 600-bpi tape	ISO/IEC 3788:1976			
6 250-bpi tape				
8 mm cassette	ISO/IEC 11319 ISO/IEC 12246			
DAT cassette				
CD-ROM	ISO 9660 ISO 10149			
WORM				
TMO				
DVD				in preparation
Bitmapped graphics and vector graphics (section 5.2)				
TIFF graphics				
GIF graphics				
JPEG graphics				
Group III fax	ITU-T group III			former CCITT
Group IV fax	ITU-T group IV			former CCITT
MPEG-1 video				
MPEG-2 video				
CGM graphics	ISO 8632			
CAD graphics				
GIS graphics				
Character sets (section 5.2.3)				
7-bit	ISO 646			
8-bit west Europe	ISO/IEC 8859-1			
8-bit Greek	ISO/IEC 8859-7			
Multi-byte	ISO/IEC 10646			
Structured text (section 5.2.3)				
SGML	ISO/IEC 8879	EN 28879		
DSSSL	DIS 10179			
SPDL	ISO/IEC 10180			
SDIF	ISO/IEC 9069			
Font information interchange	ISO/IEC9541			
Standard DTD	ISO/IEC 12083			
HyTime	ISO/IEC 10744			
HTML			W3C HTML 3.0	
ODA/ODIFF	ISO 8613 FOD 26	EN 41509 EN 41515		
Data formats (section 5.2.4)				
EDIFACT	ISO/IEC 9735	EN 29735		
STEP/ Express	ISO 10303			
IDEF0&3			IDEF	
FIMS				
Interchange protocols (section 6.2)				
HTTP			IETF RFC	
FTP			IETF RFC	
Database query (section 6.2)				
SQL	ISO/IEC 9075			version II
ISAM				
Encryption algorithms (section 6.3.2)				
DAS	ISO 8273			
DES			NIST FIPS 46-1	close to DAS
RSA				

8.4. Checklist for electronic information strategy

This annex sets out questions that may need to be dealt with when defining a strategy for electronic information. The figures in parentheses refer to the sections of these guidelines dealing with the topic in question. Future revisions will include showcases with experiences and procurement clauses.

Not every organisation will necessarily need to answer every question or wish to deal with every item (e.g. setting up or adopting a thesaurus). The list of points not covered, unresolved, undecided or rejected also provides valuable information that is an integral part of the strategy.

I — General strategy

A — Identifying those involved (4.1)
Setting up a multidisciplinary strategy team
B — List of common terms and concepts (8.1; 8.2)
C — User requirements. Identifying and monitoring (4.1)
D — Policy regarding the legal value of records (8.2)
E — Information and training policy for departments (4.1)
F — Old technology (media, documents concerned, etc.) (5.1; 5.2)

II — Managing electronic information

A — Identifying responsibilities (for each transfer) (3.2)
1. Responsibility for managing electronic information
2. Responsibility for preserving electronic information

B — Identifying and recording important electronic information (4.1; 4.2)
(size and demarcation against other electronic information)
C — Policy on contextual documentation for electronic information (2.1; 4.2)
D — Defining an appraisal scheme (4.3.6)
1. Procedure for electronic information of doubtful provenance (4.2)
2. Approval procedure for destruction or alteration (4.2)

E — Defining rules for the classification plan (4.1; 4.4)
1. Setting up a keyword classification (thesaurus)

F — Encryption strategy (6.3.2)
1. Confidentiality
2. Authentication

G — Policy on the physical transfer of electronic information (4.5)
1. List of items to be transferred
2. List of checks to be carried out for each transfer
3. Medium used for physical transfers, original medium, secure network, etc.

III — Preservation of electronic information

A — Options for preserving ‘originals’ of electronic information (4.2)
electronic records only or paper copies

B — Data media options (5.1)
1. Short-term storage
2. Long-term storage
3. Storage environment (temperature, humidity, recopying frequency, etc.)

C — Format options for saved files (5.2)
1. Bitmapped graphics (5.2.1)
   (a) Type of compression used (5.2.1)
   (b) Preserving faxes (4.3.4; 5.2.1)
2. Vector graphics (5.2.2)
   (a) Specific graphics files (CAD, GIS)
3. Audio, video and multimedia files (5.2.1)
4. Text files (5.2.3)
   (a) Character sets (default, acceptable)
   (b) Structured text formats
   (c) Whether to preserve layout or not
5. Data files (5.2.4)
   (a) Databases (low-level or proprietary).
   • Compatibility/management files
   (b) Structured text formats
   (c) Whether to preserve layout or not
6. Programs (5.2.4)

D — Policy on preserving old systems and software (4.3.4)
(if necessary). Maintenance, documentation, know-how, etc.

IV — Strategy for keeping paper documents

A — Options for preserving paper documents (4.2)
Identifying documents for scanning

B — Quality chart for documents that are to be scanned (4.3.3)

C — Optical character recognition strategy (4.3.4)
1. Options when using OCR
2. OCR procedure (4.3.4)

D — Strategy for vectorising graphics (4.3.4)

E — Procedure for grouping the various elements of electronic information (text, vector graphics, bitmapped graphics, etc.) (4.3.4)

V — Converting or preserving data formats (4.3.3)

A — Policy for documenting formats of data used by systems and software (2.4.2)

B — Options for converting or preserving old formats (4.3.6)

C — Options for adding a structure to text or data recovered (4.3.4)

D — Options for generating an anonymous base for dissemination (6.3.1)
1. Procedure for ensuring anonymity

E — Procedure for analysing loss of information due to conversion
(4.3.3)

VI — Data access and utilisation

A — Policy on access privileges (viewing and printing) (6.3.1)
B — Making information available (passive dissemination) (6.4)
1. Electronic reading room (6.4)
   (a) Consultation standards (6.2)
2. Copying electronic information unaltered for distribution (6.4)
   (a) Copy media (6.2)
3. Copying electronic information in a format for dissemination (6.4)
   (a) Media, protocols. Languages and formats (6.2)
   (b) Internet strategy (6.2)
4. Access model for automatically generating a user-friendly format (6.4)
   (a) Media, protocols. Languages and formats (6.2)
   (b) Internet strategy (6.2)

C — Promoting access (active dissemination) (6.4)

8.5. Prototype: what metadata should be created?

This annex gives an example of possible metadata based on the Dublin core metadata proposal of December 1996. More information can be found at http://www.purl.org/metadata/dublin_core.

This prototype is not the only way to proceed. It only proposes some examples of practices which may help the reader to define his own strategy.

The 15 elements are all optional and extensible. They describe the context of a specific resource.

TITLE

The name given to the resource by the creator or the publisher

CREATOR

(Author or creator)
The person(s) or organisation(s) primarily responsible for the intellectual content of the resource. For example, authors in the case of written documents, artists, photographers, or illustrators in the case of visual resources.

SUBJECT

(Subject and keywords)
The topic of the resource, or keywords or phrases that describe the subject or content of the resource. The intent of the specification of this element is to promote the use of controlled vocabularies and keywords. This element might well include scheme-qualified classification data (for example, Library of Congress classification numbers or Dewey decimal numbers) or scheme-qualified controlled vocabularies (such as medical subject headings or art and architecture thesaurus descriptors) as well.

DESCRIPTION

A textual description of the content of the resource, including abstracts in the case of document-like objects or content descriptions in the case of visual resources. Future metadata collections might well include computational content description (spectral analysis of a visual resource, for example) that may not be embeddable in current network systems. In such a case this field might contain a link to such a description rather than the description itself.

PUBLISHER

The entity responsible for making the resource available in its present form, such as a publisher, a university department, or a corporate entity. The intent of specifying this field is to identify the entity that provides access to the resource.

CONTRIBUTORS

Person(s) or organisation(s) in addition to those specified in the creator element who have made significant intellectual contributions to the resource but whose contribution is secondary to the individuals or entities specified in the creator element (for example, editors, transcribers, illustrators, and convenors).

DATE

The date the resource was made available in its present form. The recommended best practice is an eight-digit number in the form YYYYMMDD as defined by ANSI X3.30-1985. In this scheme, the date element for the day this is written would be 19961203, or 3 December, 1996. Many other schema are possible, but if used, they should be identified in an unambiguous manner.

TYPE

The category of the resource, such as home page, novel, poem, working paper, print, technical report, essay, dictionary. It is expected that resource type will be chosen from an enumerated list of types. A preliminary set of such types can be found at the following URL:
http://www.roads.lut.ac.uk/Metadata/DC-ObjectTypes.html

FORMAT

The data representation of the resource, such as text/html, ASCII, Postscript file, executable application, or JPEG image. The intent of specifying this element is to provide information necessary to allow people or machines to make decisions about the usability of the encoded data (what hardware and software might be required to display or execute it, for example).
As with resource type, format will be assigned from enumerated lists such as registered Internet media types (MIME types). In principle, formats can include physical media such as books, serials, or other non-electronic media.

IDENTIFIER

String or number used to uniquely identify the resource. Examples for networked resources include URLs and URNs (when implemented). Other globally-unique identifiers, such as International Standard Book Numbers (ISBN) or other formal names would also be candidates for this element.

SOURCE

The work, either print or electronic, from which this resource is derived, if applicable. For example, an html encoding of a Shakespearean sonnet might identify the paper version of the sonnet from which the electronic version was transcribed.

LANGUAGE

Language(s) of the intellectual content of the resource. Where practical, the content of this field should coincide with the Z39.53 three character codes for written languages.
See: http://www.sil.org/sgml/nisoLang3-1994.html

RELATION

Relationship to other resources. The intent of specifying this element is to provide a means to express relationships among resources that have formal relationships to others, but exist as discrete resources themselves. For example, images in a document, chapters in a book, or items in a collection. A formal specification of relation is currently under development. Users and developers should understand that use of this element should be currently considered experimental.

COVERAGE

The spatial and / or temporal characteristics of the resource. Formal specification of coverage is currently under development. Users and developers should understand that use of this element is currently considered to be experimental.

RIGHTS

A link to a copyright notice, to a rights-management statement, or to a service that would provide information about terms of access to the resource. Formal specification of rights is currently under development. Users and developers should understand that use of this element is currently considered to be experimental.

8.6. Prototype: how to select the right standards

This annex provides a proposed standard for each type of data. Any other choice may be made for each type, using the information given in these guidelines. It is good practice to define a recommended standard. Another way of doing it, is to list the acceptable standards for each type of file.

This prototype is not the only way to proceed. It only proposes some examples of practices which may help the readers to define their own strategy.

Type of data	Recommended standard	Comments
Character sets	ISO/IEC 8859-1	For western European countries another possibility is Unicode (ISO/IEC 10646) if other character sets are needed
Structured text	SGML	
Bitmap graphics	JPEG	
Faxes	ITU-T Group III	
Vector graphics	CGM	
Audio and video	MPEG II	
CAD/CAM	STEP	
Accounting/invoice	EDIFACT	
Other database files	Flat file, comma separator	No standard database format exists. The flat file allow long-term preservation if the structure of the database is well documented
Encrypted files programs	RSA Source or PC compatible version	No standard exists for compiled programs. The Java byte code, which is platform independent may help for long-term preservation of Java applications.
Media for long-term preservation	DVD	While DVD is rather new, it should be widely used in the near future. Its large capacity and ease of use may help in making the DVD the media for archives. Several organisations have made their own choice for the media. It is recommended to select one medium or a very small set to facilitate future use.

8.7. Index

8.8. List of figures

8.9. Bibliography

This bibliography lists only a few reference works used in producing these guidelines. A large number of internal documents from national and Community bodies were also consulted but are not listed here.

Archives in the European Union, Report of the Group of Experts on the Coordination of Archives, European Commission, Secretariat-General, Brussels • Luxembourg, 1994
ISBN 92-826-8233-1, Cat No CM-83-94-741-FR-1

Proceedings of the DLM-Forum on electronic records, Brussels 18 to 20 December 1996, *INSAR – European archives news*, Supplement II, 1997, EUR-OP, Luxembourg, 1997, p. 376, ISBN 92-828-0111-X. Cat. No CM-AC-97-S01-EN-C (DE, EN, FR)
On Internet:
http://www.echo.lu/dlm/en/home.html

***European procurement handbook for open systems* (EPHOS)**
Reference: EUR 14021

ICA guide on electronic records, International Council on Archives, Committee on Electronic Records. Paris, 1997 (EN, FR).
On Internet:
http://www.archives.ca/ica/p-er/english.html

B. Bauwens, F. Evenepoel, J. Engelen, ‘Standardisation as a prerequisite for accessibility of electronic text information for persons who cannot use printed material’, *IEEE transactions on rehabilitation engineering*, Vol. 3, No 1, pp. 84-89, 1995.

Chr. Reeves, T. Wesley, *Guidelines for accessible Web page design*,
Brochure published by the Harmony consortium (1997). Also available on the Web
http://www.esat.kuleuven.ac.be/teo/harmony/guidelines