

INTERNATIONAL
STANDARD

ISO/IEC
8613-2

Second edition
1995-04-15

Information technology — Open Document Architecture (ODA) and interchange format: Document structures

*Technologies de l'information — Architecture des documents ouverts
(ODA) et format d'échange: Structures documentaires*



Reference number
ISO/IEC 8613-2:1995(E)

AVA Documents
Document Sharing Hub

Contents

	<i>Page</i>
1 Scope	1
2 Normative references	2
2.1 Identical Recommendations International Standards	2
2.2 Paired Recommendations International Standards equivalent in technical content	2
2.3 Additional references	2
3 Definitions	2
4 Abbreviations	2
5 Conventions.....	3
5.1 Subordinate	3
5.2 Superior	3
5.3 Next and following	3
5.4 Preceding	3
6 Architectural principles	3
6.1 Architectural concepts.....	3
6.2 Structural model of a document	4
6.2.1 Specific structures	4
6.2.2 Content of a document	5
6.2.3 Generic structures	5
6.3 Descriptive representation of a document.....	6
6.3.1 Constituents of a document.....	6
6.3.2 Content portion descriptions	6
6.3.3 Object descriptions.....	6
6.3.4 Object class descriptions.....	7
6.3.5 Styles	7
6.3.6 Alternative descriptions	8
6.3.7 Document profile	8
6.3.8 Protected part descriptions	9
6.3.9 Document class descriptions	9
6.3.10 Generic-documents	9
6.3.11 External-documents	10
6.3.12 Resource-documents	10
6.3.13 Classes of document architecture.....	10
6.3.14 Sets of constituents	11

© ISO/IEC 1995

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the publisher.

ISO/IEC Copyright Office • Case postale 56 • CH-1211 Genève 20 • Switzerland

Printed in Switzerland

	<i>Page</i>
6.4 Document processing model.....	13
6.4.1 The editing process	13
6.4.2 The layout process	13
6.4.3 The imaging process	15
6.5 Roles of the document architecture constituents in document processing.....	15
6.5.1 Editing process.....	15
6.5.2 Layout process	17
6.5.3 Imaging process	18
6.6 Security protection of parts of a document	19
6.6.1 Intended and privileged recipients	19
6.6.2 Protecting parts of the document profile.....	19
6.6.3 Protecting parts of the document body	20
7 Document structures and colour.....	20
7.1 Specific structures.....	20
7.1.1 General principles	20
7.1.2 Sequential order	21
7.1.3 Content portions.....	22
7.2 Specific logical structures.....	22
7.2.1 Document logical root.....	22
7.2.2 Composite logical objects	22
7.2.3 Basic logical objects	23
7.3 Specific layout structure.....	23
7.3.1 Objects of the layout structure	23
7.3.2 Positioning layout objects	25
7.3.3 Naming of edges of frames and blocks.....	26
7.3.4 Measurement.....	28
7.3.5 Borders of frames and blocks	28
7.4 Documents containing both specific structures.....	29
7.5 Generic structures	30
7.5.1 General principles	30
7.5.2 Generator and factor sets of object class descriptions	31
7.5.3 Complete generator sets of object class descriptions	31
7.5.4 Generation of structures.....	32
7.5.5 Content in generic structures.....	32
7.5.6 Generic logical structure	32
7.5.7 Generic layout structure.....	33
7.5.8 Resource-document.....	33
7.6 Colour	34
7.6.1 Overview.....	34
7.6.2 Reference colour space	34
7.6.3 Interchange colour spaces	34
7.6.4 Colour imaging model	34
7.6.5 Colour tolerance.....	35
8 Interface between the document architecture and content architectures	35
8.1 General concept	35
8.2 Specification of a content architecture.....	36
8.3 Interface information	36
8.3.1 Interface between a content architecture and the document profile.....	37
8.3.2 Interface between a content architecture and a basic component description.....	37
8.3.3 Interface between a content architecture and a content portion.....	37

	<i>Page</i>
9 Attribute definitions	37
9.1 General principles of attributes	37
9.1.1 Attribute categorisation	37
9.1.2 Attribute values	40
9.1.3 Expressions	44
9.1.4 Colour attribute values	47
9.2 Attribute definition format	51
9.3 Shared Attributes	51
9.3.1 Identification attributes	51
9.3.2 Construction attributes	54
9.3.3 Relationship attributes	56
9.3.4 Content architecture class	59
9.3.5 Miscellaneous attributes	60
9.3.6 Security attributes	63
9.4 Layout attributes	65
9.4.1 Property attributes	65
9.4.2 Formatting attributes	71
9.4.3 Imaging attributes	75
9.4.4 Presentation attributes	81
9.5 Logical attributes	82
9.5.1 Protection	82
9.5.2 Layout style	82
9.6 Layout style attributes	82
9.6.1 Layout style identifier	83
9.6.2 Attributes which can be specified for layout styles	83
9.7 Layout directive attributes	83
9.7.1 General principles of layout directive attributes	83
9.7.2 Block alignment	84
9.7.3 Concatenation	85
9.7.4 Fill order	85
9.7.5 Floatability range	86
9.7.6 Indivisibility	87
9.7.7 Layout category	88
9.7.8 Layout object class	89
9.7.9 Logical stream category	89
9.7.10 Logical stream sub-category	90
9.7.11 New layout object	91
9.7.12 Offset	92
9.7.13 Same layout object	92
9.7.14 Separation	93
9.7.15 Synchronisation	94
9.7.16 Interactions and precedences among the layout directive attributes	95
9.7.17 Interactions among attributes affecting the layout process	98
9.8 Presentation style attributes	98
9.8.1 Presentation style identifier	98
9.8.2 Attributes which can be specified for presentation styles	98
9.9 Content portion attributes	99
9.9.1 Identification attributesContent identifier – logical;Content identifier – layout	99
9.9.2 Common coding attributes – Type of coding	100
9.9.3 Content information attributes	101
9.9.4 Coding attributes	101

	<i>Page</i>
9.10 Protected part attributes	102
9.10.1 Protected part identifier.....	102
9.10.2 Sealed document profile information.....	102
9.10.3 Enciphered information	103
10 Reference model of the document layout process.....	103
10.1 Introduction.....	103
10.1.1 Overview	103
10.1.2 Initialization	104
10.2 Content and layout structure generation	105
10.2.1 Laying out content of a document.....	105
10.2.2 Specific layout structure generation.....	107
10.3 Layout references and categories	107
10.3.1 Stream categories	107
10.3.2 Layout categories	108
10.3.3 Basic principles of layout streams.....	108
10.3.4 Layout object class.....	109
10.4 The effect of some attributes on the document layout process	111
10.4.1 New layout object	111
10.4.2 Floatability range	111
10.4.3 Same layout object.....	112
10.4.4 Indivisibility	112
10.4.5 Synchronization	113
10.4.6 Balance	113
10.5 Layout process for frames.....	114
10.5.1 Placement of frames with fixed position.....	114
10.5.2 Placement of frames with variable position	114
10.6 Allocation of areas for blocks	116
10.6.1 Determination of the available area	116
10.7 Alternative representation	118
11 Reference model of the document imaging process.....	119
11.1 Imaging order.....	119
11.2 Intersection principles	119
11.3 General rules for positioning pages on presentation surfaces	120
11.3.1 Nominal page and assured reproduction areas	120
11.3.2 Positioning of the page.....	120
11.3.3 Definition of assured reproduction area	121
11.3.4 Recto/verso pages	124
11.3.5 Positioning of pages on soft copy media	124
12 Reference model for protecting parts of a document	124
12.1 The overall model	125
12.2 The local system	125
13 Document architecture classes	126
13.1 Definition of document architecture classes	126
13.2 The constituents of document architecture classes	127
13.3 Minimum attribute sets	128
13.4 Constraints on attributes.....	129
Annex A – Notation used to represent document structures	130
A.1 Notation for structure diagrams	130

	<i>Page</i>
A.2 Notation for expressions	131
A.2.1 Conventions for production rules.....	131
A.2.2 Notation for construction expressions.....	132
A.2.3 Notation for string expressions	132
A.2.4 Notation for numeric expressions	133
A.2.5 Notation for object identifier expressions	134
A.2.6 Bindings	135
A.2.7 Notation for references to binding values	135
Annex B – Examples of document structures	136
B.1 Introduction.....	136
B.2 Notation used to specify constituents of a document.....	136
B.2.1 General	136
B.2.2 Generator for subordinates.....	137
B.2.3 Content portions.....	137
B.3 Introduction to specimen document.....	137
B.4 Specific structures	144
B.4.1 Formatted form documents with specific layout structure only.....	144
B.4.2 Processable form document with specific logical structure only	144
B.5 Processable form document with generic logical structure and generic layout structure	153
B.5.1 Introduction.....	153
B.5.2 Generic logical structure	153
B.5.3 Specific logical structure.....	156
B.5.4 Generic layout structure.....	156
B.5.5 Object descriptions and object class descriptions	156
B.6 Specific layout structure.....	156
Annex C – Examples of particular document architecture features	171
C.1 Layout of frames	171
C.1.1 Conventions	171
C.1.2 Position attribute for a frame in a fixed position.....	172
C.1.3 Single column text of variable height	173
C.1.4 “Wrapping” of text around a left aligned picture.....	174
C.1.5 Picture with text on both sides	177
C.1.6 Multi-lingual synchronized text	181
C.1.7 Footnote placement.....	182
C.1.8 Tabular layout	183
C.2 Layout of blocks	185
C.2.1 Concatenation	185
C.2.2 Offset	186
C.2.3 Separation	186
C.2.4 Layout path	187
C.2.5 Fill order	187
C.2.6 Block alignment.....	188
C.2.7 Illustration of attributes used in combination.....	189
C.3 Further constraints on layout	193
C.3.1 Layout object class.....	194
C.3.2 Synchronization	194
C.3.3 New layout object	194
C.3.4 Same layout object.....	195
C.3.5 Indivisibility	195
C.3.6 Balance.....	195

	<i>Page</i>
C.4 Texture and imaging order attributes	196
C.4.1 Example 4.1	196
C.4.2 Example 4.2	197
C.4.3 Example 4.3	198
C.4.4 Example 4.4, example of an application	201
C.5 Bindings, content generator	201
Annex D – The defaulting mechanism	203
D.1 The defaulting mechanism as applicable to each defaultable attribute	203
D.2 Determination of content portions and their associated attributes	203
D.2.1 Determination of content associated with basic logical objects	203
D.2.2 Determination of content associated with layout objects	205
Annex E – Attribute summary tables	207
Annex F – Overview of alternative description – Technical and implementation aspects	216
F.1 Substituting basic objects	216
F.2 Independence of substitutions	216
F.3 Selection of alternatives	216
F.4 Substitution in the initialization process	216
F.5 Syntactical selection of alternatives	216
F.6 Preference between several alternatives	217
Annex G – Further information on security aspects within a document	218
G.1 What can in principle be protected within a document?	218
G.1.1 What does a document contain?	218
G.1.2 What can an unauthorized recipient do with a document?	218
G.1.3 What protection can be given to a document?	218
G.1.4 Summary	219
G.2 Security features supported by the ITU-T Rec. T.410-Series ISO/IEC 8613	219
G.2.1 Features provided to an originator	220
G.2.2 Features provided to a privileged recipient	220
G.3 The kinds of protection mechanisms supported	220
G.3.1 Confidentiality	220
G.3.2 Integrity	220
G.3.3 Authenticity	221
G.3.4 Non-repudiation of origin	221
G.4 Techniques supported by the ITU-T Rec. T.410-Series ISO/IEC 8613	221
G.4.1 Techniques for confidentiality	221
G.4.2 Techniques for sealing for content integrity, authenticity and non-repudiation	221
G.5 Further details on the reference model for protecting parts of a document	222
G.5.1 The overall model	222
G.5.2 The local system	225
G.6 Document application profiles	225
Annex H – Conversions between the reference colour space and the interchange spaces	227
H.1 CIE $L^*u^*v^*$	227
H.1.1 Conversion from the reference colour space to CIE $L^*u^*v^*$	227
H.1.2 Conversion from CIE $L^*u^*v^*$ to the reference colour space	228
H.2 CIE $L^*a^*b^*$	228
H.2.1 Conversion from the reference colour space to CIE $L^*a^*b^*$	228
H.2.2 Conversion from CIE $L^*a^*b^*$ to the reference colour space	229
H.3 RGB	229
H.3.1 Conversion from the reference colour space to linear (non-gamma corrected) RGB	229
H.3.2 Conversion from linear (non-gamma corrected) RGB to the reference colour space	230
H.3.3 Conversion from non-linear (gamma-corrected) RGB to the reference colour space	230

	<i>Page</i>
H.3.4 Conversion from the reference colour space to non-linear (gamma-corrected) RGB	232
H.4 CMY(K)	234
H.4.1 Conversion from CMY(K) to the reference colour space	234
H.4.2 CMY(K) calibration data	234
H.5 Colour tolerance	235
Annex I – Definitions of colour terms	236
Annex J – Colour concepts	239
J.1 General concepts	239
J.1.1 An overview of CIE colorimetry	239
J.1.2 Appearance matching	242
J.1.3 Device spaces	242
J.2 Colour spaces	243
J.2.1 CIE uniform colour spaces	243
J.2.2 RGB (Red, Green, Blue) colour space	243
J.2.3 CMY(K) colour space	245
J.3 Gamut compression	246
J.4 Colour differences	246
Annex K – Bibliography on colour	247

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 8613-2 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 18, *Document processing and related communication*, in collaboration with ITU-T. The identical text is published as ITU-T Rec. T.412.

This second edition cancels and replaces the first edition (ISO 8613-2:1989), which has been technically revised.

ISO/IEC 8613 consists of the following parts, under the general title *Information technology — Open Document Architecture (ODA) and Interchange Format*:

- *Part 1: Introduction and general principles*
- *Part 2: Document structures*
- *Part 3: Abstract interface for the manipulation of ODA documents*
- *Part 4: Document profile*
- *Part 5: Open Document Interchange Format*
- *Part 6: Character content architectures*
- *Part 7: Raster graphics content architectures*
- *Part 8: Geometric graphics content architectures*
- *Part 9: Audio content architectures*
- *Part 10: Formal specifications*
- *Part 11: Tabular structures and tabular layout*
- *Part 12: Identification of document fragments*
- *Part 13: Spreadsheet*
- *Part 14: Temporal relationships and non-linear structures*

Annexes A and H form an integral part of this part of ISO/IEC 8613. Annexes B to G and I to K are for information only.

Introduction

This ITU-T Recommendation | International Standard was prepared as a joint publication by ITU-T Study Group 8 and ISO/IEC Joint Technical Committee 1.

At present, the ITU-T Recs. in the T.410-Series | International Standard ISO/IEC 8613 consist of:

- Introduction and general principles;
- Document structures;
- Document profile;
- Open document interchange formats;
- Character content architectures;
- Raster graphics content architectures;
- Geometric graphics content architectures;
- Formal specification of the Open Document Architecture (FODA).

(The formal specification is applicable to ISO/IEC 8613 only).

Further Recommendations | International Standards may be added to this series of ITU-T Recommendations | International Standards.

Development of this series of ITU-T Recommendations | International Standards was originally in parallel with the ECMA-101 standard: Open Document Architecture.

This series of ITU-T Recommendations | International Standards is a new edition of the CCITT T.410-Series Recommendations (1988) and ISO 8613 (1989).

Significant technical changes are the inclusion of the following amendments as agreed by ITU-T and ISO/IEC:

- Alternative Representation;
- Annex on use of MHS/MOTIS;
- Colour;
- Conformance Testing Annex;
- Document Application Profile Proforma and Notation;
- Security;
- Streams;
- Styles;
- Tiled Raster Graphics.

In addition, a number of technical corrigenda have been applied.

This ITU-T Recommendation | International Standard contains eleven annexes:

- Annex A – Notation used to represent document structures (integral);
- Annex B – Examples of document structures (non-integral);
- Annex C – Examples of particular document architecture features (non-integral);
- Annex D – The defaulting mechanism (non-integral);
- Annex E – Attribute summary tables (non-integral);
- Annex F – Overview of alternative description, technical and implementation aspects (non-integral);
- Annex G – Further information on security aspects within a document (non-integral);
- Annex H – Conversions between the reference colour space and the interchange spaces (integral);
- Annex I – Definitions of colour terms (non-integral);
- Annex J – Colour concepts (non-integral);
- Annex K – Bibliography on colour(non-integral).

INTERNATIONAL STANDARD

ITU-T RECOMMENDATION

INFORMATION TECHNOLOGY — OPEN DOCUMENT ARCHITECTURE (ODA) AND INTERCHANGE FORMAT: DOCUMENT STRUCTURES

1 Scope

The purpose of the ITU-T Rec. T.410-Series | ISO/IEC 8613 is to facilitate the interchange of documents.

In the context of these Recommendations | International Standards, documents are items such as memoranda, letters, invoices, forms and reports, which may include pictures and tabular material. The content elements used within the documents may include graphic characters, raster graphics elements and geometric graphics elements, all potentially within one document.

NOTE – These Recommendations | International Standards are designed to allow for extensions, including hypermedia features, spreadsheets and additional types of content such as audio and video.

In addition to the content types defined in these Recommendations | International Standards, ODA also provides for arbitrary content types to be included in documents.

These Recommendations | International Standards apply to the interchange of documents by means of data communications or the exchange of storage media.

These Recommendations | International Standards provide for the interchange of documents for either or both of the following purposes:

- to allow presentation as intended by the originator;
- to allow processing, such as editing and reformatting.

The composition of a document in interchange can take several forms:

- formatted form, allowing presentation of the document;
- processable form, allowing processing of the document;
- formatted processable form, allowing both presentation and processing of the document.

These Recommendations | International Standards also provide for the interchange of ODA information structures used for the processing of interchanged documents.

This ITU-T Recommendation | International Standard:

- defines a document architecture intended for representation of documents;
- defines a document processing model;
- defines the document structures, the basic constituents of the architecture and a descriptive representation of these in terms of attributes;
- defines an interface which allows the use of different content architectures with the document architecture;
- defines the reference model of the document layout process;
- defines the reference model of the document imaging process;
- defines the reference model for protecting parts of a document;
- defines three document architecture classes;
- defines a notation used for illustrating and describing document structures;
- provides examples of document structures;
- provides examples of particular document attributes.

2 Normative references

The following ITU-T Recommendations and International Standards contain provisions which, through reference in this text, constitute provisions of this Recommendation | International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Recommendation | International Standard are encouraged to investigate the possibility of applying the most recent edition of the Recommendations and Standards listed below. Members of ISO and IEC maintain registers of currently valid International Standards. The ITU-T Secretariat maintains a list of currently valid ITU-T Recommendations.

2.1 Identical Recommendations | International Standards

- ITU-T Recommendation T.411 (1993) | ISO/IEC 8613-1:1994, *Information technology – Open Document Architecture (ODA) and interchange format: Introduction and general principles*.
- ITU-T Recommendation T.414 (1993) | ISO/IEC 8613-4:1994, *Information technology – Open Document Architecture (ODA) and interchange format: Document Profile*.
- ITU-T Recommendation T.415 (1993) | ISO/IEC 8613-5:1994, *Information technology – Open Document Architecture (ODA) and interchange format: Open Document Interchange Format*.
- ITU-T Recommendation T.416 (1993) | ISO/IEC 8613-6:1994, *Information technology – Open Document Architecture (ODA) and interchange format: Character content architectures*.
- ITU-T Recommendation T.417 (1993) | ISO/IEC 8613-7:1994, *Information technology – Open Document Architecture (ODA) and interchange format: Raster graphics content architectures*.
- ITU-T Recommendation T.418 (1993) | ISO/IEC 8613-8:1994, *Information technology – Open Document Architecture (ODA) and interchange format: Geometric graphics content architectures*.

2.2 Paired Recommendations | International Standards equivalent in technical content

- CCITT Recommendation X.208 (1988), *Specification of Abstract Syntax Notation One (ASN.1)*;
ISO/IEC 8824:1990, *Information technology – Open Systems Interconnection – Specification of Abstract Syntax Notation One (ASN.1)*.

2.3 Additional references

- ISO 2846:1975, *Set of printing inks for offset printing – Colorimetric characteristics*.
- ISO 3664:1975, *Photography – Illumination conditions for viewing colour transparencies and their reproductions*.
- ISO/IEC 6937:1994, *Information technology – Coded graphic character set for text communication – Latin alphabet*.
- CIE Publication S002:1986, *Colorimetric Observers*.
- CIE Publication 15.2:1986, *Colorimetry*.
NOTE – CIE = International Commission on Illumination.
- ANSI PH2.30-1985, *Graphic Arts and Photographic Viewing Conditions for Color Prints, Transparencies and Photomechanical Reproduction*.
- SMPTE Recommended Practice RP37:1969, *Color Temperature for Color Television Studio Monitors*.
- SMPTE Recommended Practice RP145:1986, *Color Monitor Colorimetry*.

3 Definitions

For the purpose of this Recommendation | International Standard the definitions given in ITU-T Rec. T.411 | ISO/IEC 8613-1 apply.

NOTE – Definitions of some additional terms relating to colour are in Annex I.

4 Abbreviations

For the purpose of this Recommendation | International Standard, the abbreviations given in ITU-T Rec. T.411 | ISO/IEC 8613-1 apply.

5 Conventions

For the purpose of this Recommendation | International Standard the conventions given in ITU-T Rec. T.411 | ISO/IEC 8613-1 apply.

The following additional conventions are used within this Recommendation | International Standard.

5.1 Subordinate

References to the immediate subordinates of an object always use the form “immediate subordinates” or “immediately subordinate”.

When the term “subordinate” is used without this qualifier it means subordinate to any hierarchical level.

5.2 Superior

References to the immediate superior of an object always use the form “immediate superior” or “immediately superior”.

The term “nearest superior” is used in conjunction with an object and with a particular qualification to mean the first superior going up the hierarchy which satisfies the qualification, i.e. first checking the immediate superior, then its immediate superior, and then each superior in turn in this hierarchic order. For example, “nearest superior that belongs to the specified object class” means the first superior going up the hierarchy from the object which is of the specified object class.

When the term “superior” is used without either of these qualifiers it means superior to any hierarchical level.

5.3 Next and following

When the term “next” or “immediately following” is used in conjunction with an object then it means the object immediately following this object in sequential order (see 7.1.2). Thus, “next layout object” means the immediately following layout object in sequential layout order; “next logical object” means the immediately following logical object in sequential logical order.

When the term “following” is used in conjunction with an object without the qualifier “immediately” then it means an object at any position later in the sequential order than this object.

In some cases the term “next” is used together with a further qualification, for example, “next layout object with the same layout category” means the first of the following layout objects in sequential layout order which has the same layout category as the specified layout object.

5.4 Preceding

When the term “immediately preceding” is used in conjunction with an object then it means the object immediately preceding this object in sequential order (see 7.1.2). Thus, “immediately preceding layout object” means the immediately preceding layout object in sequential layout order; “immediately preceding logical object” means the immediately preceding logical object in sequential logical order.

When the term “preceding” is used in conjunction with an object without the qualifier “immediately”, then it means an object at any position earlier in the sequential order than this object.

In some cases the term “preceding” is used together with a further qualification, for example, “preceding layout object with the same layout category” means the last of the preceding layout objects in sequential layout order which has the same layout category as the specified layout object.

6 Architectural principles

6.1 Architectural concepts

For the purpose of the ITU-T Rec. T.410-Series | ISO/IEC 8613, a document is an amount of structured information that can be interchanged as a unit.

This Specification provides the means to represent the structures of documents in two major forms:

- in a *formatted form* which allows documents to be presented (for example, printed or displayed) as intended by the originator;
- in a *processable form* which allows further processing of documents by the recipient, such as editing and formatting as intended by the originator.

It also provides the means to represent a document in *formatted processable form* to satisfy both purposes.

A document contains information that relates to its content and structure.

The *content* of a document consists of any type of information that is intended for human perception, for example, content elements that can be presented in a two-dimensional form, such as printed on paper or displayed on a screen.

The structural information included in a document is provided in order to:

- delimit portions within a document, such as areas for the imaging of different types of content elements;
- delimit portions of a document that have a logical meaning, such as chapters, paragraphs;
- use different types of coding for the different content types;
- allow processing of the document.

The rules for defining the structure and representation of documents are collectively called the *document architecture*.

The document architecture comprises a structural model and a descriptive representation. The *structural model* describes the structural elements of a document and the relationships among these elements. The *descriptive representation* describes how the elements of a document and the properties of these elements are represented.

The structural model and the descriptive representation present complementary views of a document. Both are necessary in order to distinguish between the structural aspects of a document and the data structures which represent those structural aspects. Also, the descriptive representation shows how additional information not described by the structural model is represented. This information includes styles and the document profile. Styles contain information relating to the layout and presentation of a document (see 6.3.5). The document profile includes information relating to the document as a whole (see 6.3.7).

This Specification does not require all parts of the document architecture to be present in any particular document.

6.2 Structural model of a document

The structural model introduces the *structural elements* of the document architecture.

6.2.1 Specific structures

The structural model of a document provides for two different but complementary views of the content of a specific document:

- the logical view associates content with structural elements such as chapters, appendices, headings, paragraphs, footnotes and figures;
- the layout view associates content with structural elements relating to presentation media, such as pages and areas within pages.

This is illustrated in Figure 1.

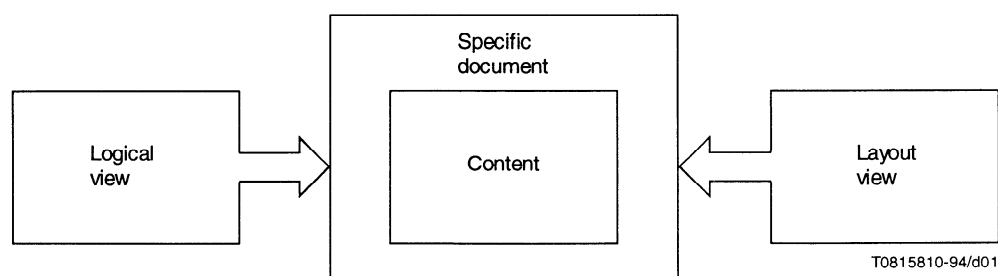


Figure 1 – Views of a document

These structural elements of a specific document are called *objects*. Each view associates the same document content with a separate structure which consists of a hierarchy of objects.

Thus

- the *specific logical structure* associates the content of a document with a hierarchy of *logical objects* and provides for the representation of documents in processable form;
- the *specific layout structure* associates the content of a document with a hierarchy of *layout objects* and provides for the representation of documents in formatted form.

The *specific structure* consists of the specific logical structure and/or the specific layout structure.

The structures are independent of the types of content within a document.

There are two types of relationships among objects in a structure, namely:

- *structural relationships* which specify the hierarchical structure of the objects;
- *non-hierarchical relationships* which specify other relationships, such as cross-references to figures or footnotes.

6.2.2 Content of a document

The structural model of a document partitions the content into structural elements called *content portions*.

The information within each content portion must pertain to a particular type of content and the structure of this information is defined by a content architecture. A *content architecture* consists of the definition of a set of *content elements*, control functions and attributes, with their coded representation, and of the rules for the application of the attributes and control functions to the content elements.

Selection of content architecture depends on the type or types of content elements to be represented. A single document may contain content portions pertaining to different content architectures. There shall be only one content architecture per content portion.

A content architecture does not identify any logical or layout objects of a document. All structural information and all logical and layout objects are specified by the logical and layout structures of the document architecture.

The document architecture supports the incorporation of the content architectures that are specified in other parts of the ITU-T Rec. T.410-Series | ISO/IEC 8613 as well as supporting the incorporation of content architectures for any other types of content. The document architecture has been designed to be separate from the content architectures. This is achieved since the structural model has been designed such that the properties of the structural elements that are used to describe a document are independent of the types of content that may be associated with those structural elements. Both document architecture and one or more content architectures are needed to represent a document.

The interface between the document architecture and the content architecture(s) is defined in clause 8.

This interface allows the use of any of the content architectures defined in other parts of the ITU-T Rec. T.410-Series | ISO/IEC 8613 with the document architecture. Other types of content may be used with the document architecture provided a content architecture is defined to this interface.

6.2.3 Generic structures

The generic structure of a document provides for the representation of characteristics common to, and relationships between, a number of objects within a document. In the most comprehensive case it provides for the representation of the common characteristics of a group of documents.

6.2.3.1 Object classes

In many documents, there may be sets of objects with common characteristics; for example:

- logical objects representing sections, consisting of a sequence of subordinate objects representing paragraphs, with the same characteristics;
- pages with the same headers and footers.

An *object class* is a structural element of the document that models such a group of common characteristics.

The ITU-T Rec. T.410-Series | ISO/IEC 8613 does not define particular object classes; however it provides the means by which object classes can be defined.

Any content portions associated with an object class are called *generic content portions*.

For convenience of reference, the term *component* is used to refer collectively to an object or an object class.

6.2.3.2 Document classes

A *document class* is used to represent the common characteristics of a group of documents; for example, a set of reports with common paragraphs and common layout. A document class can be used to maintain consistency of an entire document during editing and/or formatting of the document.

The *generic logical structure* provides for the representation of the common characteristics of logical objects of a document class and the *generic layout structure* provides for the representation of the common characteristics of layout objects of a document class.

The generic logical structure consists of all the logical object classes and associated generic content portions of a document.

The generic layout structure consists of all the layout object classes and associated generic content portions of a document.

The *generic structure* consists of the generic logical structure and/or the generic layout structure.

The ITU-T Rec. T.410-Series | ISO/IEC 8613 does not define particular document classes; however it provides the means by which document classes can be defined.

6.3 Descriptive representation of a document

The descriptive representation introduces the descriptive elements of the document architecture.

For the purpose of interchange, a document is represented as a collection of *constituents*, each of which is a set of attributes.

Within the ITU-T Rec. T.410-Series | ISO/IEC 8613 each *attribute* is identified by a name and has a value, which either represents a characteristic of a structural element or a relationship with other constituents.

The ITU-T Rec. T.410-Series | ISO/IEC 8613 defines these constituents and attributes.

Those constituents that are counterparts to the structural elements are termed *descriptions*.

6.3.1 Constituents of a document

The following *types of constituent* are defined:

- document profile;
- logical object class description;
- layout object class description;
- logical object description;
- layout object description;
- content portion description;
- presentation style;
- layout style;
- sealed document profile description;
- enciphered document profile description;
- pre-enciphered document body part description;
- post-enciphered document body part description.

Each constituent is characterised by its attributes; within a constituent all attribute names are unique.

The *document body* is a term used to describe all the constituents of a document except for the document profile.

6.3.2 Content portion descriptions

Each content portion within a document is characterised by a set of attributes called a *content portion description*.

Any content portion description associated with an object class description (see 6.3.4), for example a logo block or a standard paragraph, is called a *generic content portion description*.

6.3.3 Object descriptions

Each object within a structure is characterised by a set of attributes called an *object description*.

Each attribute has a value and may represent one of the following:

- a) a characteristic of the object itself;
- b) a structural relationship which specifies a hierarchical relationship between the object and other objects in the same structure;
- c) a non-hierarchical relationship between:
 - the object and other objects in the same structure;
 - the object and other objects in different structures included in the same document;
 - the object and object classes included in the same document.

An object description may also be referred to as a *primary description*, in particular when it is required to distinguish between object descriptions and alternative descriptions (see 6.3.6).

6.3.4 Object class descriptions

Each object class within a document is characterised by a set of attributes called an *object class description*.

An object class description for a logical object class is called a *logical object class description* and an object class description for a layout object class is called a *layout object class description*.

In general, each attribute of an object class description specifies a rule to determine the value of a corresponding attribute of an object description for an object of the object class concerned.

Object class descriptions can be used for the following purposes:

- a) to improve transmission efficiency;
- b) to maintain the internal consistency of a document when it is modified;
- c) to facilitate the creation of objects and documents.

Object class descriptions can be used either individually or collectively.

In either case, the object class descriptions may have generic content portion descriptions associated with them.

In the case of individual use, each object class description is used for transmission efficiency and/or to facilitate creation of objects. Each such object class description consists of a set of attributes representing the common characteristics of objects of the object class.

The set of object class descriptions corresponding to this case of individual use is termed a *factor set* of object class descriptions.

In the case of collective use, two situations must be considered.

In the first situation, some or all object class descriptions are related to each other in a particular hierarchical structure. This collection serves to facilitate creation of sets of hierarchically related objects within a specific structure, but need not fully specify all possible specific structures that may be created. Such a collection is called a *partial generator set* of object class descriptions.

In the second situation, all object classes are related to each other such that they fully control the generation of specific structures. This collection serves to maintain consistency of an entire document during editing or formatting of the document. During creation and modification of the document, each logical object class description specifies characteristics of the objects that may be created and how these objects may build up the possible specific logical structures of the entire document. Similarly, during document layout, each layout object class description influences the creation of the possible specific layout structures.

The set of object class descriptions corresponding to this second situation is termed a *complete generator set* of object class descriptions.

A generic logical structure or a generic layout structure whose representation includes a partial generator set of object class descriptions is termed a *partial generic logical structure* or a *partial generic layout structure*, respectively.

A generic logical structure or a generic layout structure whose representation includes a complete generator set of object class descriptions is termed a *complete generic logical structure* or a *complete generic layout structure*, respectively.

6.3.5 Styles

In addition to logical and layout component descriptions, a document may contain a number of layout styles and presentation styles which are distinct from the component descriptions.

A style is a set of attributes which may be referred to from component descriptions. The effect of such a reference is to apply the attributes of the style to the component description which contains the reference. Styles may be derived from other styles. A derived style need only specify the attributes and/or attribute values that differ from those of the style from which it was derived.

A *layout style* is a set of attributes referred to from one or more logical component descriptions which during document layout guides the creation of a specific layout structure. That is, a layout style provides information for the creation of pages and, if required, guides the division of pages into separate areas, for the layout of the document content.

A *presentation style* is a set of attributes referred to from one or more basic logical and/or layout component descriptions which guides the format and appearance of the content of the document on the presentation medium.

The separation of styles from the component descriptions allows the layout and presentation of a document to be modified without affecting the logical structure.

6.3.6 Alternative descriptions

In addition to its primary description (see 6.3.3), a basic logical object or a basic layout object may be represented by one or more alternative descriptions. An *alternative description* is intended by the originator to be used by the recipient of a document in lieu of the primary description when the recipient is not capable of processing the primary description. In the case that there are multiple alternative descriptions, a preference order is defined between these alternative descriptions. An alternative description for a basic object may specify the same or a different set of attributes from those specified by the primary description.

Alternative descriptions provide fallback mechanisms in this Specification. Possible uses include: fallback content architecture classes (e.g. providing a raster image as a fallback for geometric graphics), compatibility with several versions of ODA, compatibility with several document application profiles, and fallback for non-basic values within a document application profile.

NOTE – Document application profiles are defined in ITU-T Rec. T.411 | ISO/IEC 8613-1.

Using an alternative description for an object implies using a different set of associated content portion descriptions, if any. An alternative description for an object in conjunction with any associated content portion descriptions is collectively called an *alternative subtree*. The primary description in conjunction with its associated content portion descriptions is called the *primary subtree*. Thus when required a primary subtree may be replaced by an alternative subtree.

A fallback for a content portion description can be supplied by providing an alternative description for the basic object with which the content portion description is associated, and associating the fallback content portion description with this alternative description.

Alternative subtrees must satisfy the constraint that substituting the alternative subtree for its primary subtree shall result in a valid document for the purposes of performing a layout process and/or an imaging process.

6.3.7 Document profile

The *document profile* consists of a set of attributes which specify characteristics of the document as a whole.

For instance, the document profile indicates which of the following are present in the document:

- logical object descriptions;
- layout object descriptions;
- logical object class descriptions;
- layout object class descriptions;
- presentation styles;
- layout styles.

The document profile specifies the document architecture class used in the document. It also specifies the ODA version, document application profile, content architectures and interchange format class used in the document.

The document profile may describe the document and its history, including information for filing and retrieval and, for example, describing the fonts used in the document.

For the convenience of the recipient, the document profile may duplicate information usually found in the document content (for example, document name, author, date, etc.). However, the document profile does not include data specific to a particular mode of transmission, such as mail, message or Teletex.

A document profile may be interchanged alone to allow a sender to test the capability of the recipient, or for the recipient to get information about the document without sending the complete document.

A complete specification of the document profile is contained in ITU-T Rec. T.414 | ISO/IEC 8613-4.

6.3.8 Protected part descriptions

A *protected part description* is a set of attributes which may be referred to from the document profile.

There are four kinds of protected part descriptions:

- sealed document profile description;
- enciphered document profile description;
- pre-enciphered document body part description;
- post-enciphered document body part description.

6.3.9 Document class descriptions

A document class is specified by a *document class description*.

A document class description consists of one of the following:

- a) a document profile;
a complete generator set of logical object class descriptions;
optionally layout styles;
optionally presentation styles;
optionally generic content portion descriptions; or
- b) a document profile;
a complete generator set of layout object class descriptions;
optionally presentation styles;
optionally generic content portion descriptions; or
- c) a document profile;
a complete generator set of logical object class descriptions;
a complete generator set of layout object class descriptions;
optionally layout styles;
optionally presentation styles;
optionally generic content portion descriptions.

A document class description defines how to generate an entire specific logical and/or specific layout structure for any document of the class.

Documents may reference one externally specified document class description (see 6.3.11). In this case, the document profile indicates whether an external-document class description is referenced by the document, and if so, which one.

6.3.10 Generic-documents

A *generic-document* description consists of one of the following:

- a) a document profile;
a complete generator, partial generator or factor set of logical object class descriptions;
optionally layout styles;
optionally presentation styles;
optionally generic content portions; or
- b) a document profile;
a complete generator, partial generator or factor set of layout object class descriptions;
optionally presentation styles;
optionally generic content portions; or
- c) a document profile;
a complete generator, partial generator or factor set of logical object class descriptions;
a complete generator, partial generator or factor set of layout object class descriptions;
optionally layout styles;
optionally presentation styles;
optionally generic content portions.

A generic-document may be interchanged and used to aid in the generation of documents. Resource-documents and external-documents are generic-documents.

6.3.11 External-documents

A document that does not contain a generic structure can refer to an *external-document*. The external-document is identified in the document profile of the document that is interchanged. The reference is ignored if any generic layout structure or generic logical structure is present in the interchanged document. Styles may be present in both the interchanged document and the external-document.

The external-document can provide any or all of :

- constituents representing a complete generic logical structure;
- constituents representing a complete generic layout structure;

and optionally,

- layout styles;
- presentation styles.

An external-document contains a document profile which supplies information for the constituents and styles in the external-document, for example, a fonts list. The information in this document profile cannot be used by the interchanged document except by reference to the constituents and styles in the external-document.

If layout or presentation styles in the interchanged document and the external-document have the same identifier then the style in the interchanged document is used and the style in the external-document is ignored.

6.3.12 Resource-documents

An object class description in a document that is interchanged may contain a reference to an object class description external to the document, in a generic-document. The latter generic-document is called the *resource-document* of the interchanged document.

The object class descriptions contained in the resource-document act as models for object class descriptions in the interchanged document.

The attributes constituting an object class description in the resource-document may supply values for the corresponding attributes of those object class descriptions in the interchanged document that refer to this object class description in the resource-document.

A resource-document may contain generic content portion descriptions to be included by reference into an interchanged document referring to this resource-document.

Thus, the relationship between an object class description in the interchanged document and the corresponding object class description in the resource-document is similar to the relationship between an object description and the corresponding object class description in the interchanged document.

Styles may be present in both the interchanged document and the resource-document. If styles in the interchanged document and the resource-document have the same identifier then references from the resource-document are to the style in the resource-document, and references from the interchanged document are to the style in the interchanged document.

A resource-document is separate from any document or documents referring to it.

6.3.13 Classes of document architecture

A *document architecture class* is a set of rules for defining the structure and representation of documents in formatted form, processable form, or formatted processable form.

Clause 13 defines three document architecture classes that may be used in conjunction with document application profiles, as defined in ITU-T Rec. T.411 | ISO/IEC 8613-1.

The three classes of document architecture are:

- a) *Formatted document architecture class* which allows for document content to be presented as intended by the originator; for example, printed or displayed. A document of this class includes a document profile and constituents representing a specific layout structure. It may also include constituents representing a generic layout structure and presentation styles.

- b) *Processable document architecture class* which allows for document content to be processed; for example, edited or formatted. A document of this class includes a document profile and constituents representing a specific logical structure. It may also include constituents representing a generic logical structure, a generic layout structure, layout styles and presentation styles.
- c) *Formatted processable document architecture class* which allows for document content to be processed as well as presented as intended by the originator. A document of this class includes a document profile, constituents representing a specific logical structure, a specific layout structure and a generic layout structure. The generic layout structure can be omitted if there is a reference to an external-document containing at least a complete generic layout structure. The document may also include constituents representing a generic logical structure, layout styles and presentation styles.

A generic-document may be assigned to one of the preceding classes, by the following rules:

- a) If the generic-document contains logical object class descriptions, and no layout object class descriptions, it is assigned to the processable document architecture class.
- b) If the generic-document contains layout object class descriptions, and no logical object class descriptions, it is assigned to the formatted document architecture class.
- c) If the generic-document contains both logical and layout object class descriptions, it is assigned to the formatted processable document architecture class.

6.3.14 Sets of constituents

A document is represented by constituents, which are grouped into sets of constituents, and which have inter-relationships, as defined in the ITU-T Rec. T.410-Series | ISO/IEC 8613.

The possible types of constituent in the descriptive representation of a document are shown in Figure 2.

Figure 2 illustrates that

- a) the document consists of one document profile and optionally a number of constituents that form the document body;
- b) the document body consists of one of the following sets:
 - 1) constituents representing the generic part of the document;
 - 2) constituents representing the style constituents of the document;
 - 3) constituents representing the specific part of the document;
 - 4) constituents representing protected parts of the document.
- c) the constituents representing the generic part of the document consist of constituents representing the generic logical structure and/or constituents representing the generic layout structure;
- d) the style constituents of a document consist of layout styles and/or presentation styles;
- e) the constituents representing the specific part of the document consist of constituents representing the specific logical structure and/or constituents representing the specific layout structure;
- f) the constituents representing the generic logical structure consist of the logical object class descriptions and any associated generic content portion descriptions;
- g) the constituents representing the generic layout structure consist of the layout object class descriptions and any associated generic content portion descriptions;
- h) the constituents representing the specific logical structure consist of the logical object descriptions and any associated content portion descriptions;
- j) the constituents representing the specific layout structure consist of the layout object descriptions and any associated content portion descriptions;
- k) if both the specific logical structure and specific layout structure are present in a document, then the content portions associated with these structures are, in general, common to both;
- l) the constituents representing the protected parts of the document consist of sealed document profile descriptions and/or enciphered document profile descriptions and/or pre-enciphered document body part descriptions and/or post-enciphered document body part descriptions.

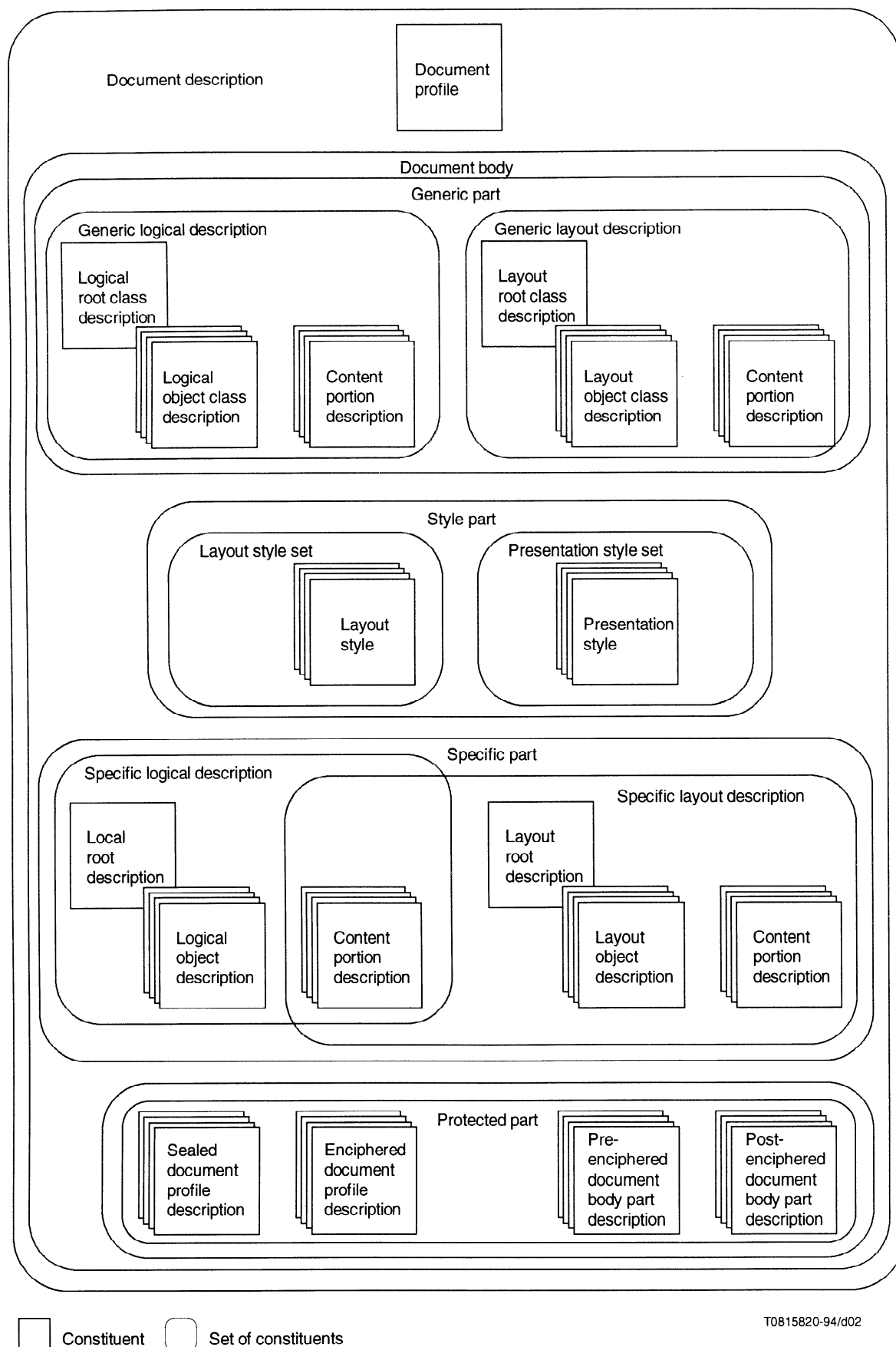


Figure 2 – Descriptive representation of a document

6.4 Document processing model

This subclause describes a conceptual model for document processing.

It addresses only those aspects of document processing that are relevant to the document architecture defined in the ITU-T Rec. T.410-Series | ISO/IEC 8613. As such, it is not a complete model of document processing since it does not specify all processing steps from document creation to document imaging.

It describes the principal operations performed on a document as a basis for understanding the semantics of the attributes defined in clause 9.

It is not intended to represent an actual implementation, nor to restrict in any way the processing that may be applied to an interchanged document.

The d

ocument processing model (see Figure 3) is summarized in this sub-clause. Three processes are shown:

- the editing process;
- the layout process;
- the imaging process.

The order of processes in the diagram is not intended to imply that they are performed sequentially in an actual implementation.

The document processing model provides for manual intervention only for carrying out editing processes on the specific logical structure and content. Manual intervention can occur, however, at many stages in the model, but it is neither explicitly allowed nor prohibited by the ITU-T Rec. T.410-Series | ISO/IEC 8613.

For example, in an actual implementation it may be possible to create and amend the logical and layout object class descriptions, and layout and presentation styles, but these processes are not included in this model since the ITU-T Rec. T.410-Series | ISO/IEC 8613 does not place any constraints on such editing.

6.4.1 The editing process

The editing process includes both document creation and document revision as these activities are indistinguishable from an architectural perspective.

This *editing process* includes the content editing process and the logical structure editing process. The *content editing process* is concerned with the creation of new content or the modification of previous content. The *logical structure editing process* is concerned with the creation of a specific logical structure or the modification of a previous specific logical structure and the allocation of content to basic logical objects. Modifications to the specific logical structure are required to conform to the rules specified in the generic logical structure, if present.

Included in the document creation and document revision activities are the creation of the generic logical structure, the generic layout structure, layout styles and presentation styles. Styles may be altered to represent changes made during the logical structure editing process. The generic layout structure may be edited in order to alter the intended layout of the document.

6.4.2 The layout process

The *layout process* includes the document layout process and the content layout process. These processes are concerned with the creation of a specific layout structure which can be used by the imaging process to present the document in human perceptible form on a presentation medium.

The *document layout process* creates a specific layout structure in accordance with the generic layout structure and information derived from the specific logical structure, the generic logical structure and layout styles (if present).

This process also determines the areas that are available within the created layout objects for the formatting of the document content (as described below) and is responsible for allocating the content to these *available areas*.

The *content layout process* is responsible for formatting (or laying out) the content portions into the available areas specified by the document layout process. This process makes use of information contained in the presentation attributes that apply to those content portions.

The document and content layout processes may interact, for instance if a content layout process cannot layout the content into the available area then it shall ask the document layout process for a larger available area and if provided with a larger available area shall repeat the content layout process.

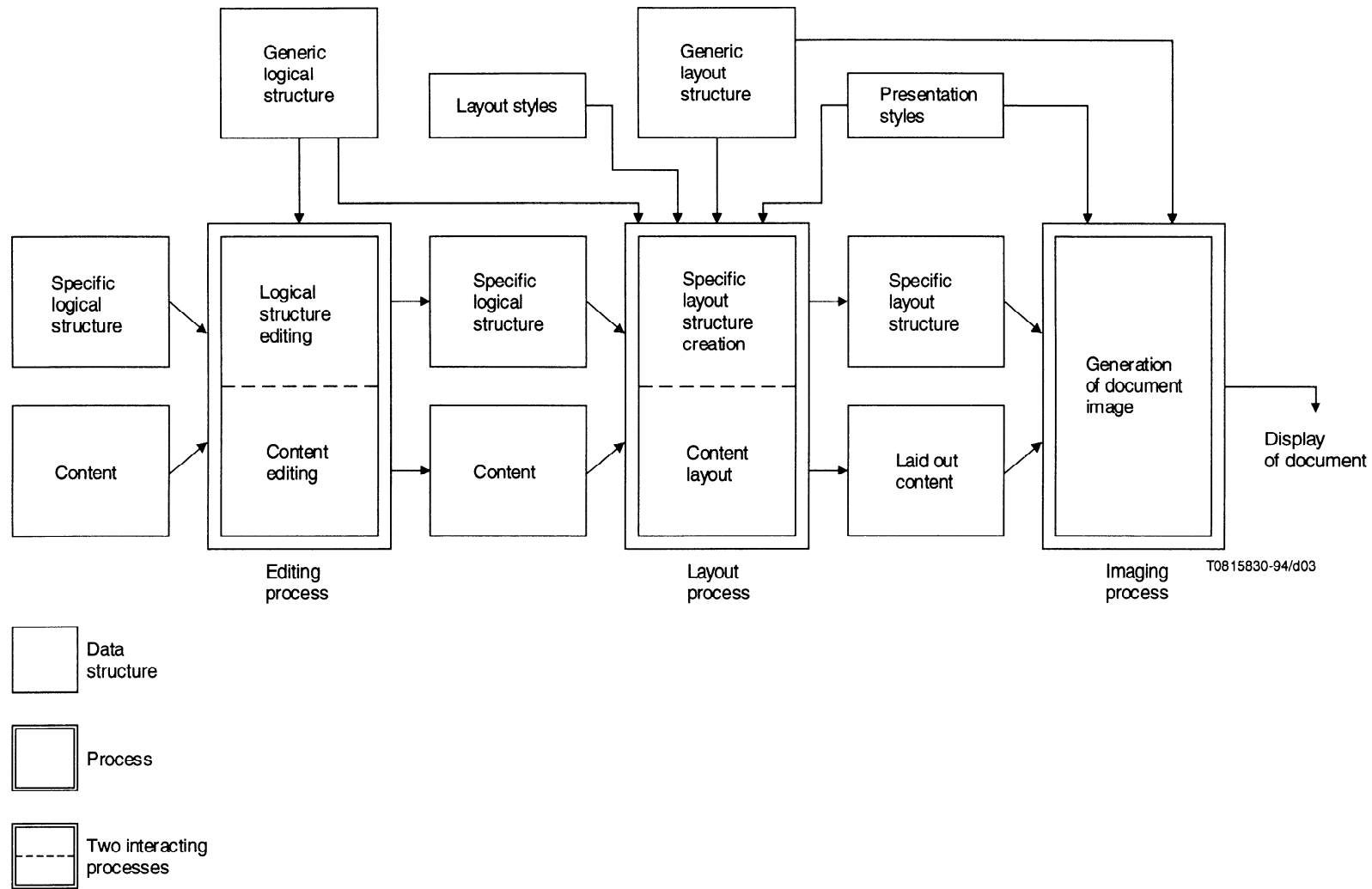


Figure 3 – Document processing model

During the layout process, presentation attributes may be derived from presentation styles referenced from the specific logical structure, from the generic layout structure, and from the generic logical structure. Derivation of presentation attributes from the logical structures has precedence over derivation from the generic layout structure. That is, those presentation attributes derived from the logical structures that differ from those derived from the generic layout structure shall be specified explicitly in the specific layout structure by the layout process.

The document layout process is described in clause 10. The content layout process depends on the content architecture pertaining to the content being laid out and is described in those parts of the ITU-T Rec. T.410-Series | ISO/IEC 8613 that are concerned with particular content architectures.

The role of the constituents that represent the generic logical structure is different in the editing process and the layout process. In the editing process, the logical object class descriptions are used to construct the specific logical structure. In the layout process, a logical object class description is used as a source of attributes and content, if any, that is common to the logical objects of the class.

6.4.3 The imaging process

The *imaging process* consists of taking a specific layout structure and a corresponding generic layout structure (if present), with associated formatted content portions and information contained in presentation styles, and displaying it on a suitable presentation medium.

Some aspects particular to the imaging process are described in clause 11. However, the imaging process is a locally defined process. Hence, apart from defining the input information required by this process, which comprises the specific and generic layout structures, referenced presentation styles and the formatted content portions, this process is not formally defined in the ITU-T Rec. T.410-Series | ISO/IEC 8613.

During the imaging process, presentation attributes are derived from the layout structures and referenced presentation styles. Any presentation attributes specified by the logical structures and referenced presentation styles are ignored.

The role of the constituents that represent the generic layout structure is different in the document layout process from that in the imaging process. In the document layout process, the layout object class descriptions are used to construct the specific layout structure. In the imaging process, a layout object class description is used as a source of attributes and content, if any, that is common to the layout objects of the class.

Some aspects of imaging depend on the content architecture pertaining to the content to be imaged and are described in those parts of the ITU-T Rec. T.410-Series | ISO/IEC 8613 that are concerned with particular content architectures.

6.5 Roles of the document architecture constituents in document processing

6.5.1 Editing process

This subclause describes the role of the various parts of the document architecture when the editing process is applied to documents of processable or formatted processable class.

6.5.1.1 Generic logical structure

The generic logical structure can be used to control the logical structure editing process that can be carried out on a document. This is accomplished by providing references from logical object descriptions to logical object class descriptions.

If a complete generator set of logical object class descriptions is present in the document then there is a reference from every logical object description to a logical object class description.

A complete generator set of logical object class descriptions controls the specific logical structure which is produced during the document creation and editing process. This is achieved by ensuring that logical objects are only created, deleted or modified according to corresponding logical object classes in the generic logical structure. In addition, the creation of a new logical object is facilitated because a logical object class description serves as a template for the logical object description.

A partial generator set of logical object class descriptions is similar to a complete generator set in facilitating the editing operations; it differs in that it guides rather than constrains the editing operations and may only apply to portions of the specific logical structure rather than all of it.

If a factor set of logical object class descriptions is present in the document, then the object classes in that structure can be used to facilitate the creation of objects in the specific logical structure that have common properties. This is achieved by providing references from a logical object description to a logical object class description. This provides a means of reducing the amount of common information in objects of the specific logical structure; this is referred to as “factorization” of information. A factor set of logical object class descriptions has no control over the structure of the specific logical structure.

6.5.1.2 Specific logical structure

The logical structure editing process involves making changes in the specific logical structure. These changes may be as follows:

- creating or deleting a logical object;
- changing the position of an object in the specific logical structure;
- changing the characteristics of an object by adding, deleting or modifying attributes specified by the object description.

Changing the characteristics of an object can also involve:

- changing the layout styles referenced by the object description;
- changing, adding or deleting a default value for an attribute in a default value list for a logical object description.

6.5.1.3 Content

The content editing process involves making changes to the content. Representing this can involve making changes to the presentation styles applicable to the basic logical object descriptions.

In order to edit the content information it must be in processable or formatted processable form. The content is changed by

- a) adding, modifying or deleting one or more content elements;
- b) adding, modifying or deleting embedded control functions.

Editing of formatted form content is outside the scope of the ITU-T Rec. T.410-Series | ISO/IEC 8613. However, all three forms of content may be specified by the content portions of the logical structure.

The editing algorithms used to change the content are not described in this model as they are outside the scope of the ITU-T Rec. T.410-Series | ISO/IEC 8613.

6.5.1.4 Generic layout structure

The generic layout structure, if present in the document, plays no direct role in the editing process.

However, the generic layout structure may be edited in order to alter the intended layout of the document. This process is of local concern only and is outside the scope of the ITU-T Rec. T.410-Series | ISO/IEC 8613.

6.5.1.5 Specific layout structure

The specific layout structure, if present in the document, plays no direct role in the editing process.

However, the specific layout structure which will be produced by the document layout process may be changed as a consequence of the editing process modifying the specific logical structure or the document content.

6.5.1.6 Layout styles

Layout styles are not used to represent controls on the editing process.

Layout styles may be altered to represent changes made during the logical structure editing process (see 6.5.1.2). This involves the addition, deletion or modification of the layout directive attributes contained in the layout styles. This can affect the layout characteristics of all the logical objects whose representation refers to the layout styles that are changed and can affect the relationships that the objects have with other objects in the document. Changing the layout styles may cause changes in the specific layout structure produced by the document layout process.

6.5.1.7 Presentation styles

Presentation styles are not used to represent controls on the editing process.

Presentation styles may be altered to represent changes made during the content editing process. This involves the addition, deletion and modification of attributes contained in the presentation styles. This can affect the layout and imaging of the content associated with all the basic logical objects whose representation refers to the presentation styles that are changed. For some presentation attributes, the same effects may be produced by editing the control functions that are embedded in the content but such editing will only alter the layout and imaging of the content in which the edited control functions are embedded.

6.5.1.8 Alternative descriptions

Alternative descriptions usually play no direct role in the editing process but may be derived automatically by the originating system from the primary descriptions. It is then the responsibility of the originating system to keep the alternative descriptions consistent with the primary description. Alternative descriptions may also have to be manually derived from the primary description (for example, descriptive text indicating the contents of the primary description); in this case it is the responsibility of the originator to maintain consistency during the editing process.

6.5.2 Layout process

This subclause describes the role of the various parts of the document architecture when the layout process is applied to documents of processable or formatted processable class. A model of this process is described in clause 10.

6.5.2.1 Generic logical structure

During the layout process, the generic logical structure may provide layout styles, presentation styles and generic content portions which may affect the creation of the specific layout structure.

Also, object classes may be referenced by the attribute "logical source" that is contained within layout object class descriptions. This results in the creation of additional layout objects that have no correspondence with any objects in the specific logical structure.

6.5.2.2 Specific logical structure

During the layout process, the sequential logical order of the objects in the specific logical structure determines the sequence in which the content of the document is considered by the layout process.

In addition, the logical objects can provide layout directive attributes which direct the document layout process and presentation attributes which direct the content layout process. These layout directive attributes and presentation attributes are specified by referencing a layout style or a presentation style, respectively.

6.5.2.3 Layout styles

During the layout process, layout styles provide layout directive attributes which direct the document layout process.

6.5.2.4 Presentation styles

During the layout process, presentation styles provide presentation attributes which direct the content layout process.

6.5.2.5 Content

During the layout process, the content is allocated to basic layout objects. The division of the content into content portions may be modified so that it is consistent with both specific structures. In addition, the content layout process may insert control functions into the content to facilitate the imaging process.

6.5.2.6 Generic layout structure

During the layout process, a complete generator set of layout object class descriptions must be available to determine a specific layout structure for the document.

The construction expressions specified by the layout object class descriptions determine all permissible specific layout structures which may be created by the layout process. Which of these permissible structures is used is determined from the specific logical structure, the generic logical structure, the content, and the layout and presentation styles.

6.5.2.7 Specific layout structure

The specific layout structure results from applying the document and content layout processes to the specific logical structure and the content, guided by the generic logical structure, the generic layout structure, layout styles and presentation styles.

6.5.2.8 Generic content

Generic content portions associated with logical object classes may be in formatted, processable or formatted processable form. When a logical object class specifying generic content portions is referred to from a basic logical object then when that basic logical object is encountered in the layout process, the generic content is laid out in accordance with the document and content layout processes and a new basic layout object is created and added, together with the created content portions, to the specific layout structure.

Generic content portions associated with layout object classes may be in formatted or formatted processable form. The layout of such content portions during the layout process does not cause additional content portions to be added to the specific layout structure. However, the content portions are indirectly associated with the specific layout structure by reference to the appropriate layout object class description.

6.5.2.9 Alternative descriptions

Alternative descriptions do not influence the reference layout process. If a system is to lay out a document that contains primary descriptions which it is not capable of processing, it may substitute alternative descriptions for those primary descriptions prior to the layout process.

A system that provides a layout process where both primary and alternative descriptions influence the layout process is outside the scope of this Specification.

6.5.3 Imaging process

This subclause describes the role of the various parts of the document architecture when the imaging process is applied to documents of formatted or formatted processable class.

6.5.3.1 Content

In the imaging process, the content, together with the specific layout structure, is used to produce an image of the document perceptible to a human.

After layout, the content may be either in formatted form or in formatted processable form. Both forms of content are suitable for imaging.

6.5.3.2 Generic layout structure

During the imaging process, the generic layout structure may provide, for the layout objects in the specific layout structure, any combination of the following:

- a) attributes that direct the imaging of the content;
- b) generic content portions.

6.5.3.3 Specific layout structure

In the imaging process, the specific layout structure, together with the content, is used to produce a human-perceptible image of the document.

The sequential order of the objects in the specific layout structure determines the precedence for imaging the content of the document by the imaging process. It is also possible to override the precedence of the layout objects by specifying an imaging order which is different from the sequential layout order.

6.5.3.4 Layout styles

In the imaging process, layout styles play no role.

6.5.3.5 Presentation styles

In the imaging process, the presentation attributes of the presentation styles referenced by the layout structures may affect the generated image of the content.

6.5.3.6 Alternative descriptions

Alternative descriptions do not influence the reference imaging process. If a system is to image a document that contains primary descriptions which it is not capable of processing, it may substitute alternative descriptions for these primary descriptions prior to the imaging process.

6.6 Security protection of parts of a document

This Specification distinguishes between the following two sets of security protected parts of a document:

- a) Parts of a document profile description.
- b) Parts of the document body consisting of complete object descriptions, object class descriptions, layout styles and presentation styles. A complete composite object description comprises the object descriptions for all subordinate objects and content portion descriptions for all associated content portions.

Two concepts of document security exist:

- a) Information indicating how the whole document shall be handled as a single unit, according to the security policy of the security domain to which the originator belongs. The originator is responsible for making the indication; the actual security handling of the document is outside the scope of both originator and recipient, and outside the scope of the ITU-T Rec. T.410-Series | ISO/IEC 8613.
- b) Information to be interchanged between the originator and the recipient on how security aspects of parts of the document shall be handled. The handling of this component of security is under the control of the originator and the recipient.

For parts of the document, concept b) covers the properties of:

- confidentiality;
- integrity;
- authenticity, including signature and non-repudiation of origin.

It is not the purpose of the ITU-T Rec. T.410-Series | ISO/IEC 8613 to specify any particular security scheme or method, but rather to provide the means in the document for a variety of possible security implementations as required by various security policies.

Two cryptographic techniques may be used to provide the security protections in the ITU-T Rec. T.410-Series | ISO/IEC 8613:

- encipherment of clear text to provide for confidentiality and possibly integrity of data;
- production of cryptographically derived information to provide for integrity and authenticity of data.

There are two phases, a generation phase and an interpretation phase, involved with security protection of parts of a document.

The generation phase enciphers or seals parts of the document and creates security information that is added to the document. The protected part descriptions are created in this phase. These descriptions contain the enciphered and the sealed versions of the protected parts of the document.

The interpretation phase decipheres enciphered protected parts or validates seals that have been created in the generation phase.

6.6.1 Intended and privileged recipients

Two categories of recipients are defined, namely intended recipients and privileged recipients.

An intended recipient of a document is a recipient that is expected by the originator to receive or have access to the document.

A privileged recipient of a document is a recipient that, in addition to being an intended recipient, has the right to perform certain security-related operations intended for that particular recipient, such as to interpret specified enciphered parts of the document and to perform integrity and authenticity checks on specified parts of the document.

6.6.2 Protecting parts of the document profile

Protected parts of a document profile are specified in two sets of document profile descriptions, the set of enciphered document profile descriptions and the set of sealed document profile descriptions.

The sealed document profile descriptions are for integrity, authenticity and non-repudiation of origin, one for each seal of parts of the document profile.

The enciphered document profile descriptions are for confidentiality, one for each encipherment of parts of the document profile.

All information about each sealed or enciphered document profile description is found in the document profile.

6.6.3 Protecting parts of the document body

The parts of the document body that may be protected are object class descriptions, object descriptions, layout styles and presentation styles.

Confidentiality is based on encipherment. Integrity, authenticity and non-repudiation of origin are based on a seal.

For encipherment:

- The protection of a composite object description implies that itself and all its subordinates and all the content portions directly referred to from any of its subordinates are protected.
- The protection of a basic object class description implies that the basic component itself and all content portions directly referred to from it are protected.
- The protection of a basic object description implies that for processable form and formatted form documents, itself and all the content portions directly referred to from it are protected. In a formatted processable form document all content portions directly referred to from a protected basic object description in one of the structures will be protected, but it may be that only some of the content portions referred to from a basic object description in the other structure are protected.
- The protection of a composite object class description, layout style or presentation style implies that all of it is protected.

For sealing:

- The protection of a composite object description implies that itself, all its subordinates and all the content portions directly referred to from any of its subordinates are protected.
- The protection of a basic object class description implies that the basic component itself and all content portions directly referred to from it are protected.
- The protection of a basic object description implies that for processable form and formatted form documents, itself and all the content portions directly referred to from it are protected. In a formatted processable form all content portions directly referred to from a protected basic object in one of the structures will be protected, but it may be that only some of the content portions referred to from a basic object description in the other structure are protected.
- The protection of a composite object class description, layout style or presentation style implies that all of it is protected.

For confidentiality enciphered document body parts are defined.

There are two sets of enciphered document body part descriptions: one set for pre-enciphered document body parts and one set for post-enciphered document body parts.

A pre-enciphered document body part description is provided for each encipherment performed before the layout process, and a post-enciphered document body part description is provided for each encipherment performed after the layout process.

All information about each pre- or post-enciphered document body part is found in the document profile.

For integrity, authenticity and non-repudiation of origin seals are provided. The seals and all information about them are found in the document profile.

7 Document structures and colour

7.1 Specific structures

7.1.1 General principles

The specific layout and specific logical structures of a document are hierarchical structures of objects.

The object at the highest level in the hierarchy of the specific layout structure is called the document layout root and the object at the highest level in the hierarchy of the specific logical structure is called the document logical root.

A *composite object* is an object that has one or more subordinate objects. The structural relationships of a composite object identify its immediate subordinates.

The minimum number of hierarchical levels below the highest level in either structure is one. Thus, the document root object is always a composite object. The actual number of levels is variable and depends upon a given document.

In general, at any level in a specific structure, the subordinates of any composite object can consist of any number and combination of basic objects and composite objects.

Basic objects are objects that have no subordinate objects.

Basic objects are also distinguished from composite objects since basic objects are the only objects with which the content of a document is *associated* (see 7.1.3). Every basic object has content, either in the form of one or more specific content portions, in the form of an expression for generating content or derived from an object class to which the basic object belongs.

Each object in a structure is of a certain object type. The object types that can occur within a specific logical structure or a specific layout structure are defined in 7.2 and 7.3, respectively. The *object type* determines which attributes are applicable to an object description.

The particular attributes and attribute values comprising the object description characterise the object, that is, the attributes specify the characteristics of the object itself and specify the relationships that it has with other components in the document.

Each object in a structure is uniquely identified within that structure. The structural relationships between a composite object and its immediate subordinates are defined in the composite object description.

7.1.2 Sequential order

In the case that more than one immediate subordinate is identified by a composite object then the composite object description specifies an ordering of these immediately subordinate objects. This subordinate ordering is used to define a unique sequential ordering of all the objects in a structure.

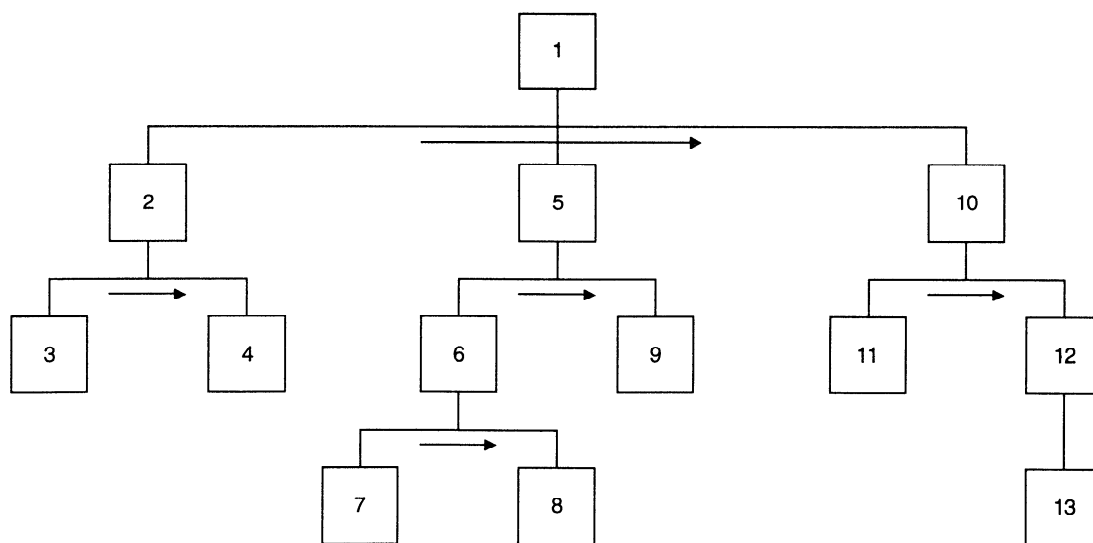
This sequential order is such that each object in the structure is succeeded by all of its immediate subordinates, before any other objects with the same immediate superior. Each of the immediate subordinates is followed by all of its immediate subordinates, before proceeding to the next immediate subordinate in sequence. The immediate subordinates occur in the subordinate ordering defined within the object description.

The sequential order of the specific logical structure is called the sequential logical order; the sequential order of the specific layout structure is called the sequential layout order.

The *sequential logical order* defines the order in which the logical objects are intended to be processed by the layout process (see clause 10).

The *sequential layout order* defines the order of precedence for imaging unless this is overridden by the specification of an imaging order in the object description (see 11.1).

An example of the sequential order of a structure is shown in Figure 4, where the numbers indicate the sequential order.



Note – Arrows define sequential order.

T0815840-94/d04

Figure 4 – Example of sequential order

7.1.3 Content portions

The content of a document is divided into content portions to allow the document architecture to address each content portion as a unit. The amount of content to be placed into a content portion may vary from no content elements to the entire document content. The maximum number of consecutive content elements that have the same properties and characteristics will typically be placed together in one content portion. For example, a content portion may represent a heading, paragraph, picture or perhaps just that amount of content that needs to have particular constraints regarding its imaging or processing.

Each content portion is structured according to a single content architecture. When a document contains only a single specific structure, either a specific layout structure or a specific logical structure, each content portion in the specific structure is associated with a single basic object.

Each basic object may have more than one associated content portion, provided that all of these content portions pertain to the same content architecture. In the case that more than one content portion is associated with a basic object then the object description specifies the ordering among these content portions.

7.2 Specific logical structures

The specific logical structure provides a means of modelling the structure of a document in terms of logical objects which have meanings that are significant to the application or user.

For example, the specific logical structure can be used to model a document in terms of chapters, sections, paragraphs and footnotes. The ITU-T Rec. T.410-Series | ISO/IEC 8613 does not standardise such particular kinds of logical objects. However, it does provide the means by which any such construction can be defined in terms of logical objects to be interpreted as chapters, sections, paragraphs or footnotes, etc.

The significance of this form of definition is that instead of standardising a few particular kinds of logical objects it allows any number of differing kinds of logical objects to be constructed, so as to meet a wide variety of needs.

NOTE 1 – Particular kinds of logical objects are standardised by the various Recommendations | International Standards for document application profiles. Each document application profile defines particular kinds of logical objects, to meet the needs of that profile. (Document application profiles are defined further in ITU-T Rec. T. 411 | ISO/IEC 8613-1.)

Logical objects provide the means to specify the specific logical structure of a document that is relevant to a particular application of the ITU-T Rec. T.410-Series | ISO/IEC 8613. For example, the specific logical structure may be required in the processing of a document, such as editing and laying out the document.

Clause 9 defines the attributes that are used to describe the characteristics of logical objects, including the use of presentation styles and layout styles.

The objects that can occur within a specific logical structure of a document are of the following object types:

- document logical root;
- composite logical object;
- basic logical object.

The allowable hierarchical relationships between logical objects are shown in Figure 5.

NOTE 2 – The notation used in Figure 5 is that defined in Annex A.

7.2.1 Document logical root

The *document logical root* is the highest level object in the hierarchy of the specific logical structure. It is a composite object whose immediate subordinates consist of any number and combination of basic logical objects and composite logical objects.

7.2.2 Composite logical objects

A *composite logical object* is a composite object that is immediately subordinate either to the document logical root or to a composite logical object of a hierarchically higher level. Its immediate subordinates consist of any number and combination of composite logical objects and basic logical objects. Thus, the number of hierarchical levels between a basic logical object and the document logical root may vary from one basic logical object to another within any particular document.

The use of composite objects is optional. Content portions cannot be directly associated with composite logical objects.

The hierarchical relationships between a composite logical object and its subordinates can express a logical relationship between that composite logical object and its subordinates that has a significance relevant to a particular application of the ITU-T Rec. T.410-Series | ISO/IEC 8613. For example, these relationships may be used to specify which sections, paragraphs and diagrams relate to a certain chapter in a document.

7.2.3 Basic logical objects

A *basic logical object* is a basic object that is immediately subordinate either to the document logical root or to a composite logical object.

A basic logical object is a container for portions of the document content. Zero, one or more content portions are associated with a basic logical object. In the case that zero content portions are associated, the content is either specified in the form of an expression for generating content or is derived from a basic object class description.

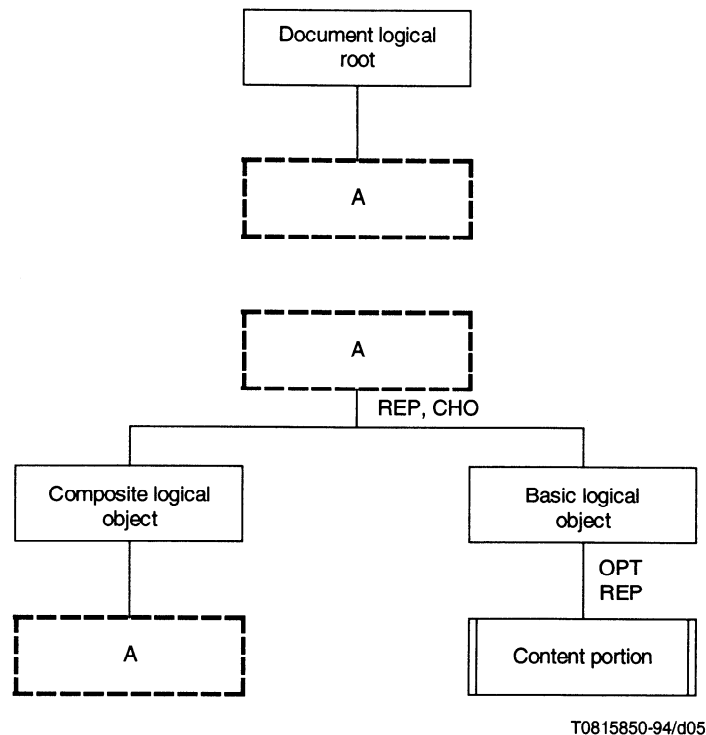


Figure 5 – Permissible specific logical structures

7.3 Specific layout structure

The specific layout structure provides a means of modelling the structure of a document in terms of layout objects which have meanings that are significant for the layout process and the imaging process.

Clause 9 defines the attributes that are used to describe the characteristics of layout objects, including the use of presentation styles.

7.3.1 Objects of the layout structure

The objects that can occur within the specific layout structure of a document are of the following object types:

- document layout root;
- page set;
- composite or basic page;
- frame;
- block.

The allowable hierarchical relationships between layout objects are shown in Figure 6.

NOTE – The notation used in Figure 6 is that defined in Annex A.

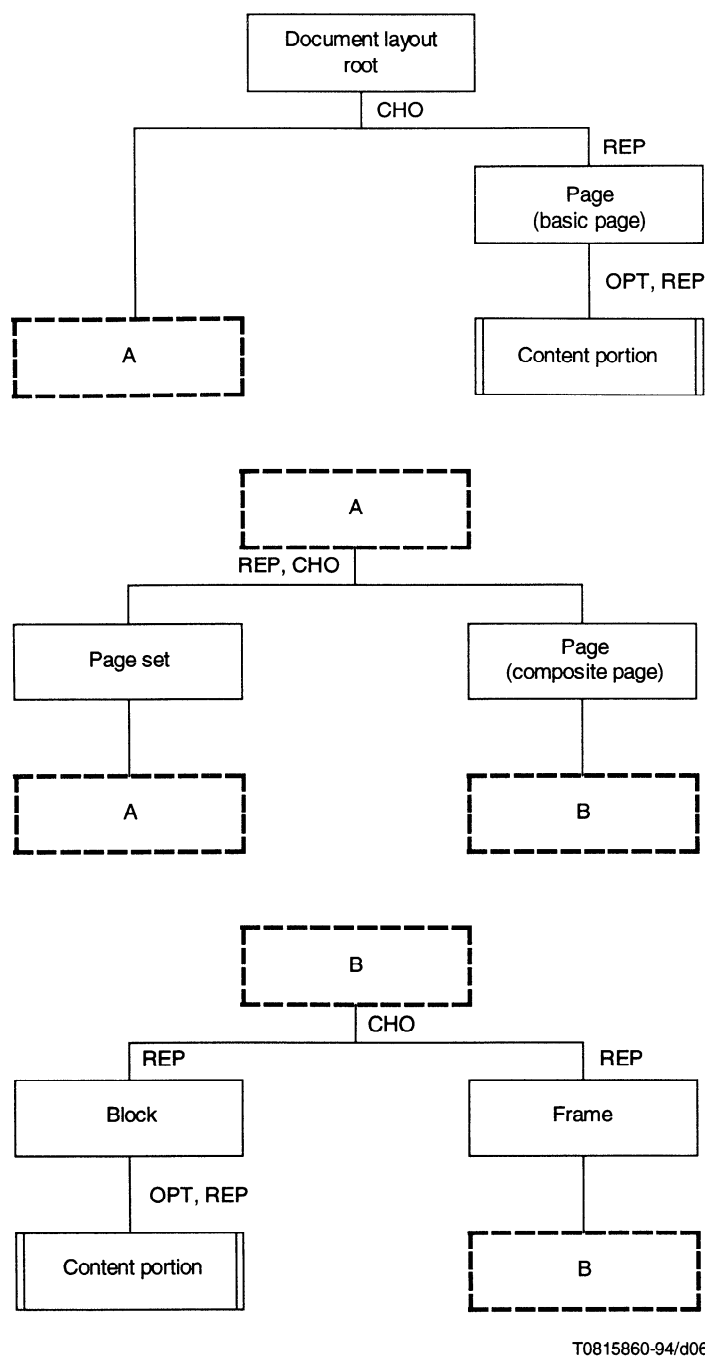


Figure 6 – Permissible specific layout structures

7.3.1.1 Document layout root

The *document layout root* is the highest level object in the hierarchy of the specific layout structure. It is a composite object whose immediate subordinates consist of any number and combination of page sets and pages.

7.3.1.2 Page set

A page set can be used to identify a number of page sets or pages (or a combination of both) as a group, for example, the pages which contain a particular section of a document.

A *page set* is a composite object that is immediately subordinate to either the document layout root or to a page set of a hierarchically higher level. Its immediate subordinates consist of any number and combination of page sets and pages. Thus the number of hierarchical levels between a page and the document layout root may vary from page to page within any particular document.

7.3.1.3 Page

A *page* is a rectangular area used as the reference area for positioning and imaging the content of the document. Its size may be smaller than, equal to or greater than the size of the presentation medium.

A page is immediately subordinate either to the document layout root or to a page set.

A page can be a basic object or a composite object and is correspondingly termed either a *basic page* or a *composite page*.

The pages within a layout structure can be either basic pages or composite pages, but not a combination of these.

A basic page has zero, one or more content portions associated with the page. In the case that zero content portions are associated, the content is derived from a basic object class description.

A composite page has as immediate subordinates either any number of frames or any number of blocks.

Basic pages and composite pages whose immediate subordinates are blocks shall only be included within documents of the formatted document architecture class.

7.3.1.4 Frame

A *frame* is a rectangular area that is contained entirely within the area of the object to which it is immediately subordinate. It is positioned so that its sides are parallel to the sides of the enclosing page. Frames are used to define areas for laying out the content associated with composite pages.

A frame is a composite object that is immediately subordinate to a composite page or to an enclosing frame. The immediate subordinates of a frame consist of either any number of frames or any number of blocks. Thus the number of hierarchical levels between a block and the enclosing page may vary from block to block within any particular page.

A *lowest level frame* on any branch of the layout structure is a frame which does not have any subordinate frames. Only lowest level frames can contain blocks.

7.3.1.5 Block

A *block* is a rectangular area that is contained entirely within the area of the object to which it is immediately subordinate. It is positioned such that its sides are parallel to the sides of the enclosing page. A block is an area for the positioning and imaging of portions of the document content.

A block is a basic object that is immediately subordinate to a page or a frame. Zero, one or more content portions are associated with a block. In the case that zero content portions are associated, the content is derived from a basic object class description.

7.3.2 Positioning layout objects

This subclause defines the rules for positioning and dimensioning frames and blocks within pages. These layout objects are used to layout and image the content of documents.

7.3.2.1 Page co-ordinate system

The positions of all layout objects subordinate to pages are specified, directly or indirectly, by means of an orthogonal *page co-ordinate system*. The origin of the page co-ordinate system is the top left corner of the page. The horizontal axis coincides with the top edge and the vertical axis coincides with the left edge of the page. The horizontal and vertical axes determine the *horizontal direction* and *vertical direction*, respectively, of the subordinate layout objects, as shown in Figure 7. This figure also defines names for the corners of a layout object: *top left corner*, *top right corner*, *bottom left corner* and *bottom right corner*. Horizontal positions are measured positively from the vertical axis to the right and vertical positions are measured positively from the horizontal axis downward.

The origin of the page co-ordinate system is the reference point used for positioning.

7.3.2.2 Positioning of frames

The reference point for positioning a frame is the top left corner of that frame. Frames are positioned relative to the reference point of the object to which they are immediately subordinate. Thus, frames that are immediately subordinate to a page are positioned relative to the origin of the page coordinate system. Frames which are immediately subordinate to another frame are positioned relative to the reference point of that frame.

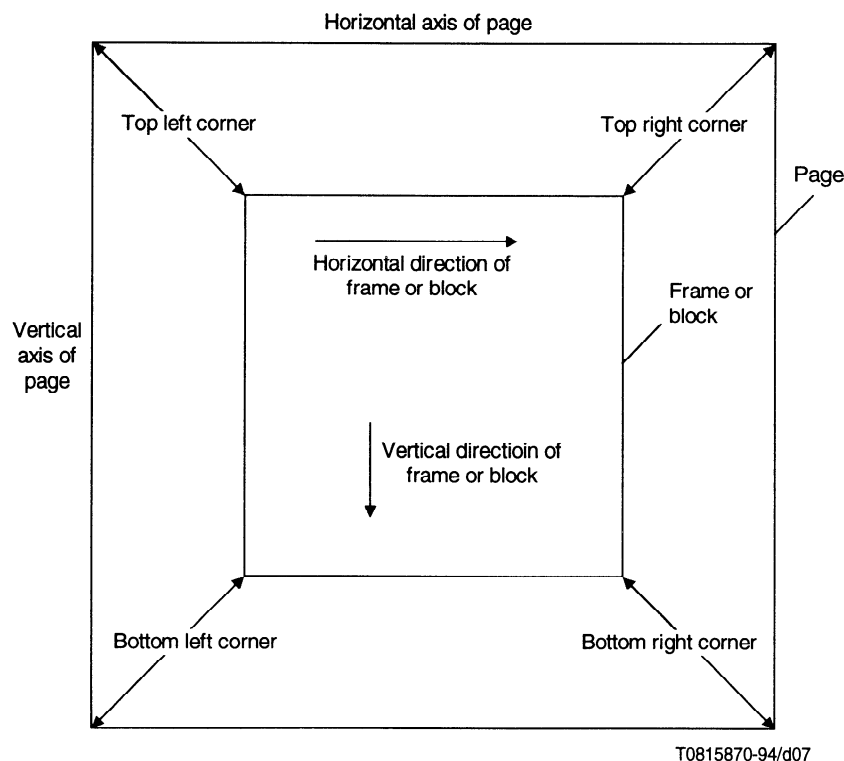


Figure 7 – The horizontal and vertical directions of layout objects

7.3.2.3 Positioning of blocks

The reference point for positioning a block is the top left corner of that block. Blocks are positioned relative to the reference point of the layout object to which they are immediately subordinate.

Thus, blocks that are immediately subordinate to a page are positioned relative to the origin of the page coordinate system. Blocks that are immediately subordinate to a frame are positioned relative to the reference point of that frame.

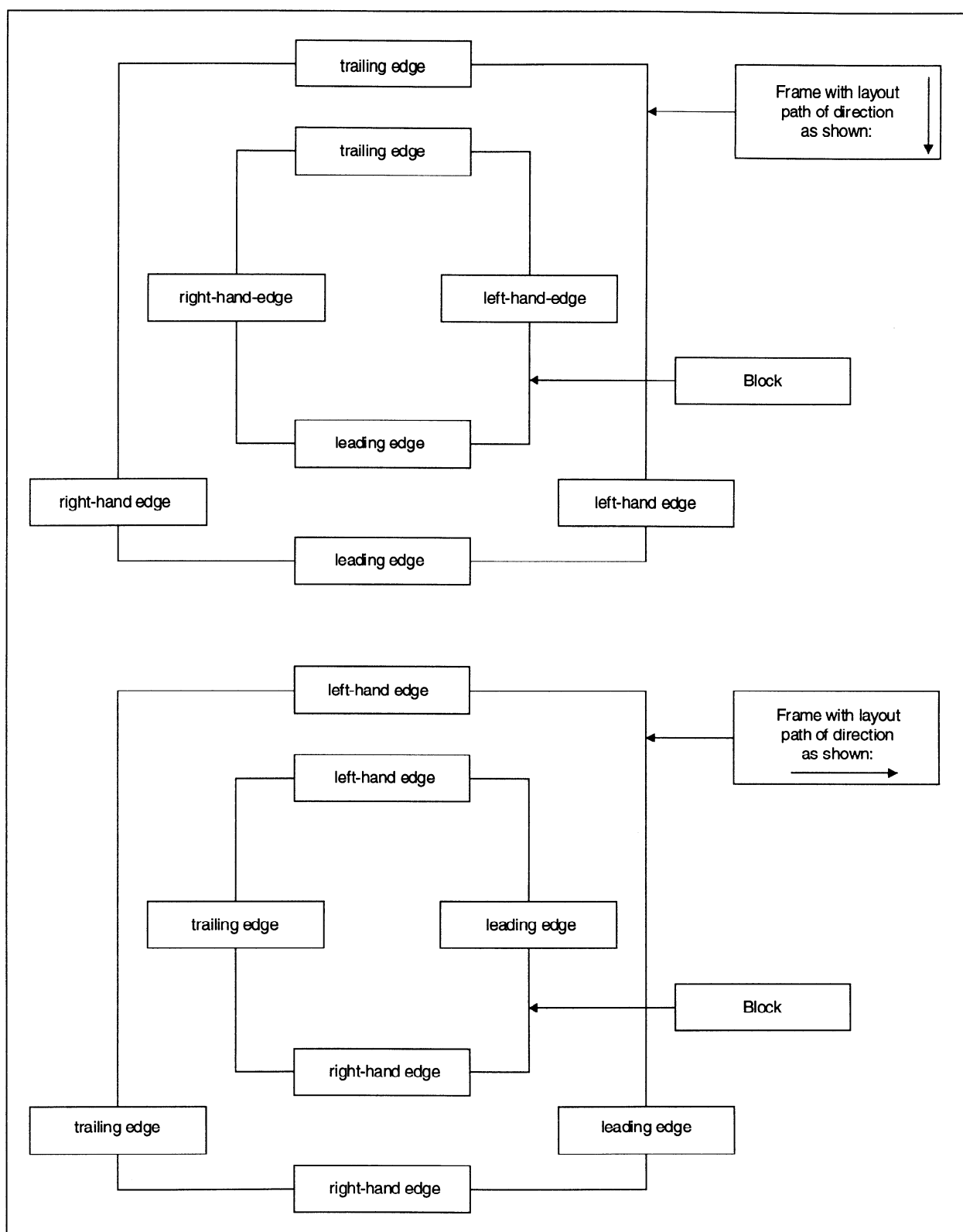
7.3.3 Naming of edges of frames and blocks

In the following definition, the layout path (see 9.4.2.2, 10.5 and 10.6) referred to is that specified for a frame, or, in the case of a block, for its immediately superior frame.

For a frame or block, the leading and trailing edges are defined as the two opposite edges of the same frame or block that are orthogonal to the direction of layout path, such that the direction from the trailing edge to the leading edge is in the same direction as the layout path.

The left-hand and right-hand edges of a frame or block are defined as the two opposite edges of the same frame or block that are parallel to the direction of layout path, such that the direction from the right-hand edge to the left-hand edge is at an angle of 90° anti-clockwise relative to the direction of the layout path.

The names of the edges of frames and blocks are illustrated in Figure 8.



T0815880-94/d08

Figure 8 – Naming of the edges of blocks and of their immediately superior frames

7.3.4 Measurement

7.3.4.1 Basic measurement unit

For the purpose of conveying the originator's intentions, all dimensions and positions are expressed in basic measurement units. The value of the basic measurement unit (BMU) is equal to 1/1200 of 25.4 mm. A locally defined scaling factor may be used to map the document to a particular imaging device.

7.3.4.2 Scaled measurement unit

All attributes and numeric control function parameters that specify absolute or relative positions and dimensions are expressed as integral multiples of a scaled measurement unit (SMU) that is equal to m/n basic measurement units. The *unit scaling factor* is specified in the form of two integers m and n by the document profile attribute "unit scaling" (see ITU-T Rec. T.414 | ISO/IEC 8613-4). When the attribute "unit scaling" is not specified, the scaled measurement unit used in the positioning and dimensioning attributes is equal to the basic measurement unit.

7.3.5 Borders of frames and blocks

Borders can be specified to be around the edges of blocks or within the edges of frames.

A *border* consists of the border freespace and the border line (see Figures 9 and 10).

The border line is described by specifying its width and line type. Line type can take the values 'solid', 'dashed', 'dot', 'dash-dot', 'dash-dot-dot' or 'invisible'.

The border freespace is described by specifying its width.

The border may be specified for a particular set of edges of the frame or block, or for all edges. The characteristics of the border for each of the edges may have a separate specification.

7.3.5.1 Borders of blocks

A border around a block is illustrated by Figure 9.

In the case of a block, the border freespace provides a region which surrounds the block between the block boundary and the border line.

The border surrounds the block and does not intrude on the block. The border is contained entirely within the area of the object to which the block is immediately subordinate. This means that the document layout process shall take the border into account in determination of the available area.

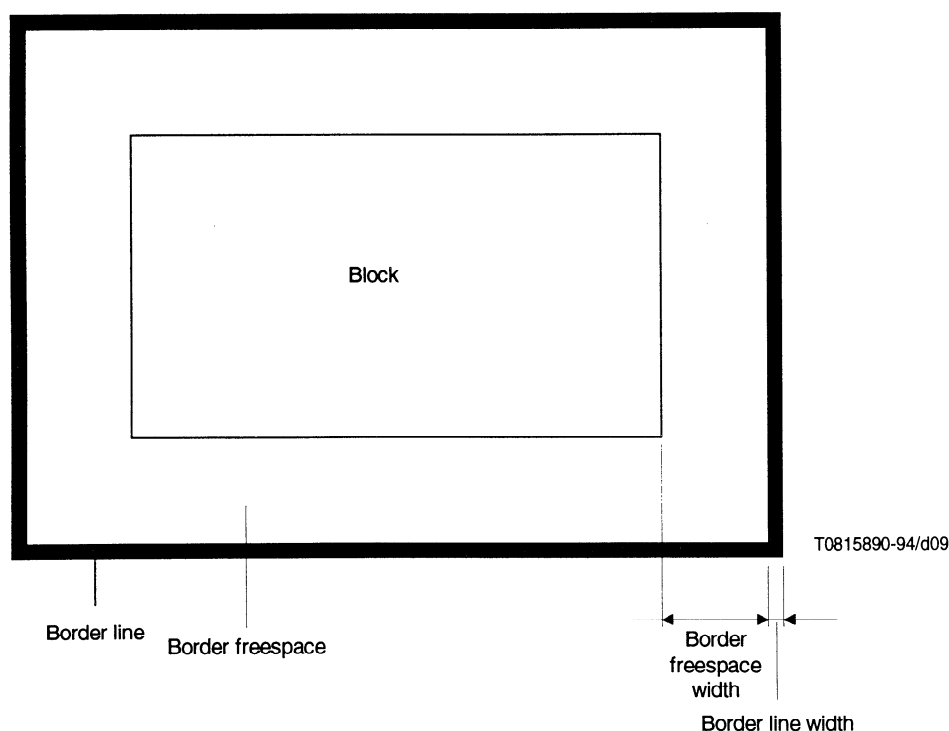


Figure 9 – Border around a block

7.3.5.2 Borders of frames

A border within a frame is illustrated by Figure 10.

In the case of a frame, the border freespace provides a region which is within the frame between the frame boundary and the border line.

The border reduces the available area within the frame for layout purposes. This means that the document layout process shall take the border into account in determination of the available area.

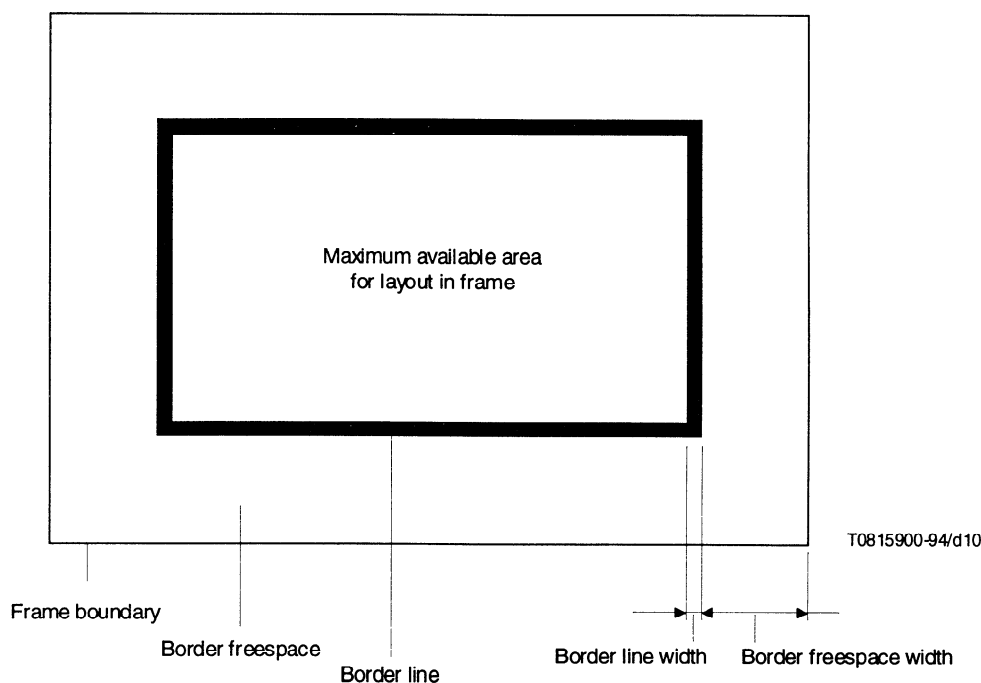


Figure 10 – Border within a frame

7.4 Documents containing both specific structures

When a document contains both a specific layout structure and a specific logical structure, each content portion in the specific structure is in general associated with both of the structures. However, some of the content portions can be associated with the specific layout structure only. These are the content portions that:

- represent formatted content corresponding to generic content portions associated with a basic logical object class of the generic logical structure or resource-document;
- are created as a result of the attribute “content generator”;
- are created as a result of the attribute “logical source” of a layout object class description.

The allocation of content portions to basic objects in the specific layout structure need not correspond to the allocation of content portions to basic objects in the specific logical structure.

For example, a basic logical object representing a paragraph would need to contain at least two content portions in the case where the paragraph was split across a page boundary. This is shown in Figure 11 if the middle basic logical object represents the split paragraph and the composite layout objects represent the two separate pages.

This is in contrast to the situation when a document contains only one specific structure, in which case there is no need for dividing the content associated with a basic component into more than one content portion.

Figure 11 also illustrates the correspondence between logical and layout objects in the case that there is both a specific logical structure and a specific layout structure.

The presentation attributes for a content portion are specified by the specific layout structure and referenced presentation styles, any specified by presentation styles referenced by the specific logical structure are ignored.

In the imaging process, the logical structure and referenced styles are ignored.

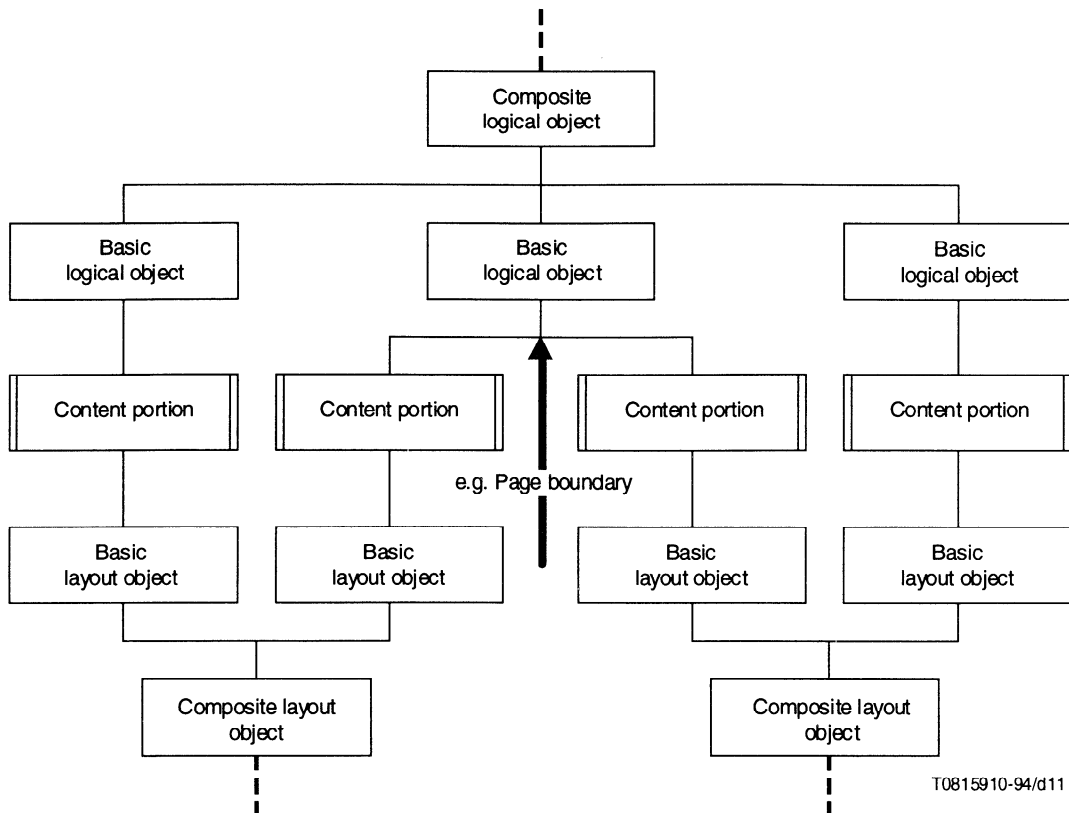


Figure 11 – Example of relationship between logical and layout objects and associated content portions

7.5 Generic structures

7.5.1 General principles

The generic structure of a document describes characteristics common to a number of objects within a document. In the most comprehensive case it describes the common characteristics of a document class.

The generic structure can be used to

- improve transmission efficiency by factorisation;
- maintain the internal consistency of a document by providing the recipient with the structural information necessary to edit and/or lay out the document as intended by the originator;
- facilitate the creation of objects and documents by the recipient as prepared by the originator.

The generic structure consists of a generic logical structure and/or a generic layout structure.

Within the set of constituents representing the generic structure each object class description consists of attributes which parallel the attributes of object descriptions. These attributes can be used to determine the value(s) for the attributes of the corresponding object descriptions. The object class descriptions may also contain references to layout styles and presentation styles.

If an attribute value is specified explicitly in an object description then this overrides any value that may be derived from the corresponding attribute of an object class description.

Each object class in a generic structure specifies the object type of all the objects that are derived from the object class. There may be many object classes of the same object type. The object types that can occur are specified in 7.5.6 and 7.5.7.

Each object class in a generic structure is uniquely identified within that structure.

Object classes for basic objects may specify content, either in the form of generic content portions or by specifying an expression to generate content.

7.5.2 Generator and factor sets of object class descriptions

A complete generator set of logical object class descriptions can be used to control the editing process. In a document with such a complete generator set every logical object description references a logical object class description.

In addition, the generation of immediate subordinates is completely derived from the referenced logical object class descriptions.

A partial generator set of logical object class descriptions can be used to guide the generation of sub-hierarchies of the specific logical structure during the editing process. In a partial generator set, some composite logical object class descriptions may guide the possible subordinates for objects of the class, while others do not.

A factor set of logical object class descriptions can be used to guide the creation of logical objects during the editing process. In a factor set, the composite logical object class descriptions do not constrain the possible subordinates for objects of the class. Such a factor set need not include an object class description corresponding to the document logical root.

A complete generator set of layout object class descriptions can be used to control the layout process. In a document with such a complete generator set every reference from the layout directive attributes, either by layout reference or layout category, is satisfied in the layout object class descriptions.

If the constituents representing the specific layout structure are present together with such a complete generator set then every composite layout object description references a layout object class description; and the generation of immediate subordinates is completely derived from the referenced layout object class descriptions. A block may, but need not, reference a layout object class description.

A partial generator set of layout object class descriptions can be used to guide the generation of sub-hierarchies of the specific layout structure. In a partial generator set, some composite layout object class descriptions may guide the possible subordinates for objects of the class, while others do not.

A factor set of layout object class descriptions can be used to guide the creation of layout objects. In a factor set, the composite layout object class descriptions do not constrain the possible subordinates for objects of the class. Such a factor set need not include an object class description corresponding to the document layout root.

In any of the preceding cases, object class descriptions achieve factorisation by holding information that is common to a number of object descriptions. Such factorisation of attribute values, by avoiding replication, can result in improved transmission efficiency and can also facilitate the creation of objects by the recipient. Such factorisation applies to content when an object class description either specifies generic content portion description(s) or specifies an expression to generate content.

7.5.3 Complete generator sets of object class descriptions

A complete generator set of layout object class descriptions form the nodes of a single directed graph.

A complete generator set of logical object class descriptions form the nodes of a set of directed graphs. This set consists of a primary graph and, optionally, one or more secondary graphs, each corresponding to the logical object class descriptions referenced by the attribute "logical source" (see 9.4.2.5).

Each directed arc of each graph connects the node from which it starts with each of the nodes which represent a possible class of immediately subordinate objects that can be generated. For any node each possible class of immediately subordinate objects is represented by a directed arc starting from the node and ending on a node representing the class of the immediately subordinate objects.

The rules describing the possible hierarchical relationships between a composite object and its immediate subordinates, which correspond to these arcs, are defined in the object class description of the composite object.

Each graph is connected such that there is a single root node which is such that:

- it is possible to reach every other node on a path through the graph following the directed arcs;
- there is no path through the graph following the directed arcs from which it is possible to reach this single root node.

The single root node of the graph represents the object class description of the object type document layout root in the set of constituents representing the generic layout structure.

The single root node of the primary graph represents the object class description of the object type document logical root in the set of constituents representing the generic logical structure.

The single root node of each secondary graph represents an object class description which is identified by the attribute “logical source” of one or more layout object class descriptions.

Some paths may pass through a particular node a multiple number of times. Some of the directed arcs may start and end on the same node.

The non-terminal nodes, that is the nodes from which some directed arc starts, which in all cases include the document root itself, represent object classes for composite objects.

In the case of the terminal nodes, that is the nodes from which no directed arc starts, the situation depends on which generic structure is represented by the complete generator set of object class descriptions.

If the generic logical structure is represented, then the terminal nodes represent the object classes for basic logical objects.

If the generic layout structure is represented, then the terminal nodes represent object classes for basic pages, for blocks, or for frames which will always be the lowest level frame in the particular branch of the hierarchy (see 7.5.7).

Generic content portions are referenced only from object classes for basic objects and each generic content portion shall be referenced by just one object class.

7.5.4 Generation of structures

A complete generator set of object class descriptions and associated generic content portions representing a generic structure is used to maintain the internal consistency of a document. This is achieved by defining which specific structures are possible in a given document class and which object classes are possible in parts of a document (see 9.3.2.1). When creating, editing or laying out a document, a generic structure can be used to control the generation and modification of the specific structure and thereby preserve the intentions of the originator, in this case the creator of the document class description.

Generation and modification of the specific structure is controlled by ensuring that the attributes of each object class description are used to specify default values for attributes in the corresponding object descriptions.

Recursion is permitted in the generic logical structure and in the generic layout structure, thus objects of the same object class may be hierarchically related.

The ITU-T Rec. T.410-Series | ISO/IEC 8613 does not define any particular generic structures or object classes. However, it allows the definition of a potentially infinite variety of document classes.

7.5.5 Content in generic structures

An object class description for a basic object can specify content in one of three ways:

- by specifying one or more generic content portions;
- by specifying a content generator;
- by referencing an object class in a resource-document which has content specified.

A content generator allows the content information to be specified by an expression which is evaluated during the layout process. Together with specification of a content generator, a content portion description may also be present, in order to allow for specification of attributes of the content portion.

When an object class description specifies content, content portion descriptions can be derived for the specific structure.

A basic object class description with a generic content portion description may be referred to by more than one basic object description of the same specific structure and of the same object class, which permits content sharing.

7.5.6 Generic logical structure

The object classes that can occur within the generic logical structure of a document are for objects of the following object types:

- document logical root;
- composite logical object;
- basic logical object.

The possible hierarchical relationships between logical object classes are as follows:

- the object class for the document logical root can specify that there shall be any number and combination of immediately subordinate composite logical objects and basic logical objects;
- an object class for a composite logical object can specify that there shall be any number and combination of immediately subordinate composite logical objects and basic logical objects;
- an object class for a basic logical object can optionally have content specified; if content is specified then it may be in the form of one or more generic content portions, specification of a content generator, or a reference to an object class in a resource-document which has content specified.

Object classes of these object types serve as templates for objects of corresponding types in the specific logical structure.

7.5.7 Generic layout structure

The object classes that can occur within the generic layout structure of a document are for objects of the following object types:

- document layout root;
- page set;
- composite or basic page;
- frame;
- block.

The object classes for pages within a generic layout structure may either be basic pages or composite pages, but not a combination of these.

The possible hierarchical relationships between layout object classes are as follows:

- the object class for the document layout root can specify that there shall be any number and combination of immediately subordinate pages or page sets;
- an object class for a page set can specify that there shall be any number and combination of immediately subordinate pages or page sets;
- an object class for a composite page can specify that there shall be either any number of immediately subordinate frames or any number of immediately subordinate blocks;
- an object class for a frame can specify that there shall be either any number of immediately subordinate frames or any number of immediately subordinate blocks;
- an object class for a basic page or a block always has content specified; this may be by specifying one or more generic content portions, by specifying a content generator, or by referencing an object class in a resource document which has content specified.

Object classes of these object types serve as templates for the corresponding objects in the specific layout structure. Basic objects created by the layout process, for the content associated with the logical structures, do not reference layout object classes.

Object classes for basic pages and object classes for composite pages that specify immediately subordinate blocks shall only be included within documents of the formatted document architecture class.

7.5.8 Resource-document

A logical object class description may contain a reference to a logical object class description in the resource-document (see 6.3.12). A layout object class description may contain a reference to a layout object class description in the resource-document. In both cases, a subset or all of the attributes may be present in the object class description of the interchanged document only implicitly, by reference to the object class description of the resource-document. This may include the attribute “content portions”, in which case content portions may be present in the interchanged document only implicitly, by reference to the content portion descriptions of the resource-document.

The association between the interchanged document and the resource-document is achieved through the use of resource names. The resource-document includes, within the document profile, a table which maps resource names onto the object class identifiers of that resource-document. An object class description within the interchanged document may then refer to an object class description within the resource-document by using one of these resource names.

Attributes in the resource-document that are used by reference as attributes of object class descriptions in the interchanged document shall not make use of object class identifiers.

7.6 Colour

7.6.1 Overview

One goal of the ITU-T Rec. T.410-Series I ISO/IEC 8613 is to support efficient device independent interchange of documents containing coloured layout components or coloured contents. It enables the originator to transmit a precise specification of intent with respect to colour.

In order to achieve this goal, a colorimetrically precise reference colour space is used to specify interchange colour spaces.

7.6.2 Reference colour space

The reference colour space is based on the XYZ tristimulus values of the 1931 CIE Standard Colorimetric Observer, described in CIE Publication S002. All interchange spaces are defined in terms of this space.

NOTES

1 The reference colour space is colorimetrically precise and covers all perceivable colours. It is based on properties of the human visual system, determined by extensive experiments in colour matching, rather than on the properties of any particular device.

2 It is recognised that the general problem of appearance matching has not been completely solved. However, this Specification uses the best available and internationally recognised approach, which is the CIE system of colorimetry.

3 The reference colour space is normalised such that the Y tristimulus value is 1 for the reference white to allow for simplicity of conversion from colorimetric values to other colour spaces. This differs from the CIE recommendation of normalising Y of the perfect white reflecting diffuser (reference white) to exactly 100. See H.1 and H.2.

7.6.3 Interchange colour spaces

The interchange colour spaces for document level attributes as well as for character content architecture, raster graphics content architecture and geometric graphics content architecture are RGB and CMY(K). CMY(K) references both the CMY colour space and the CMYK colour space. The CMY colour space does not include a black component. The CMYK colour space has a place holder for a black value which may be zero. RGB allows the use of the reference colour space as well as other colour spaces based on self-luminous primary colours. In addition to these colour spaces, CIELUV and CIELAB are permitted for raster graphics content architecture.

Multiple interchange colour spaces are allowed to accommodate many different sources of colour data. All allowed interchange colour spaces have the common feature that they are defined in terms of the reference colour space. Transformations from each interchange colour space to the reference space are documented in Annex H. These transformations are defined in the data stream by additional interchange data called calibration data.

NOTE – The capability for more than one interchange space is required because:

- content architectures have different requirements;
- different applications have different requirements for representations of colour;
- conformance is required with other standards that are related to a content architecture.

7.6.4 Colour imaging model

The following is a conceptual model of one way in which interchanged colour content may be imaged.

An imaging system converts a colour value specified by one of the interchange colour spaces to the reference space. It then converts it from the reference space to the device space for its imaging device. This applies for both input and output devices. An imaging system may also utilize the colour values directly without conversion. See Figure 12 for a description of these colour space relationships at the document architecture level. This Figure also represents the common colour spaces used across all content architectures.

NOTES

1 Although this is conceptually two operations, it can be implemented as one transformation.

2 One practical difficulty which can arise during this process is the need to produce colours which are outside the gamut of the imaging device. In addition, an imaging device may have a limited ability to reproduce colour. The methods for addressing these problems are left for future study. See also Annex J.

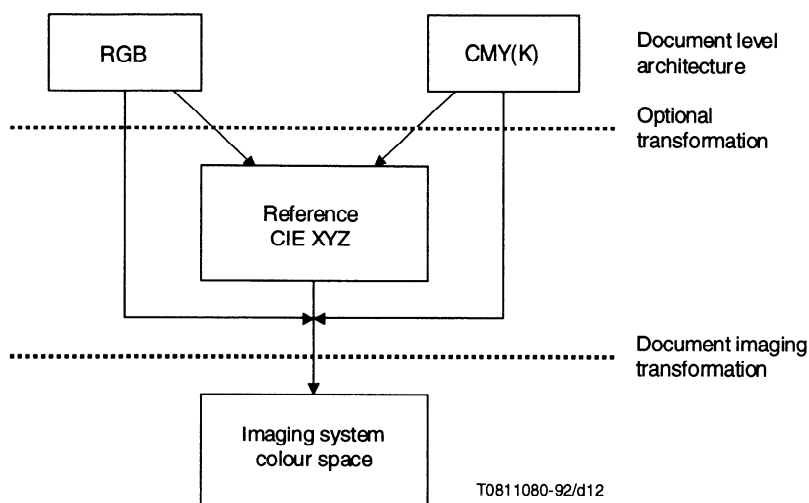


Figure 12 – Relationships among the colour spaces for the document level architecture

7.6.5 Colour tolerance

In some instances, the originator of a colour description may specify some tolerance or permit some variability in the interpretation of colour values. An example is the traditional usage within CGM of RGB values which do not refer to any particular set of RGB primaries or to any particular reference white. (See H.3.1 and H.3.2). Therefore, this Recommendation | International standard uses the attribute “colour tolerance” to allow the originator to specify colour differences, the amount by which colour values can vary during presentation (or processing) and still satisfy the original intent. The colour differences are specified using one of the two CIE-recommended uniform colour spaces – CIELUV and CIELAB (see Annex H and CIE Publication, 15.2, 2nd Edition, 1986).

8 Interface between the document architecture and content architectures

8.1 General concept

A document is defined by means of:

- The *document architecture* – A finite set of constituents together with their characteristics and rules that establish their relationships.
- The *content architectures* – A finite set of content elements, (for example, graphic characters, pels), together with their characteristics and rules that establish their relationships.

Content elements and control functions make up content portions governed by a content architecture. Every content portion is associated with a basic component governed by the document architecture.

One content element is thus specified within the document through a specific content architecture together with the document architecture.

The terms basic values and non-basic values are used in this clause with the following meanings:

- *basic values* of attributes, control function parameters and other capabilities are those that are unconditionally allowed in document interchange in the context of a particular document application profile;
- *non-basic values* of attributes, control function parameters and other capabilities are only allowed in document interchange in the context of a particular document application profile, if their use is declared in the document profile.

The document architecture and any content architectures are connected through an interface, as shown in Figure 13.

The interface is specified by describing which attributes defined in a content architecture have influence on constituents of the document architecture (document profile, basic component descriptions, content portions).

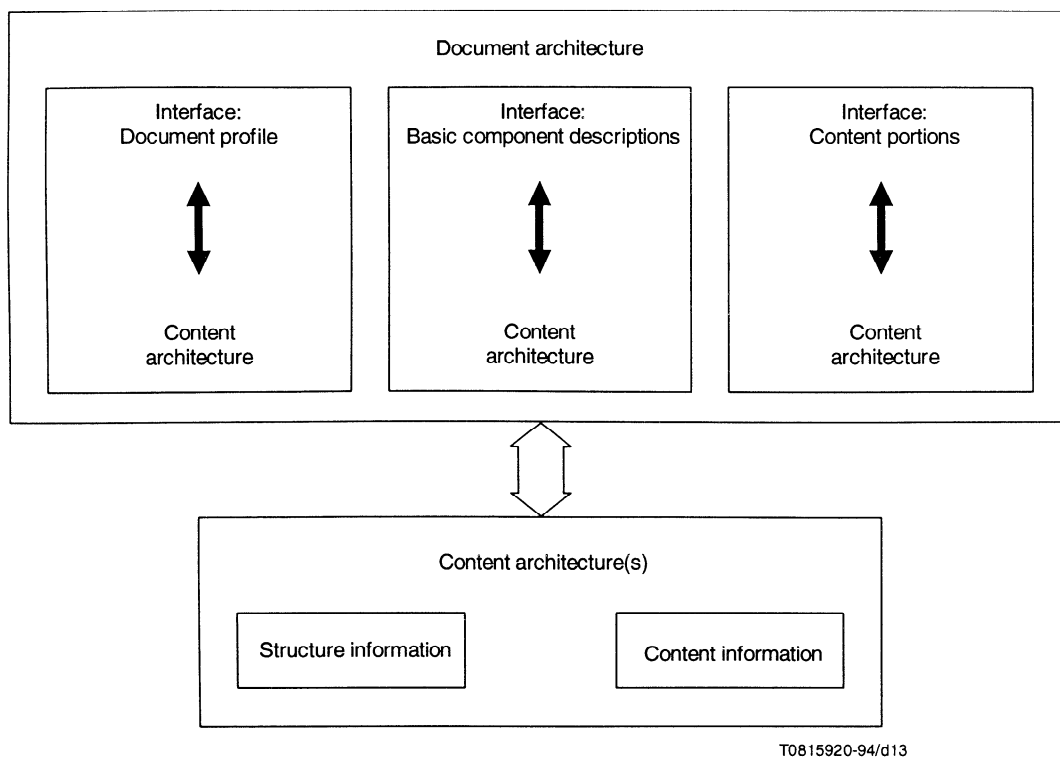


Figure 13 – Document architecture – Content architecture interface

8.2 Specification of a content architecture

The specification of a content architecture consists of three categories of information:

- structure information that identifies the class of content architecture, its internal structure rules and the positioning and imaging rules;
- content information that determines what information comprises the content portion. It includes a set of content element repertoires along with a default repertoire and the set of control functions available in the content architecture;
- information that specifies the basic values, default values and non-basic values of presentation attributes, coding attributes and control function parameters.

This information is required in order to link the document architecture to content architecture(s), through the interface information.

8.3 Interface information

A content architecture has the following three interfaces to the document architecture:

- the *document profile*, which includes attributes that identify the content architecture class(es) used within the document and attributes that specify the use of any non-basic content architecture features;
- an *object description* or an *object class description* for a basic object, which include attributes that identify the content architecture class and presentation attributes;
- a *content portion description*, which includes attributes that identify the type of coding and coding attributes.

The definition of each content architecture specifies the information associated with these attributes as described in the following sub-clauses.

8.3.1 Interface between a content architecture and the document profile

The definition of a content architecture shall include the following information:

- the value(s) of the attribute used to identify the content architecture class in the document profile;
- the format of the attribute used to specify non-basic values of presentation attributes in the document profile;
- the format of the attribute used to specify default values of presentation attributes in the document profile;
- the format of the attribute used to specify non-basic values of coding attributes in the document profile;
- the format of the attribute used to specify default values of coding attributes in the document profile.

The distinction between basic and non-basic values of presentation attributes and coding attributes is not specified as a part of the definition of a content architecture. This distinction is made in the definition of a document application profile in accordance with ITU-T Rec. T.411 | ISO/IEC 8613-1.

8.3.2 Interface between a content architecture and a basic component description

A definition of a content architecture shall include the following information:

- the value(s) of the attribute used to identify the content architecture class in a basic component description;
- the formats, permissible values and recommended default values of the presentation attributes;
- the presentation attributes for which a document application profile may define non-standard default values;
- possible interactions between presentation attributes and document architecture attributes.

8.3.3 Interface between a content architecture and a content portion

The definition of a content architecture shall include the following information:

- the value(s) of the attribute used to identify the type of coding in a content portion;
- the formats, permissible values and recommended default values of the coding attributes;
- the coding attributes for which a document application profile may define non-standard default values;
- possible interactions between control functions and presentation attributes.

9 Attribute definitions

9.1 General principles of attributes

Characteristics of structural elements of a document and relationships between structural elements are represented by constituents which are sets of attributes. Each attribute is identified by a name and has a value that describes the characteristic or relationship. Attributes are also used to identify constituents.

This clause contains definitions of all attributes defined in the document architecture. The attribute definitions specify the range of values that each attribute can assume.

9.1.1 Attribute categorisation

Attributes are categorised according to the constituent to which they apply, as follows:

- document profile attributes;
- component description attributes;
- layout style attributes;
- presentation style attributes;
- content portion description attributes;
- protected part attributes.

The constituents termed descriptions are those which are counterparts of the structural elements (objects, object classes and content portions).

Document profile attributes are defined in ITU-T Rec. T.414 | ISO/IEC 8613-4. The general principles of other attribute categories are described in this sub-clause and the attributes are defined in 9.3 to 9.10.

NOTE – The tables in Annex E summarise all attributes defined in this clause, in accordance with their categorisation.

9.1.1.1 Attributes of components

Attributes of component descriptions are further categorized as follows:

- *Shared attributes* – These can be included in both logical and layout component descriptions (see 9.3).
- *Layout attributes* – These can be included in layout component descriptions only (see 9.4).
- *Logical attributes* – These can be included in logical component descriptions only (see 9.5).

Some attributes of components can be included in composite component descriptions only, some in basic component descriptions only and some only in component descriptions of a particular object type.

In addition, some attributes may be included in object class descriptions only, some in object descriptions only and some in both object class descriptions and in object descriptions.

All attributes, except the presentation attributes (see 9.1.1.4), are independent of the content architecture pertaining to any component descriptions.

9.1.1.2 Layout style attributes

A layout style consists of the attributes:

- layout style identifier;
- user-visible name;
- user-readable comments;
- application comments;
- sealed;
- a set of layout directive attributes.

Layout styles are referred to from logical component descriptions. Such a reference is made using the attribute “layout style” with a value equal to the value of the layout style identifier. The layout style identifier uniquely identifies the layout style within the document.

A layout style that explicitly specifies all of the appropriate attributes that relate to that layout style is termed a root layout style. Any number of additional layout styles may be derived from a root layout style. The derived layout styles will specify only the attributes and/or attribute values that differ from the root layout style. This provides a means for factoring attributes thus preventing the necessity of copying the same attributes in similar layout styles. Any number of levels of derived layout styles may be provided by first specifying a layout style derived from the root layout style and then specifying other layout styles derived from derived layout styles.

A layout style may be referenced by more than one logical component description. The effect of a reference to a layout style is to apply its layout directive attributes to that logical component description which contains the reference. The layout directive attributes are a set of attributes which specify information for the document layout process (see clause 10). Precedence rules are specified in 9.1.2.4, 9.1.2.6 and 9.7.16.

Layout directive attributes are defined in 9.7.

Thus, layout styles affect the layout of objects, not their content and are independent of particular content architectures. Presentation styles affect the layout and imaging of the content associated with basic objects and hence are content architecture specific. Therefore, there is no conflict between the two.

9.1.1.3 Presentation style attributes

A presentation style consists of the attributes:

- presentation style identifier;
- user visible name;
- user-readable comments;
- application comments;

- transparency;
- border;
- sealed;
- colour;
- colour of layout object;
- object colour table;
- content background colour;
- content foreground colour;
- content colour table;
- sets of presentation attributes particular to each content architecture.

A presentation style may be referred to from a basic logical or basic layout component description. Such a reference is made using the attribute “presentation style” with a value equal to the value of the presentation style identifier. The presentation style identifier uniquely identifies the presentation style within the document.

A presentation style may be referenced by more than one component description.

A presentation style that explicitly specifies all of the appropriate attributes that relate to that presentation style is termed a root presentation style. Any number of additional presentation styles may be derived from a root presentation style. The derived presentation styles will specify only the attributes and/or attribute values that differ from the root presentation style. This provides a means for factoring attributes thus preventing the necessity of copying the same attributes in similar presentation styles. Any number of levels of derived presentation styles may be provided by first specifying a presentation style derived from the root presentation style and then specifying other presentation styles derived from derived presentation styles.

The effect of a reference to a presentation style is to apply its presentation attributes, transparency, colour and border to the basic component which contains the reference. In addition, presentation attributes can be specified for basic layout components. Precedence rules are specified in 9.1.2.4 and 9.1.2.6.

Each set of presentation attributes corresponds to a particular content architecture class.

Presentation styles affect the layout and imaging of the content associated with basic objects and hence are content architecture specific. Layout styles affect the layout of objects, not their content. Therefore, there is no conflict between the two.

9.1.1.4 Presentation attributes

Presentation attributes, which can be applied to basic components only, are used to specify the properties of the content portion(s) associated with that component. The presentation attributes specify information for the content layout process and content imaging process.

The presentation attributes specify the initial conditions concerned with the presentation of the content portion(s). The presentation attributes are specified either in a presentation style or, for layout components, in a basic layout component description.

The set of presentation attributes that is applicable depends upon the particular content architecture class specified for the basic component. The content architecture class attributes specify the content architecture class of the associated content portion(s), and therefore which set of presentation attributes are applicable to the content (see 9.3.4).

The presentation attributes are defined in those Specifications of the ITU-T Rec. T.410-Series | ISO/IEC 8613 that are concerned with individual content architectures.

9.1.1.5 Content portion description attributes

A content portion description consists of the attributes:

- content identifier-logical;
- content identifier-layout;
- type of coding;
- alternative representation;
- coding attributes;
- content information.

These attributes uniquely identify the content portion description, specify the type of coding used to code the content information and specify an alternative representation that may be imaged in lieu of content information when a recipient is not capable of decoding and/or imaging the content portion.

Content portion identifiers and alternative representation are fully specified in 9.9. Other attributes of content portions depend on the content architecture and details are specified in those Specifications of the ITU-T Rec. T.410-Series I ISO/IEC 8613 that deal with individual content architectures.

9.1.1.6 Protected part attributes

There are four kinds of descriptions of protected parts of a document, one for sealed information and three for enciphered information. They are:

A sealed document profile description consists of the attribute “sealed document profile information”, which is a document profile, with a similar structure to a regular document profile. The only differences are that:

- every attribute is optional;
- only the attributes that are sealed shall be present.

It is also possible to seal absent attributes.

An enciphered document profile description consists of two attributes:

- *Protected part identifier.*
- *Enciphered information* – The value of this attribute is the result of an encipherment of the confidentially protected part of the document profile. The confidentially protected part of the document profile has a structure similar to a regular document profile. The differences are that:
 - every attribute is optional;
 - only the attributes that are confidential shall be present.

A pre-enciphered document body part description consists of two attributes:

- *Protected part identifier.*
- *Enciphered information* – The value of this attribute is the result of encipherment of the confidential part of the document body applied before the layout process has been performed.

A post-enciphered document body part description consists of two attributes:

- *Protected part identifier.*
- *Enciphered information* – The value of this attribute is the result of encipherment of the confidential part of the document body applied after the layout process has been performed.

9.1.2 Attribute values

9.1.2.1 General principles

Each constituent consists of a set of attributes; the attributes in the set are said to be *specified* for the particular constituent.

In the case of object descriptions, the complete set of attributes describing the characteristics of the object need not be specified for the constituent; such attributes can also be derived from attributes specified for other constituents, by using defaulting rules specified in 9.1.2.4. In such cases, the attributes which are specified for the constituent, together with those derived for the constituent, are said to form the set of attributes which *apply* to the constituent (or that are applied to, or are applicable to, the constituent).

Individual attributes may be structured into parameters, sub-parameters, sub-sub-parameters, etc., each attribute definition includes a specification of any such structuring.

Attributes may be formed from a set consisting of a list of attributes. At each point in this specification when an attribute is formed from a set of attributes, a list of the attributes that may be included in the set is explicitly specified.

The definition of an attribute also indicates whether the value(s) of that attribute may consist of one or more elements of data or expression(s).