TECHNICAL REPORT

ISO/IEC

TR

24754-2

First edition
2011-10-15

Information technology — Document description and processing languages — Minimum requirements for specifying document rendering systems —

Part 2:
Formatting specifications for document rendering systems

Technologies de l'information — Description des documents et langages de traitement — Exigences minimales pour spécifier les systèmes de conversion de documents —

Partie 2: Spécifications de formatage pour les systèmes de conversion de documents

Reference number
ISO/IEC TR 24754-2:2011(E)

© ISO/IEC 2011

Contents

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. 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.

In exceptional circumstances, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide to publish a Technical Report. A Technical Report is entirely informative in nature and shall be subject to review every five years in the same manner as an International Standard.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 24754-2 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 34, *Document description and processing languages*.

ISO/IEC 24754 consists of the following parts, under the general title *Information technology — Document description and processing languages — Minimum requirements for specifying document rendering systems*:

⎯ *Part 1: Feature specifications for document rendering systems*

⎯ *Part 2: Formatting specifications for document rendering systems* [Technical Report]

Introduction

The scope of ISO/IEC 24754-1 makes reference to a set of style specifications an originator provides to a recipient to reconstruct the presentation of a structured document. To avoid problems with presentation, the originator and recipient must negotiate the available functionality in a document rendering system. Through such negotiation, the exchanged style specifications can avoid making reference to functionality not available in the document rendering system. This manages the expectations of the originator and the recipient regarding the final rendering.

ISO/IEC 24754-1 presumes the originator (the end user with the data that is to be formatted) has created the required style specifications for the recipient (the compositor writing the stylesheet or other instructions for the document rendering system) to render the document as desired. No guidelines are included in ISO/IEC 24754-1 regarding the components of a formatting specification reflecting all of the style specifications suitable for specifying the use of facilities available in the document rendering system.

This part of ISO/IEC 24754 provides such guidance. Utilizing the terminology of ISO/IEC 24754-1, it describes the context of use of the features listed in ISO/IEC 24754-1 so as to direct the end user in the authoring of a useful formatting specification for developers and compositors to implement. Beyond those terms specified in ISO/IEC 24754-1, this part of ISO/IEC 24754 incorporates guidelines and terminology derived from ISO/IEC 9541 (various parts), ISO/IEC 10179, and ISO/IEC TR 19758.

The relationship between this part of ISO/IEC 24754 and ISO/IEC 24754-1 is depicted in Figure 1, where the end user must write a suitable formatting specification for the compositor to create the necessary stylesheet (layout instructions) for a document rendering system to produce the appropriate formatted result for the end user.
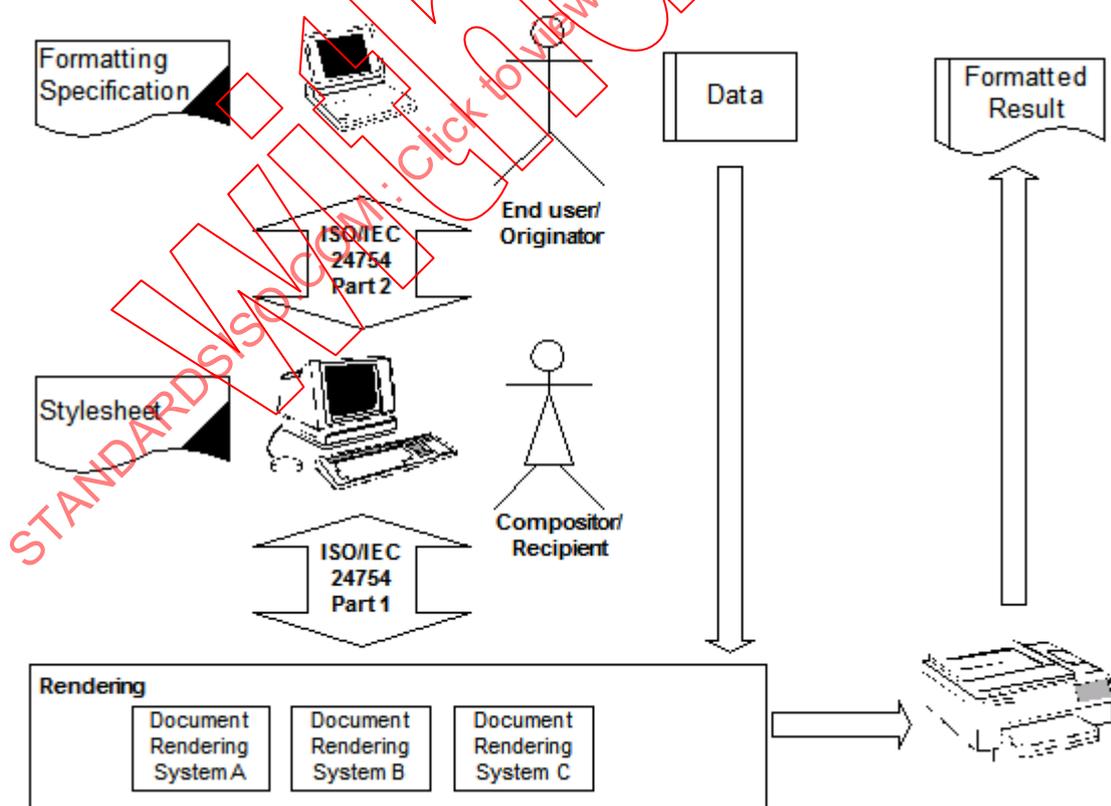

Figure 1 — Relationship between ISO/IEC 24754-1 and ISO/IEC TR 24754-2

Information technology — Document description and processing languages — Minimum requirements for specifying document rendering systems —

Part 2:
Formatting specifications for document rendering systems

1 Scope

This part of ISO/IEC 24754 provides guidelines in the writing of formatting specifications targeted for document rendering systems supporting functionality described by ISO/IEC 24754-1. There is no scope of conformance constraints against which the use of these guidelines are measured. Possible user requirements for publishing layouts are unbounded. It is the user's obligation to describe their formatting requirements to be fulfilled by a developer or a compositor using a document rendering system. Using incomplete formatting specifications can delay achieving the desired rendered result.

This part of ISO/IEC 24754 is a specification of questions for the writer of a formatting specification. It is assumed that the writer of a formatting specification is unaware of the needs of the stylesheet writer in meeting the needs of the specification writer. For a writer aware of such needs, this part of ISO/IEC 24754 is a useful checklist of the issues to be considered.

This part of ISO/IEC 24754 outlines various considerations users should assess when analysing their rendering expectations. Properties described by this part of ISO/IEC 24754 should be specified in sufficient detail such that a developer or compositor has the information required to engage the necessary facilities in the document rendering system.

This part of ISO/IEC 24754 does not presume any particular stylesheet language with which to specify layout and formatting properties. Various stylesheet languages support different degrees of functionality described herein. Not all languages can assume to offer control over the properties described by this part of ISO/IEC 24754, yet it behoves the user to specify as much detail as possible so as to equip the developer or compositor for their task. By knowing as many nuances of layout as desired, there are more strategies available with which to accomplish a layout as close as possible to the requirements.

Many components of this part of ISO/IEC 24754 are presented to the reader as questions in the first person, so as to ask the reader to reflect on their requirements when answering the question.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

NOTE   Each of the following documents has a unique identifier that is used to cite the document in the text. The unique identifier consists of the part of the reference up to the first comma.

ISO/IEC 24754-1, *Information technology — Document description and processing languages — Minimum requirements for specifying document rendering systems — Part 1: Feature specifications for document rendering systems*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 24754-1 and the following apply.

3.1
stylesheet
external style and layout specification expressing properties of document rendering

NOTE   Any expression language can be used and no particular language is presumed by this part of ISO/IEC 24754.

4 General properties of the formatting specification

4.1 Supported output device

A formatting specification should describe the common layout aspects and distinguish the different layout aspects between the various target media for transformation. These distinctions may be necessary based on the target audience of the content for each medium. These distinctions may be necessary based on the rendering features for each medium.

Where possible, reducing the distinctions and establishing common layout aspects between media for the same audience may promote both consistency in presentation and commonality in development of the stylesheet.

4.2 Dimension specifications

All dimensions, extents, sizes and spacing values specified in the document should use units of measure appropriate to their application. For example page sizes, margins and large areas are often specified in imperial or SI units. Font sizes, line sizes and smaller areas are often specified in typographic units such as points (rounded to 1/72nd of an inch) and picas (12 points).

One should avoid using device pixels in any length specification as this varies based on the choice of document rendering system properties and final form resolution.

4.3 Mockups

A mockup layout illustrating the expected result using representative data is a useful guideline in the interpretation of formal formatting specifications. Nevertheless, one should not rely solely on a mockup to convey all the nuances that would be better described in formal prose.

Depending on the technology used to create the mockup, the compositor may be able to easily inspect layout nuances in the properties of the formatted sample. For example, hand-writing XSL-FO to create the mockup, or using a desktop publishing application to compose a sample.

4.4 Supported document formats

A formatting specification should describe in detail the document format and vocabulary being used to express the source content being processed by the stylesheet.

In particular, when using XML as the source document format, all details of the XML vocabulary should be well explained. When using a generic XML vocabulary it may be necessary to document conventions of use (and, more importantly, to diligently follow such conventions when composing the source documents) such that the rendering intent can be properly inferred.

4.5   Supported style languages

While in theory a formatting specification should purely be reflecting the intent of the user for rendering, there may be pragmatic or practical limitations imposed by the style language used with the document rendering system. Recognizing this, it is still in the best interests of all parties for users to express their full intent for formatting, documenting all of their expectations in the result. Developers and compositors can then interpret these requirements to choose either the appropriate style language to use, or at best the most appropriate features of the available style languages from which to choose.

4.6   Rendering media dimension

A formatting specification for paged layout should describe the expected media dimensions such as A4 (210mm by 297mm), US Letter (8.5in by 11in), another internationally accepted dimension by its name, or a custom dimension that may be necessary. The image area is typically bound within the media dimensions using margins but for the requirement of supporting bleeds (i.e. content beyond the physical page cutting dimensions; e.g. coloured thumb indexes with or without cut indentations) may be specified to be larger than the media dimension. The paper placement (vertical/portrait or horizontal/landscape) should be specified so as not to be unclear.

ISO 3535 recommends a margin of at least 20mm from the start-side of the flow of text on a line (e.g. left-side for left-to-right line flow scripts) and at least 10mm from the before-side of the flow of text on a page (e.g. top-side for top-to-bottom block flow scripts).

The image area is typically measured as the result of taking the margin area extents away from the media dimension, however for flexibility in disseminating a consistently sized result to audiences with different expectations for media dimensions, ISO 6422 recommends a common image area for content suitable for both A4 and US Letter is 183mm (short side) by 262mm (long side).

4.7   Colour support

Different colour-management systems are supported on various document rendering systems. The specification of colour can typically be translated into equivalent specifications of different colour schemes. Examples of colour schemes are CMYK, RGB, HSV, HSL, CYCA, CYCB, Pantone, etc. The formatting specification should document the use of a particular colour scheme for use within the specification, as well as any constraints on the necessity to use or not to use any particular colour scheme (perhaps due to licensing constraints).

Where possible, the formatting specification should indicate a fallback colour to use for each colour specified. A common fallback mechanism is the rudimentary red/green/blue RGB web colour system using three values between 0 and 255. A document rendering system not supporting the desired colour scheme can then fall back and use the alternative where required.

4.8   Dynamic content generation

It is essential to distinguish in sample mockups which content is generated or supplied by the stylesheets and which content is sourced from the input XML documents being processed by the stylesheets.

One technique for distinguishing generated and source content in mockup examples is to use colour or some other decoration (e.g. italic, underscore, etc) provided one establishes the conventions so that the differences in appearance are not interpreted as instructions for formatting. The convention is applied to sourced content, leaving generated content as it is intended to appear in the final result. This may not work well if the sourced content itself has formatting properties distinguished from the generated content, as the distinguishing presentation conventions obscure the intended presentation of the sourced content.

Another technique for distinguishing generated and source content in a mockup layout is to replace sourced content with placebo character strings rather than representative character strings. This technique has the added benefit of possibly conveying meaning in the placebo character strings.

3

NOTE   In this clause the diagrams and associated content utilize such a placebo definition by representing all content of the source document by abbreviated references. An example of such a reference as used in this clause is "!17!". The use of exclamation marks prevents misinterpreting abutted citations to source information. In these examples numbers are used between the exclamation marks to indicate source content. A similar convention of using letters in exclamation marks is used to cite formatting instructions to the compositor.

This part of ISO/IEC 24754 does not prescribe any particular abbreviation algorithm or syntax, nor even the use of abbreviations.

4.8.1   Generated content

Generated content is calculated by and supplied by the stylesheet either as fixed boilerplate or conditional generated content based on the presence or absence of source content.

Fixed boilerplate might be the document heading or fixed strings used to indicate headings triggered by the presence of a particular section element in the source. At times boilerplate is quite structured itself, with different components of the boilerplate emphasized or displayed quite differently (e.g. headings with two components, the main component bold and underline and a contiguous but less emphasized component trailing behind the heading).

Generated content might be a lengthy checklist of items, some of which might be checked by the presence of source content, but all needing to be listed for completeness for the reader. This kind of generated content is often found in supplemental source files where enumerations of available values are stored. It is important to have complete examples of the enumerations and to detail where the content is to be located and how it is to be processed.

4.8.2   Sourced content

Sourced content is information copied directly from elements or attributes of the source input files and added to the result either a verbatim or manipulated fashion.

The locations of sourced content should be accurately conveyed in the formatting specification. When working with XML, this can be done with XPath addresses.

4.8.3   Example documentation

Consider the following example mailing label that is part of a formatted result. It combines both generated and sourced information, with necessary instructions regarding the generated content. The callouts are labeled with letters to cite specific formatting instructions to the compositor. The sourced content is cited using numbers and is presented using the expected visual formatting properties.
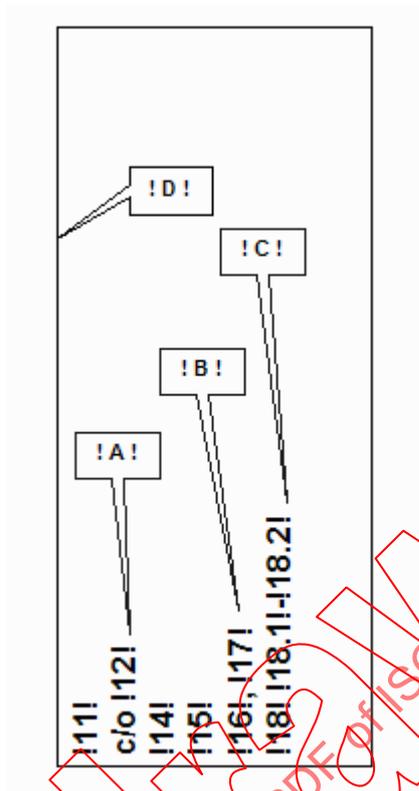
Figure 2 — Combining generated and sourced content

The associated formatting specification instructions in prose might read as follows:

— the mailing label is comprised of the bill-to information as follows:

```
!11! /invoice/bill-to/name

!12! /invoice/bill-to/contact

!14! /invoice/bill-to/addressline1

!15! /invoice/bill-to/addressline2

!16! /invoice/bill-to/city

!17! /invoice/bill-to/stateprov

!18! /invoice/bill-to/country

!18.1! /invoice/bill-to/country/@postalcode

!18.2! /invoice/bill-to/country/@postalsubcode
```

— the generated content instructions are as follows:

A - when the contact is present, prefix the contact with the sequence "c/o"; if absent present nothing on the line but leave the line blank should the reader wish to annotate the label and return it

5

B - always follow the city with the sequence "," and when the province/state is present, display it, otherwise put all of the country line information on the same line as the city line following the comma

C - when the postal sub-code is present, separate it from the postal code with a hyphen

D - to meet up with the envelope window, the mailing label is 5cm by 1.2cm in size, positioned 1cm from the top of the page and 1cm from the right of the page, on the first page only; note that margins for the entire first section of content shall be reduced by .5cm; if the first section of content is shorter than the mailing label, the second section of content is started below the mailing label with the full column width as margins

4.9 Supported coded character sets and encoding schemes

It is important to match the user's coded character set requirements with any limitations in the document rendering system. The user should indicate in the formatting specification the character encoding of the sourced content.

Where possible, a CREPDL (ISO/IEC 19757-7) specification of expected character support in the document rendering system would be a very useful tool to the developer and compositor.

4.10 Composition of combined characters

Where applicable, any special composition requirements for combining syllabic characters should be specified. This will help the developer and compositor engage such combinations if available in the document rendering system.

4.11 Directionality of text

Where the language of the text is not specified in the source content, the formatting specification should indicate the text directionality governing the page geometry.

When using XML for source content the `xml:lang=` is a suitable indication of the predominant language governing the directionality of text for the geometry of the page. The customary application of this attribute is in the document element of the source content. Typical values suggesting right-to-left directionality for page geometry include `xml:lang="he"`, `xml:lang="ar"` or `xml:lang="ur"` for, respectively, Hebrew, Arabic and Urdu. Other values suggest left-to-right directionality for page geometry.

4.12 Rendering of white-space characters

A formatting specification should include any special instructions regarding the handling of white-space characters in the composition results. Certain sourced content might preserve white-space characters found between words while most of the sourced content might require such characters to simply be collapsed into a single space. For example, white-space characters in a program listing presented in a monospace font may need to be preserved precisely as entered, while sequences of white-space characters in a block of lines presented in a proportional font may need to be collapsed to a single space.

4.13 Font substitution

The formatting specification should provide the list of font resources that the developer or compositor is permitted to substitute, or is explicitly not permitted to substitute, in the formatted result.

Because the acceptable differences of the rendering result vary based on the intended purpose of document, the specification should indicate an appropriate choice of a default value (substitutable or non-substitutable) based on the greater impact on font resources, and then the lesser list of the font resources with non-default values can be specified.

For any non-substitutable font resources, the font resource file or the identifiers of the particular font resource are expected also to be provided. This identifier is dependent on the font substitution mechanism supported by the document rendering system.

Where font file formats have more detailed information on the typeface, such as its font version number, this should be included in the formatting specification. This is important information even if the rendering system does not have an interface to reflect such details, as it may help the compositor fulfill the task. See also clause 4.15.

4.14 Font-resource portability

In a formatting requirement where the font is non-substitutable, the formatting specification should indicate that the font resources are to be embedded in the document. In addition, the method of how to obtain the font resource should be clarified (e.g. which URI to use for downloading the resource).

When specifying font embedding, there are several levels to be considered, including: embedding the entire font file (so as to enable revising the document); embedding only those glyphs used by the document (so as to reduce the size of the formatted result); or transforming the glyphs to equivalent scalable vector or rasterized graphic representation (because of the limitations of the font license). The formatting specification should indicate the expected or required embedding methods in consideration of the purpose of the document and of the constraints of the font licenses.

4.15 Embedding of non-standard characters and glyphs

Where substitutable font resources are to be used, the formatting specification should indicate the reference of each non-standard character and glyph in the font. Most fonts indicate the character sets in use and provide the standard glyphs for the indicated character code points. But often fonts will include non-standard characters and glyphs at the character code points that are not used by the indicated character sets.

To avoid the incorrect rendering by an unexpected glyph in the substituted font, the formatting specification should be prepared with an awareness of the coverage of the characters and glyphs used by the substitutable font resources. If non-standard characters are included in these substitutable font resources, the formatting specification should indicate the expected method to interchange the non-standard characters and their corresponding glyph, such as through font embedding, rendering an inline picture, the use of a ISO 10036 identifier, etc.

5 Description of pagination and rendering of page layout

The following list of properties in a formatting specification for a document rendering system is not comprehensive or exclusive. Any differences between pages or information on pages needs to be described, and these are only guidelines of things that may differ, and they attempt to bring to light properties of the layout that the reader may be unaware of.

5.1 Pages and geometry

All possible layout geometries for the rendered pages need to be enumerated and given semantic labels. Determine the number of unique combinations of margins and running headers.
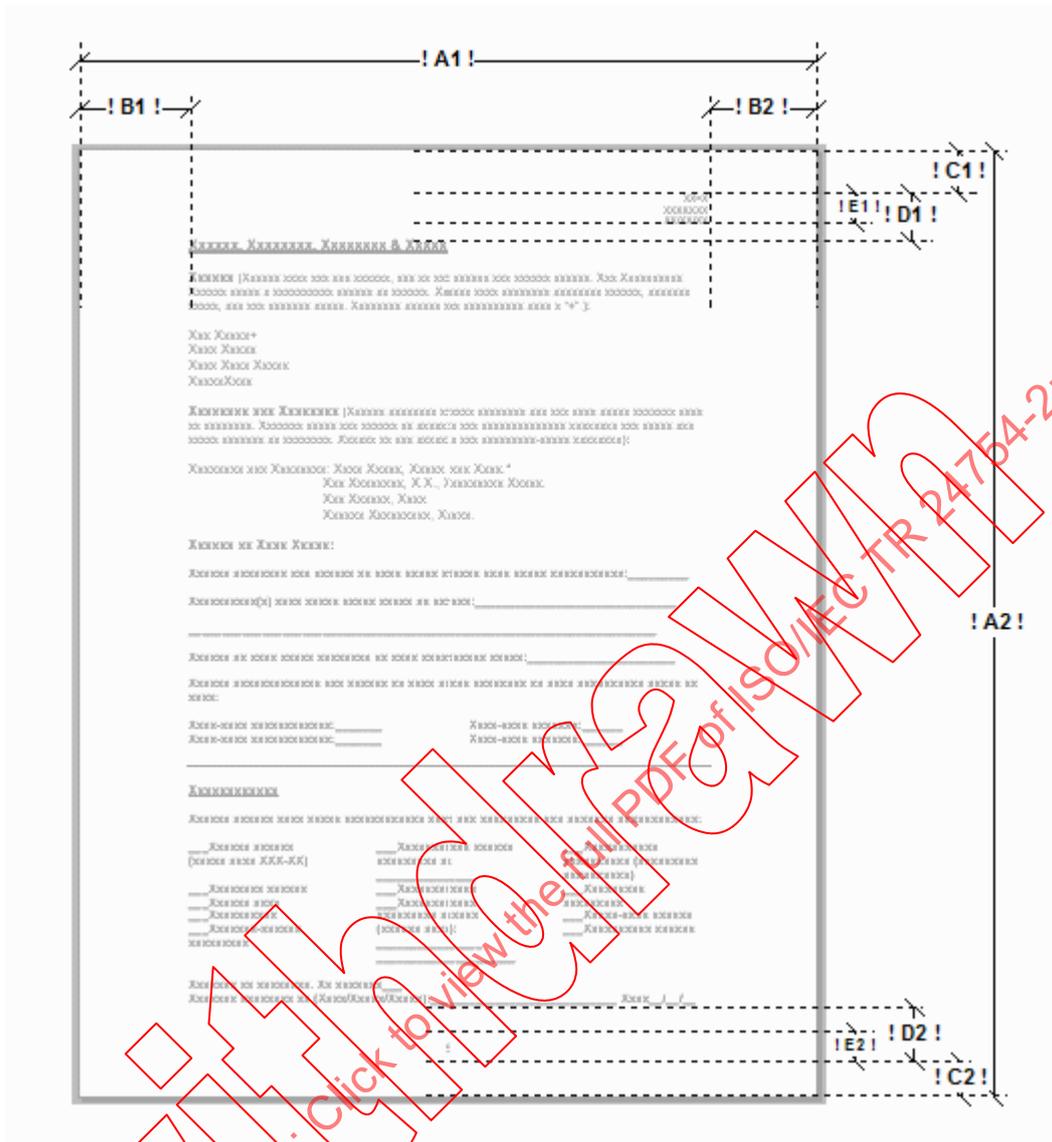
Figure 3 — Page geometry

5.1.1 Page dimensions

⎯ `A1` - page width

⎯ `A2` - page height

⎯ are there multiple page sizes used in the document?

⎯ any double-sized "right hand folds"?

5.1.2 Horizontal page margins

⎯ `B1` - left margin

⎯ `B2` - right margin

⎯ are there differences between odd and even pages, perhaps due to the binding?

⎯ do the reader need a mirror layout specifying margins on the inside edges and the outside edges of a folded open document?

5.1.3 Vertical page margins

— C1 - top page margin

— C2 - bottom page margin

— are there differences between the first page and subsequent pages of the document?

— are there differences between the first page and subsequent pages of a section of the document?

— are there differences on the last page of the document or a section?

— are there differences when there is only a single page in the document or a section?

5.1.4 Vertical body margins

— D1 - top body margin - the distance between the top of the header and the top of the body

— D2 - bottom body margin - the distance between the bottom of the body and the bottom of the footer

5.1.5 Header and footer extents

— E1 - header extent - the fixed height of the header (must be less than or equal to D1)

— E2 - footer extent - the fixed height of the footer (must be less than or equal to D2)

5.2 Column geometry and layout grid

Does the documentation predominantly use flowed columns (where an indeterminate amount of information flows from one column to the next)? At what points are all of the columns spanned with title, graphic or any other information? Are there any requirements to balance the lengths of multiple columns across the page when the last content does not fill the page?

Where information is determinate, is there any preset layout grid within which the information is placed? Can the preset layout be expressed as individual and spanned cells of an invisible border table? Need any edges of the table be visible, and if so with any kind of repetitive pattern (e.g. dots, dashes, etc.)?

6 Layout independent formatting

Does the body contain any out-of-line information and, if so, might its presence adjust margins temporarily?

In the following diagram there is an address label arbitrarily positioned at the top right of the page in order to align with the address window of the envelope being used for the printed document. The dimensions of this label impact on the margins of the content in the top two-thirds of the page. The content after the horizontal rule is then allowed to occupy a wider area of the page, as it falls after the bottom edge of the label.
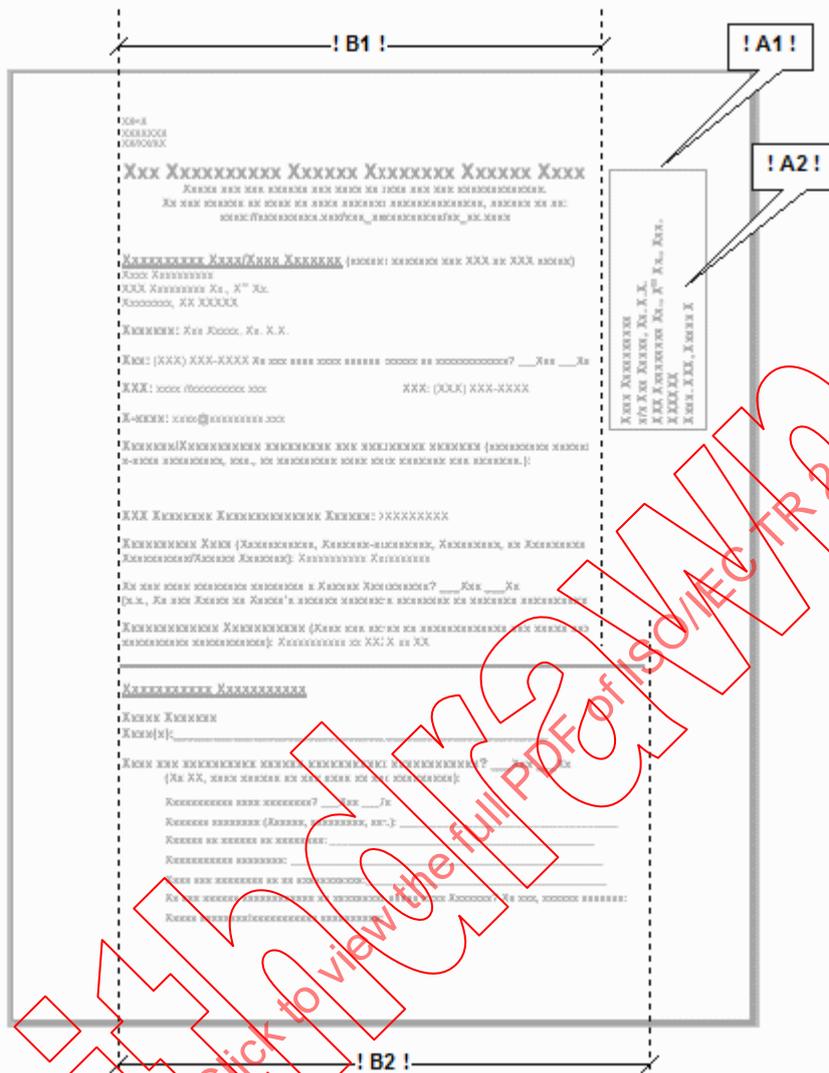
Figure 4 — Body details

6.1 Absolutely-positioned and out-of-line constructs

—— A1 - are there any out-of-line constructs? What are the dimensions of each? What is the placement of each on the page? Is the placement relative to the flow or relative to the page?

—— A2 - what are the formatting rules for the construct and its content? Is the construct bordered? What are the margins of the information kept inside of the construct?

6.2 Impacts on flow margins

—— B1 and B2 - does the presence of an object impact on the margins of information in the body? Where do any changes get restored to the full width of the column?

6.3 Floats

Do any of the information components belong out of the flow and are to be positioned at the top or bottom of each page? Note that floating to the bottom is only supported if there are no footnotes in the information.

When floats are on a page, what style of separator is desired between the set of floats and the body of the content?

6.4 Notes

Notes in source content are typically authored in the place of their reference in the body of the text. How the note is laid out on the page is then flexible, such as an out-of-line footnote with an in-line footnote citation, an endnote flowed at the end of a section with an inline end note citation, or a flowed note between blocks on the page.

6.4.1 Footnotes

Are there any footnotes, acronym expansions or glossary definitions to be taken out of the flow and positioned at the bottom of the page? If so, how are the citations to be identified and sequenced? How are citations offset from the surrounding text for more visibility?

When footnotes are on a page, what style of separator is desired between the set of footnotes and the body of the content?

6.4.2 Side notes

Are there any intrusions in the body content in which a side note is to be flowed? Does the body content flow underneath the side note or is it constrained to the edge of the side note such that the area after the side note is not occupied?

If a single line includes two side notes, do the stack into the body content further reducing the body width, or do they stack in the block progression direction with the minimum impact on the body intrusion?

6.4.3 End notes

Are notes to taken out of the flow and collected at the end of sections of content (e.g. chapter notes or section notes)? If so, how are the citations to be identified and sequenced? How are citations offset from the surrounding text for more visibility?

6.4.4 Flowed notes

Unlike footnotes and end notes that jump out of the flow, flowed notes appear on the canvas within the flow of text. How are the margins for flowed notes distinguished from the margins of surrounding content? How are flowed notes titled? Are there fixed or variable annotations presented along with the title of notes?

Are flowed notes to be rendered with multiple lines within a single line of the block? An example would be a two-line half-height font rendering of the note, presented between single-line full-height parentheses.

6.4.5 Interline notes

In CJK formatting an interline note is composed within a line space at the edge of the inline text, in parallel with the inline text. What size is used to render such notes? How are such notes aligned with the content being annotated?

6.4.6 Ruby annotations

Are ruby annotation compositions an aspect of the formatted result for CJK content? Is the composition mono-ruby or group-ruby? On which side or both sides of the text are the ruby characters rendered?

7   Tables

There are typically three uses for tables: tuples of block-aligned areas, a Cartesian coordinate layout of cells corresponding to row and column indexes, and serpentine lists of items populated in columns to save space on a page.

In these requirements, it is assumed that the column widths used in the table are fixed for the entire table. If one needs varying widths of columns for the rows of the table, this will need to be described in detail in order to establish a strategy for layout that is atypical of standardized tables.

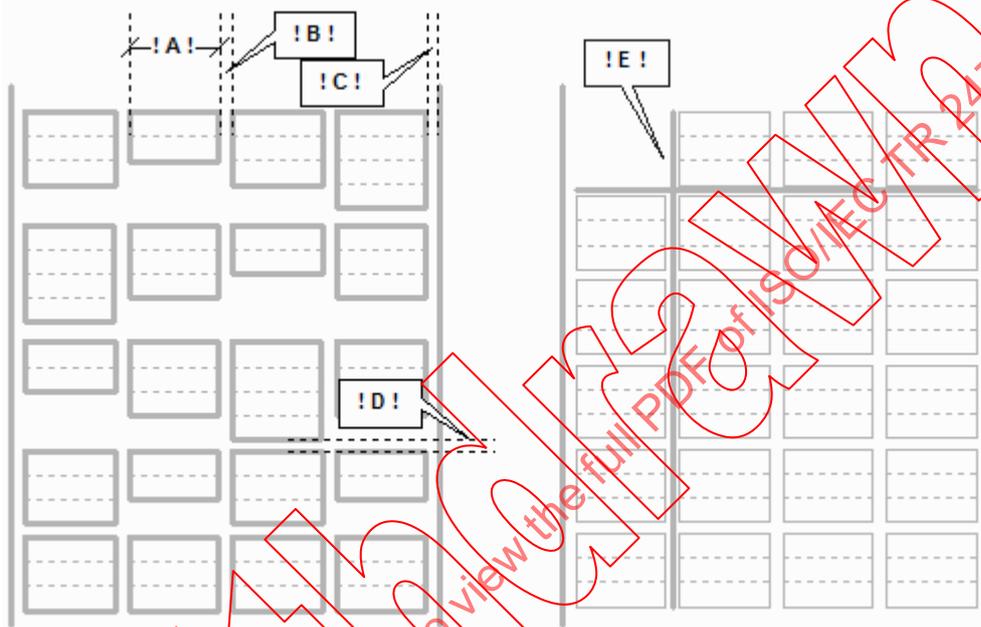### 7.1   Traditional table layouts

Figure 5 — Traditional table layouts

⎯ A - for each of the columns in the table, what is the width of that column (for the entire table)?

⎯ B - between each of the columns of the table, are the edges coincident (overlapping one on top of the other) or is there a fixed separation between all of the columns of the table?

⎯ C - what is the distance between the table and both of the margins?

⎯ D - between each of the rows of the table, are the edges coincident (overlapping on top of the other) or is there a fixed separation between all rows of the table?

⎯ E - are there any decorations around the table or inside of the table? Are any or all of the cells bordered with any particular visible rule pattern?

7.2   Serpentine lists in tables

In the following page layout, there are two instances of what appears to be a three-column table on the page. The content of each table is a lengthy list of some kind, containing many members that would when otherwise formatted occupy too much length and leave too much white-space on the page. By organizing these lengthy lists into columns, real-estate is saved and the reader of the document can navigate the many members much more quickly.
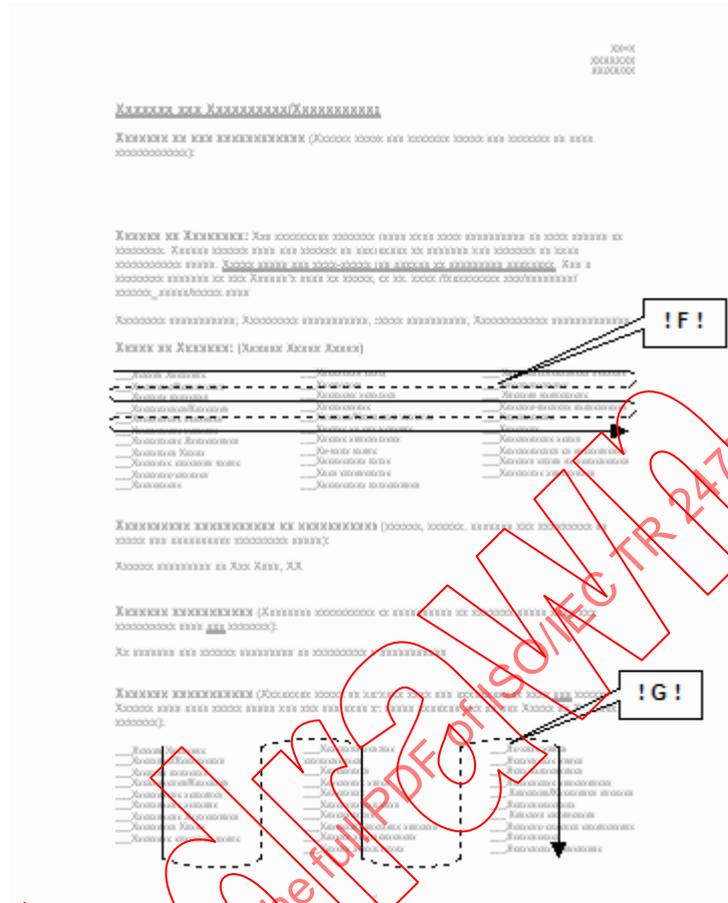
Figure 6 — Serpentine lists

— `F` - is a serpentine list populating the table columns in a row-dominant fashion?

— `G` - is a serpentine list populating the table columns in a column-dominant fashion?

8 Lists

Many choices are available in traditionally formatted lists. A list is a collection of items of block level bodies, each distinguished by a label of some kind. Lists may be bulleted (unordered) or enumerated (ordered).

The edges of the labels and body are typically fixed for all items of the list, but may for special requirements be specified on a per item basis. For the typical case, the values requested below ensure a consistent display of list members across differing column widths.

The rules for each of the item's label and the item's body are the block rules described above. Many separate blocks may be used in either the label or the body block area of a list item.

See the discussion of tables for lengthy serpentine lists that are arranged in temporary columns (table cells) in the middle of the flow to preserve real estate on the page.
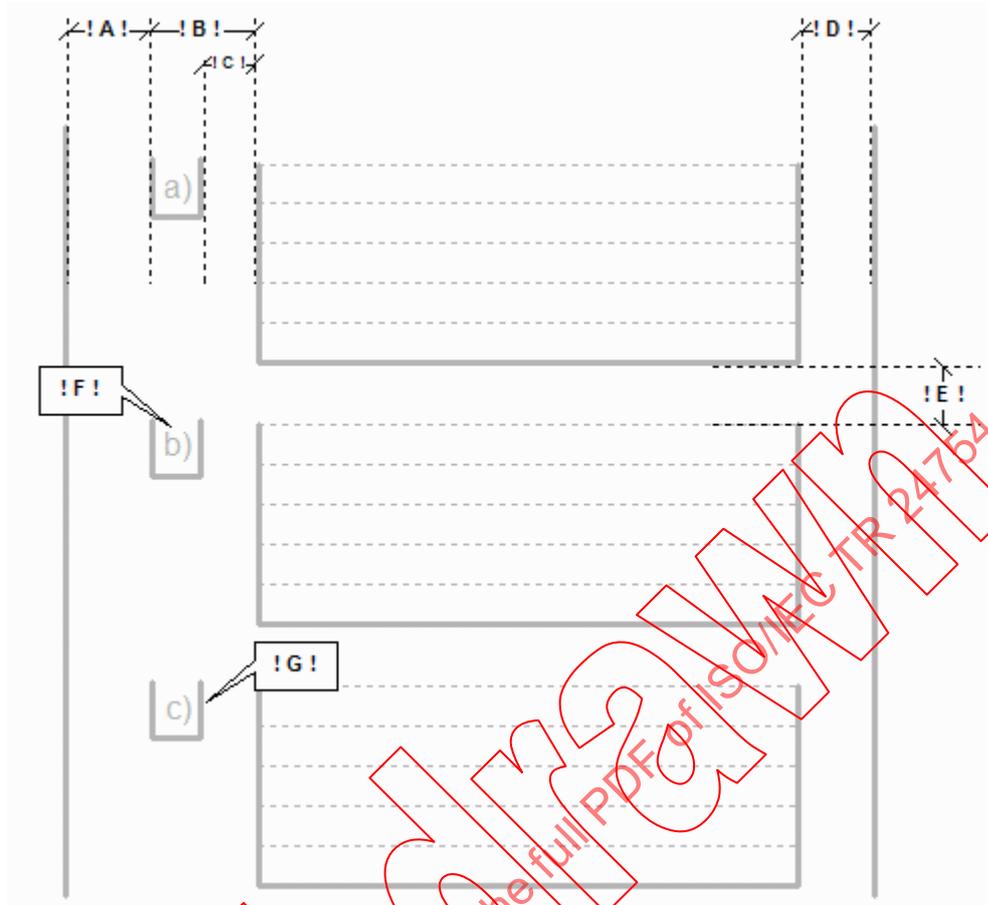
Figure 7 — Lists

8.1 List item dimensions

— A - what is the distance between the left margin of the column and the start of the label area?

— B - what is the distance between the start of the label area and the start of the body area (this will set the start of the body area to be the sum of A plus B)?

— C - what is the distance between the end of the label area and the start of the body area?

— D - what is the distance between the end of the body area and the right margin of the column?

8.2 List item identification

— F - how is each list item identified? What is the basis for any sequencing of enumerated (ordered) lists? Is there any punctuation used with the sequencing? Is any particular glyph or graphic used for itemized (unordered) lists?

— G - how is each list item label aligned within the label area? Is the alignment to the start of the label to ensure a common distance in all items between the margin and the label? Is the alignment to the end of the label to ensure a common distance in all items between the label and the body? Is the alignment to the center of the label to ensure a common center point between all labels of all items?

9 Block formatting

9.1 Blocks

Blocks of lines are the most frequently occurring construct in most printed outputs.

— which blocks have hanging indents where wrapped lines are indented further than the first line of a block, and how deep is that indent?

— what is the justification for a given block?

   e.g. FL/RR - "flush left with ragged right"

   e.g. Flandr - "flush left and right" (justified)

— what is the minimum number of lines left behind in a block that wraps the bottom of a column (orphan count)?

— what is the minimum number of lines at the end of a block that has wrapped to the top of a column (widow count)?

— is the line height determined by the tallest item on the line (a graphic, a character with a font larger than the block's font size, a superscript or subscript, etc.)? Should superscripts and subscripts be discounted from adjusting the line height? Should the line height be fixed for all lines in a block regardless of the content on the line (even distance between all text baselines)?

Block decorations can serve many purposes.

— does the block or line contain a horizontal rule for the purposes of inviting the reader to annotate the document?

— is the decoration exclusive of any document data that appears on the line, or in addition to any document data that appears on the line?

Often there is a requirement to flow blocks in parallel along the block progression direction, not quite like table columns but more like page columns.

— what are the respective widths of the blocks/columns?

— how wide is the space that separates the blocks/columns?

— is the flow of blocks in one column synchronized in any way with the flow of blocks in another column (e.g. aligned paragraph headings in parallel language translations)?

9.2 Gaps in the flow

A gap in the flow in the block direction is often required in order to leave room for the user to add their own information, or to give the impression that a lot of information is anticipated but may not be fulfilled.

In the following diagram illustrating perhaps a quiz page with questions and space for answers, what are the expectations for the gaps between content in the block-progression direction?
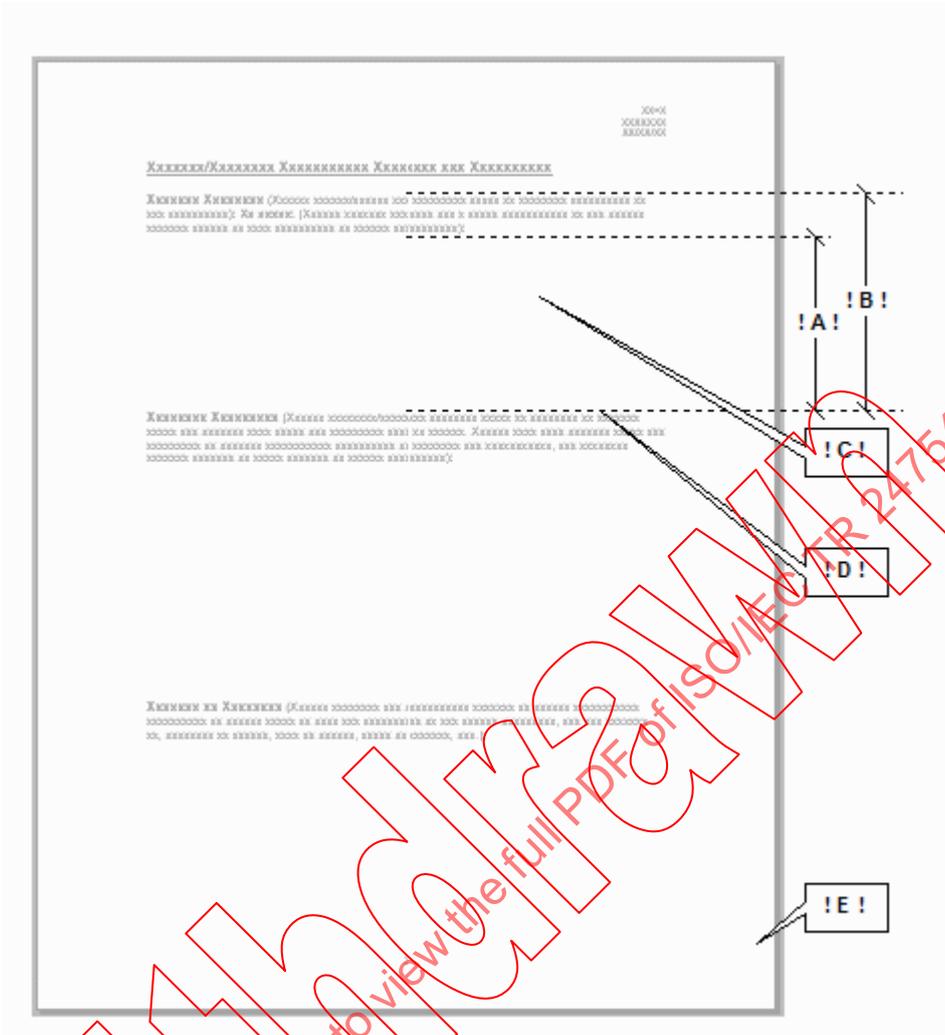
Figure 8 — Flow gaps

— A - is the size of the gap exclusive of the preamble?

— B - is the size of the gap inclusive of the preamble?

— C - is there sourced content that belongs in the gap if present in the source? If absent, are there any generated indications for the reader that the sourced content is absent?

— D - is the size only a minimum, such that it grows when the content is too long to fit? If it does grow, is there a maximum size to which it is allowed to grow? If it does not grow, is the content to be clipped without error, or is an error to be signaled?

— E - if a preamble and the size of the gap do not fit at the bottom of the page, are both moved to the top of the next page?

9.3   Keeps and breaks

When section headers are on a page, what are the rules for keeping their contents together in a column or in a page? Perhaps are there rules for keeping the first text of a section together with its respective heading? If so, how much of the content needs to be kept together?

9.4 Widows and orphans

What widows and orphans issues are important to the publication? When the end of a multi-line block flows to the next page, how many lines need to be kept together? If the lines of a section are comprised of many one-line blocks, how many of these need to be kept together when widowed from the rest of the section contents? Of the blocks that are candidates for widow-processing, are any of them conditionally present, or are they always present in the data?

10 Decoration

10.1 Graphics

Many documents incorporate graphic images for either decoration (e.g. logos) or for conveying information (e.g. photographs, charts, etc.).

⎯ in what data formats are the graphics stored?

⎯ do the graphics have captions, and if so, how are the captions to be formatted in relation to the graphic?

⎯ are there any distinctions between inline, block-level and out-of-line graphic images used in the document? How are each formatted and are out-of-line graphics displayed at the top or bottom of the formatted page?

⎯ if the graphic is to be synthesized dynamically (such as with Scalable Vector Graphics (SVG) or MathML), what is the algorithm for deriving the image from the information?

⎯ is there any scaling information relevant to the graphic (important when needing to express effective scaling results in the output)

10.2 Leaders and rules

When specifying gaps in a line in the inline direction such gaps may be elastic or inelastic. They may have a particular pattern or appearance, such as a dot-leader, a double-line rule or a grooved rule.

An elastic leader or rule can be specified with a minimum length, an optimum (initial) length and a maximum length. An inelastic rule is specified with identical values for such facets.

11 Character-level processing

11.1 Hyphenation and breaks

Does the publication have specific hyphenation requirements? What languages are being used and can tools be found to support the hyphenation of these languages? Do different areas of the document have different hyphenation requirements? Must hyphenation be turned off at any time?

Does the publication include very long strings (e.g. Internet URI strings) that may exceed the length of a line? What are the rules for breaking the string when the string length exceeds the line length?

11.2 Fonts, styles and leading

Consistency in the usage of fonts serves the document reader well. Inadvertent unnecessary changes in font use can mislead the reader to imply significance where none is intended. To help ensure consistency, it would help if every use of font, in every situation of the document, be identified from a named list of style combinations. This process also helps to minimize the number of style combinations that might, when unchecked, result in a distracting display of character presentations.

It is best to create a table of style combinations with a name and the collection of properties for each entry in the table. For the first occurrences of each use of a style combination at documented points in the information, indicate enough of the names of the style combinations to establish a pattern with which all of the remaining blocks can be inferred. Examples of style names might be `"heading1"`, `"heading1-supplemental"`, `"heading2"`, `"footer-info"`, `"header-info"`, `"mailing-label"`, `"check-table"`, `"boilerplate"`, `"customer-data"`, etc.

When setting the properties, consider the following facets:

— size of the text (e.g. `12pt`)

— line height (e.g. `14pt` which with a font size of `12pt` implies a `2pt` leading, which is implemented in XSL-FO as a 1pt half-leading)

— font family or a prioritized list of fallback font families should a particular implementation not support the desired font family

  e.g. `"Times"`, `"Times New Roman"`, `"serif"`

  e.g. `"Helvetica"`, `"Tahoma"`, `"sans-serif"`

  e.g. `"Courier"` (serif), `"Lucida Console"` (sans-serif), `"monospace"`

— any variant to the font (e.g. small caps, drop caps)

— weight of the font (e.g. bold)

— posture of the font (e.g. italic)

— decoration or scoring of the font (e.g. underline, strikethrough)

— superscript/superior and subscript/inferior font sizes and the corresponding amount of shift desired

— foreground colour and background colour

— any bordered edges?

11.3 Bidirectionality of character flow

Where in a line does a collection of generated content need to be protected from sourced content through the use of embedding levels of directionality?

12 Navigational aids

12.1 Headers, footers and headings

Navigational aids help the reader of the document easily find a given document in a set or find their place within a single document.

In the following diagram different headers and footers are seen on a set of sequential pages. Citations in the body are shown, such as headings, where the heading content is reflected in the headers. The same considerations for footers should be made as those made for headers. Page numbering is also depicted.
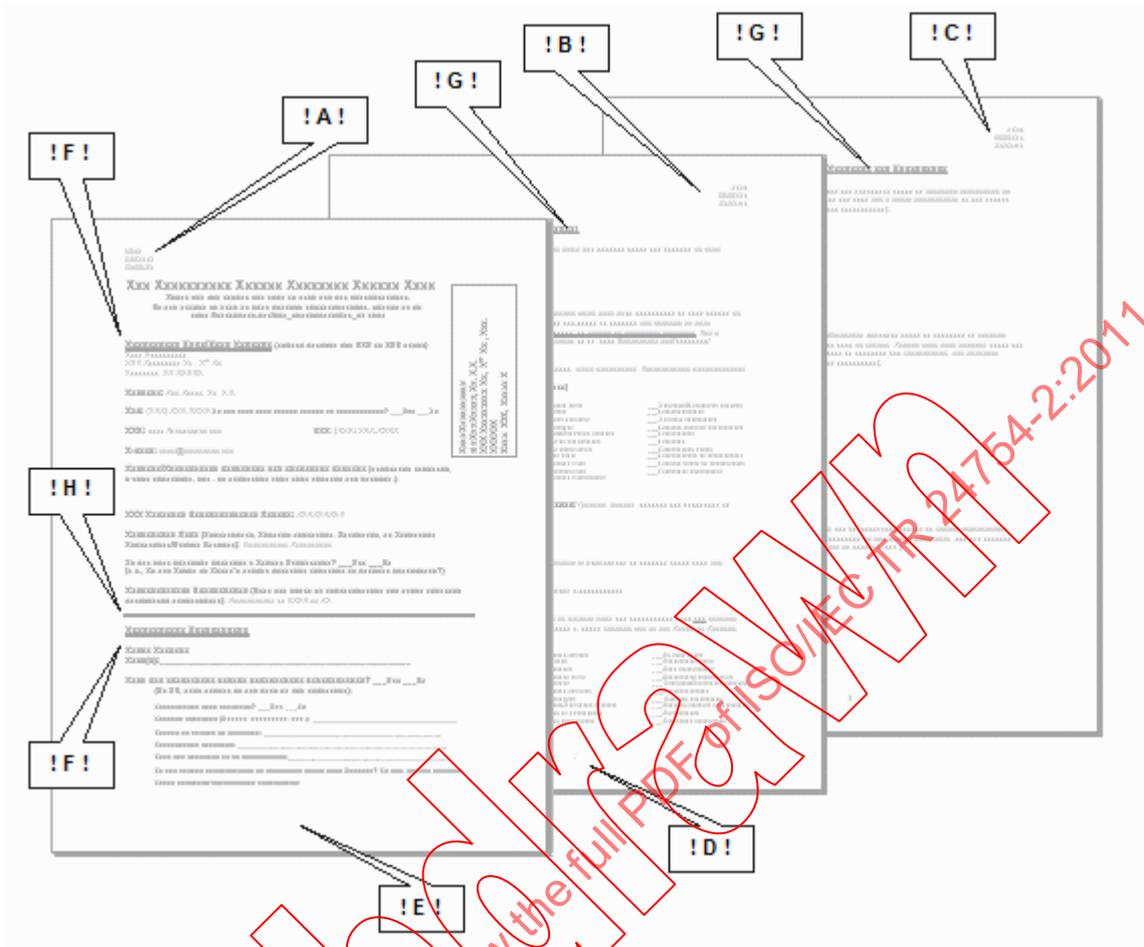
Figure 9 — Headers, footers and headings

12.1.1 Page header and footer content

— A - what information is found in the running headers and footers? Is any of it static (e.g. document type)? Is any obtained from the source files (e.g. document name)? Is any obtained during invocation (e.g. execution date)?

— B - are the running headers and footers always in the same location for every page? Is there an odd/even page requirement for the inside or outside edge of a double page presentation? Is the positioning only different for the first page and the same for all subsequent pages? Is it different on the last page? Are there any requirements for a "(continued...)" signal to show up in a header or footer when in the middle of the content of a section?

— C - how is the header and footer content aligned vertically within the extents and horizontally within the margins? Is the header content flush against the top of the header and the footer content flush against the bottom of the footer?

— D - is there a page number and, if so, where is it located? How are pages numbered (e.g. decimal, alphabetic, roman numerals, etc.)?

— E - is the page number displayed on the first page of the document?