# IEC 62605

Edition 1.0 2011-07

# INTERNATIONAL STANDARD

# NORME INTERNATIONALE





**Multimedia systems and equipment – Multimedia e-publishing and e-books – Interchange format for e-dictionaries**

**Systèmes et équipements multimédia – Publication et livres électroniques multimédia – Format d'échange pour les dictionnaires électroniques**



IEC 62605:2011

IEC Central Office
3, rue de Varembé
CH-1211 Geneva 20
Switzerland
Email: inmail@iec.ch
Web: www.iec.ch

## About the IEC

The International Electrotechnical Commission (IEC) is the leading global organization that prepares and publishes International Standards for all electrical, electronic and related technologies.

## About IEC publications

The technical content of IEC publications is kept under constant review by the IEC. Please make sure that you have the latest edition, a corrigenda or an amendment might have been published.

- Catalogue of IEC publications: www.iec.ch/searchpub
The IEC on-line Catalogue enables you to search by a variety of criteria (reference number, text, technical committee,…). It also gives information on projects, withdrawn and replaced publications.
- IEC Just Published: www.iec.ch/online_news/justpub
Stay up to date on all new IEC publications. Just Published details twice a month all new publications released. Available on-line and also by email.
- Electropedia: www.electropedia.org
The world's leading online dictionary of electronic and electrical terms containing more than 20 000 terms and definitions in English and French, with equivalent terms in additional languages. Also known as the International Electrotechnical Vocabulary online.
- Customer Service Centre: www.iec.ch/webstore/custserv
If you wish to give us your feedback on this publication or need further assistance, please visit the Customer Service Centre FAQ or contact us:
Email: csc@iec.ch
Tel.: +41 22 919 02 11
Fax: +41 22 919 03 00

---

## A propos de la CEI

La Commission Electrotechnique Internationale (CEI) est la première organisation mondiale qui élabore et publie des normes internationales pour tout ce qui a trait à l'électricité, à l'électronique et aux technologies apparentées.

## A propos des publications CEI

Le contenu technique des publications de la CEI est constamment revu. Veuillez vous assurer que vous possédez l'édition la plus récente, un corrigendum ou amendement peut avoir été publié.

- Catalogue des publications de la CEI: www.iec.ch/searchpub/cur_fut-f.htm
Le Catalogue en-ligne de la CEI vous permet d'effectuer des recherches en utilisant différents critères (numéro de référence, texte, comité d'études,…). Il donne aussi des informations sur les projets et les publications retirées ou remplacées.
- Just Published CEI: www.iec.ch/online_news/justpub
Restez informé sur les nouvelles publications de la CEI. Just Published détaille deux fois par mois les nouvelles publications parues. Disponible en-ligne et aussi par email.
- Electropedia: www.electropedia.org
Le premier dictionnaire en ligne au monde de termes électroniques et électriques. Il contient plus de 20 000 termes et définitions en anglais et en français, ainsi que les termes équivalents dans les langues additionnelles. Egalement appelé Vocabulaire Electrotechnique International en ligne.
- Service Clients: www.iec.ch/webstore/custserv/custserv_entry-f.htm
Si vous désirez nous donner des commentaires sur cette publication ou si vous avez des questions, visitez le FAQ du Service clients ou contactez-nous:
Email: csc@iec.ch
Tél.: +41 22 919 02 11
Fax: +41 22 919 03 00

# IEC 62605

Edition 1.0 2011-07

# INTERNATIONAL STANDARD

# NORME INTERNATIONALE





## Multimedia systems and equipment – Multimedia e-publishing and e-books – Interchange format for e-dictionaries

## Systèmes et équipements multimédia – Publication et livres électroniques multimédia – Format d'échange pour les dictionnaires électroniques

INTERNATIONAL ELECTROTECHNICAL COMMISSION

COMMISSION ELECTROTECHNIQUE INTERNATIONALE

PRICE CODE
CODE PRIX XF

ICS 33.160.60; 35.240.20; 35.240.30

ISBN 978-2-88912-552-4

CONTENTS

INTERNATIONAL ELECTROTECHNICAL COMMISSION

# MULTIMEDIA SYSTEMS AND EQUIPMENT – MULTIMEDIA E-PUBLISHING AND E-BOOKS – INTERCHANGE FORMAT FOR E-DICTIONARIES

## FOREWORD

1) The International Electrotechnical Commission (IEC) is a worldwide organization for standardization comprising all national electrotechnical committees (IEC National Committees). The object of IEC is to promote international co-operation on all questions concerning standardization in the electrical and electronic fields. To this end and in addition to other activities, IEC publishes International Standards, Technical Specifications, Technical Reports, Publicly Available Specifications (PAS) and Guides (hereafter referred to as "IEC Publication(s)"). Their preparation is entrusted to technical committees; any IEC National Committee interested in the subject dealt with may participate in this preparatory work. International, governmental and non-governmental organizations liaising with the IEC also participate in this preparation. IEC collaborates closely with the International Organization for Standardization (ISO) in accordance with conditions determined by agreement between the two organizations.
2) The formal decisions or agreements of IEC on technical matters express, as nearly as possible, an international consensus of opinion on the relevant subjects since each technical committee has representation from all interested IEC National Committees.
3) IEC Publications have the form of recommendations for international use and are accepted by IEC National Committees in that sense. While all reasonable efforts are made to ensure that the technical content of IEC Publications is accurate, IEC cannot be held responsible for the way in which they are used or for any misinterpretation by any end user.
4) In order to promote international uniformity, IEC National Committees undertake to apply IEC Publications transparently to the maximum extent possible in their national and regional publications. Any divergence between any IEC Publication and the corresponding national or regional publication shall be clearly indicated in the latter.
5) IEC itself does not provide any attestation of conformity. Independent certification bodies provide conformity assessment services and, in some areas, access to IEC marks of conformity. IEC is not responsible for any services carried out by independent certification bodies.
6) All users should ensure that they have the latest edition of this publication.
7) No liability shall attach to IEC or its directors, employees, servants or agents including individual experts and members of its technical committees and IEC National Committees for any personal injury, property damage or other damage of any nature whatsoever, whether direct or indirect, or for costs (including legal fees) and expenses arising out of the publication, use of, or reliance upon, this IEC Publication or any other IEC Publications.
8) Attention is drawn to the Normative references cited in this publication. Use of the referenced publications is indispensable for the correct application of this publication.

International Standard IEC 62605 has been prepared by technical area 10: Multimedia e-publishing and e-book, of IEC technical committee 100: Audio, video and multimedia systems and equipment.

The text of this standard is based on the following documents:

| FDIS | Report on voting |
|---|---|
| 100/1829/FDIS | 100/1863/RVD |

Full information on the voting for the approval of this standard can be found in the report on voting indicated in the above table.

This publication has been drafted in accordance with the ISO/IEC Directives, Part 2.

The committee has decided that the contents of this publication will remain unchanged until stability date indicated on the IEC web site under "http://webstore.iec.ch" in the data related to the specific publication. At this date, the publication will be

- reconfirmed,
- withdrawn,
- replaced by a revised edition, or
- amended.

# INTRODUCTION

Markets for multimedia e-book and e-publishing require standardization of formats for e-book data interchange among associated people; authors, data preparers, publishers and readers. The formats are classified into submission format, interchange format and reader's format. The submission format supports an interaction between authors and data preparers. The reader's format depends on e-publishing equipment. The interchange format provides an interchange format for data preparers and publishers and therefore should be e-publishing equipment independent.

The International Electrotechnical Commission (IEC) draws attention to the fact that it is claimed that compliance with this document may involve the use of patents.

Sharp Corporation has informed IEC that it has patent applications or granted patents.

The IEC takes no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured the IEC that he/she is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with the IEC. Information may be obtained from:

Sharp Corporation
Platform Technology Development Center
2613-1, Ichinomoto-cho, Tenri, Nara
632-8587 JAPAN

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those identified above. The IEC shall not be held responsible for identifying any or all such patent rights.

ISO (www.iso.org/patents) and IEC (http://patents.iec.ch/) maintain on-line data bases of patents relevant to their standards. Users are encouraged to consult the data bases for the most up to date information concerning patents.

# MULTIMEDIA SYSTEMS AND EQUIPMENT – MULTIMEDIA E-PUBLISHING AND E-BOOKS – INTERCHANGE FORMAT FOR E-DICTIONARIES

## 1 Scope

This International Standard specifies the interchange format for e-dictionaries among publishers, content creators and manufacturers.

This International Standard does not address the following aspects:

- data formats for reading devices;
- elements necessary for final print reproduction only;
- rendering issues related to physical devices;
- security issues such as DRM for document.

## 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.

IEC/TS 62229:2006, *Multimedia systems and equipment – Multimedia e-publishing and e-book – Conceptual model for multimedia e-publishing*

## 3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

**3.1**
**manufacturer**
organization or person that manufactures hardware and/or software of the e-book

## 4 Position and requirements for interchange format for E-dictionaries

### 4.1 Interchange format for e-dictionaries in contents creation/distribution model

The conceptual model for multimedia e-publishing (IEC/TS 62229) defines a contents creation/distribution model shown in Figure 1.

**Author** <--(1)--> **Data preparer** <--(2)--> **Publisher** --(3)--> **Reader**

*IEC 1383/11*

**Key**

(1) content data in submission format

(2) content data in interchange format

(3) content data in reader's format

**Figure 1 – Contents creation/distribution model**

It should be noted that the role of manufacturers of e-dictionary hardware and software overlaps that of the publisher in Figure 1. Therefore, a slightly modified model will be assumed for this International Standard, as shown in Figure 2.

**Author** <--(1)--> **Data preparer** <--(2)--> **Publisher (manufacturer)** --(3)--> **Reader**

IEC 1384/11

**Figure 2 – Contents creation/distribution model** (modified)

This International Standard specifies the interchange format between data preparers and publishers, i.e. a format for (2) in Figure 2 though it may be used as a reader's format.

### 4.2 Requirements for interchange format for e-dictionaries

An interchange format for e-dictionaries needs to address the following.

- Description of keywords, links from the keywords to entries (link data) and the order of the entries.
- Description of articles for each entry (entry data). This includes text, image, and other multimedia functionalities generally required for e-books.
- Description of bibliographical data and other data. This should include the name of the author and the publisher, the title of the content and the explanatory note. The relationship between these concepts is visually represented in Figure 3.
- Description of contents written in various languages.

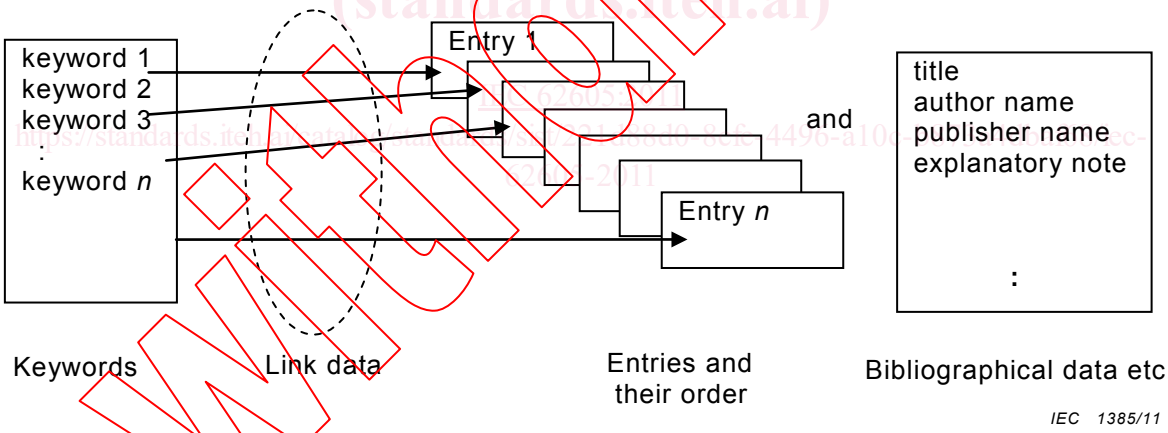


**Figure 3 – Relationship between concepts**

## 5 File format

This International Standard is based on XMDF (as described in IEC 62448:2009, Annex B) and LeXML. The standard format is represented in XML and hereafter called XMDF-LeXML format.

NOTE LeXML is proposed by Digital ASSIST Ltd. Its original specifications are found at http://www.d-assist.com/index.html.

## 6 Semantics

Elements of the XMDF-LeXML format can be rendered in accordance with appropriate style specifications, which are outside the scope of this International Standard.

## 7 File format details

This Clause specifies the XMDF-LeXML format mentioned in Clause 5.

The XMDF-LeXML format is an interchange format for e-dictionaries multimedia e-book data interchange, targeted at data preparers and publishers rather than the reader, with an emphasis on mobile devices as a target platform. Much like HTML, this format does not split the document in fixed pages, but determines the layout according to the viewer device's display size, the font in use, and so on. In this standard, such contents will be referred to as flowing content, as opposed to paged content.

## 8 Overview of the format's structure

Flowing contents are usually composed of several concatenated flows. This standard makes no particular requirement concerning the way the flowing content should be split into individual flows. This decision is left to the data preparer, to accommodate the various type of contents. For instance, a newspaper may have one flow per article, a novel one per chapter, and so on. It is also possible not to split the content, and to have only one flow. However, it should be noted that particularly large flows, or an extremely large number of flows, may impact on runtime performance, depending on the specific version of the viewer in use, the available memory, and so on.

The XML tree structure of the format is shown in Figure 4.



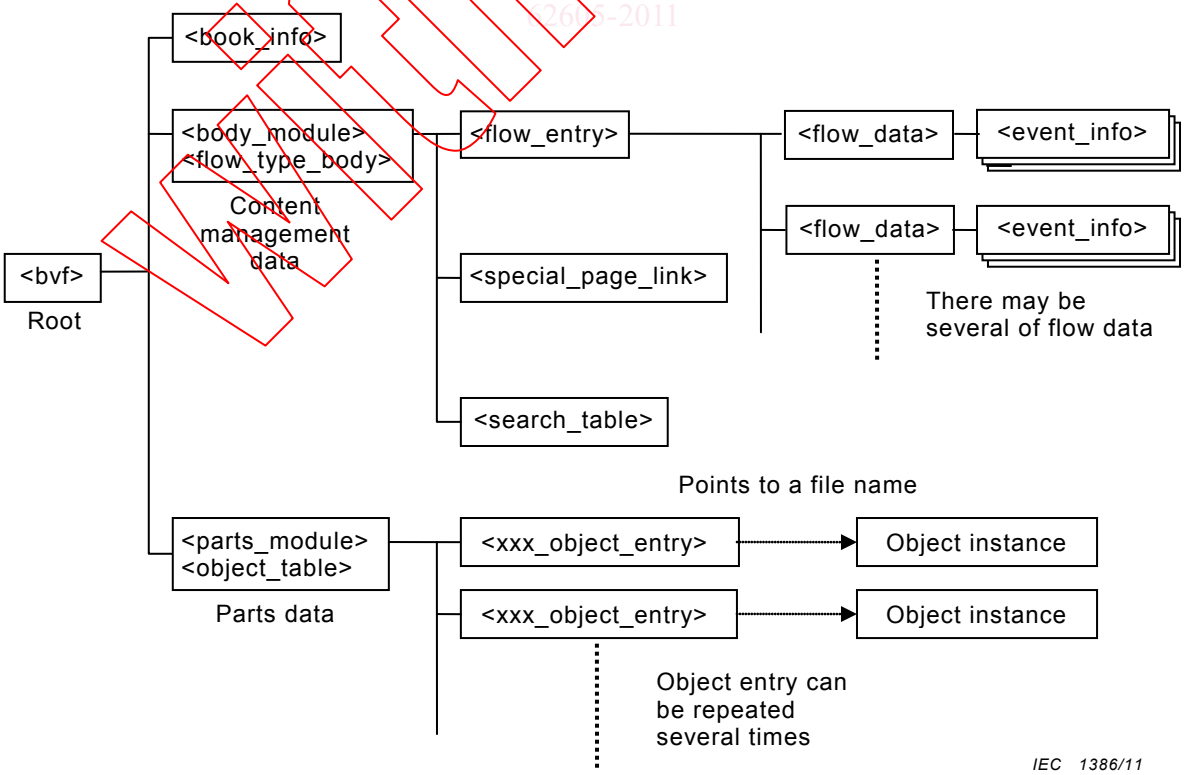


**Figure 4 – XML tree structure**

The actual contents of each flow, in other words, what will be displayed by the viewer, is recorded in the *object instance*. The *object instance* is registered in *object_entry*, and associated with an ID number and other auxiliary data, turning it into playable / displayable data. *Flow_data* determines its content by pointing at such registered objects. In addition, information on functionalities such as page link is recorded in *event_info*.

The main part of this standard is generic, and may be used for any country and language. However, some parts may have language specific behavior. To keep the main part of the specification reasonably light and focused, those localization-related issues are detailed in Clause A.2. The main text will refer to them, when appropriate.

# 9 Elements and attributes

## 9.1 General

The different types of values that may be used in the various tags or attributes are explained below. The elements and attributes detailed below will be valid throughout this standard, and will be referenced by other constructs. In the following explanations, alphanumeric characters refer to numerals from 0 to 9 and alphabetic letters from a to z and A to Z.

## 9.2 Page_ID

Page_ID specifies a unique identification number for the flow data of the flowing contents. It is a string starting by the "PG" characters, followed by alphanumeric characters.

Example:

```
<flow_data flow_id="PG0002" ... />
```

## 9.3 Object_ID

Object_ID specifies a unique identification number for objects used in the flowing contents. It is a string starting by the "OB" characters, followed by alphanumeric characters.

Example:

```
<dynamic_text_object_entry id="OB0ue4".../>
```

## 9.4 Char_ID

Char_ID specifies an identification number for positions (character strings, etc.) within text and dictionary data objects. It is an alphanumeric string which is to be given uniquely in the *text* (see 10.6.2) *and dictionary data object instance* (see 10.6.3).Char IDs with the same value in different object instances are regarded as separate and don't affect each other.

Example:

```
<trigger_pointer id="OB29s0/CR0de4"/>
Click<char_id char_id="CR0001">here</char_id>for details.
```

## 9.5 Reading

For sorting purposes, reading may be useful to specify the reading of each word. Restricting the characters allowed for this purpose to a limited set makes it easier to define the sorting method. Such characters should be determined on a per language basis. All languages can

use the characters listed in Table 1 as a common base, while the localization (see Clause A.2) will describe the language specific extensions to it.

**Table 1 – Base characters for reading**

| Name | Corresponding characters [a)] |
|---|---|
| Basic alphabet | A to Z (0x0041 to 0x005A)<br>A to z (0x0061 to 0x007A) |
| Numerals | 0 to 9 (0x0030 to 0x0039) |
| Others | space (0x0020), ((0x0028), ) (0x0029) |
| a) All values are in Unicode. | |

Example:

```
<title reading="PI">π</title>
```

### 9.6 Filename

Filenames should be written using the following convention. The path is relative to the file in which this reference is made. Network paths should not be used. For portability concerns, it is recommended that only ACSII characters be used. Both the slash and backslash characters are acceptable as directory separators. It is also recommended not to use excessively long filenames, as those might not be supported by the host operating system.

Example:

```
<dynamic text object entry src="sect1.xml" type="text/x-bvf-text" id="OB0ue4"/>
```

### 9.7 Standard character

The standard character set of the document, as set by the default_ccs attribute of the <bvf> tag (see 10.2), is to be chosen from a well defined list, so as to ease the development of viewing software. However, this list may change for different localized versions of the XMDF-LeXML format. Any e-book data has to define its standard character set as one of or a combination of character set(s) listed in Table 2 and those defined for a specific localization (see A.2.2).

**Table 2 – Standard character set**

| Character set name | Description |
|---|---|
| "ISO 646-IRV" | Characters in the range of US-ASCII |

### 9.8 Standard character string

A string composed of *Standard characters* is called a *Standard character string*. Unless specified otherwise, the spacing characters (space (0x0020), linefeed (0x000D, 0x000A, 0x0D0A), tabulation (0x0009)) are to be handled as follows:

Space (0x0020) to be displayed as is.

Linefeed (0x000D, 0x000a, 0x0D0A) not to be displayed, but simply ignored.

Tabulation (0x0009) to be displayed as if it were a single space.

Furthermore, because of restriction in the XML format, linefeeds (0x000D, 0x000A, 0x0D0A) and tabulations (0x0009) in attribute values should be replaced by spaces when converting to the distribution format.

### 9.9 Extended character

Characters which have Unicode code points while not being among those listed below are referred to as *Extended characters*.

*Standard characters*

Surrogate pair range (0xD800 to 0xDFFF)

BOM (Byte Order Mark) (0xFFFE,0xFEFF)

NON CHARACTER (0xFFFF)

Control characters (characters between 0x0000 and 0x001F except tabulation (0x0009) and linefeed (0x000A, 0x000D), as well as DEL (0x007F)).

If an e-book indeed uses any *Extended character* in its data, the name of a charset covering those *Extended characters* should be appended to the *default_ccs* attribute of the <bvf> tag. Note that all *Extended characters* used in the document do not need to be covered by the same character set, as it is possible to specify several ones.

### 9.10 Extended character string

A string composed of *Standard characters* and *Extended characters* is called an *Extended character string*. Unless specified otherwise, the spacing characters (space (0x0020), linefeed (0x000D, 0x000A, 0x0D0A), tabulation (0x0009)) are to be handled the same way as in *Standard character strings*.

### 9.11 External character

To display a character which is neither a *Standard character* nor an *Extended character*, it is possible to use the <external_char> tag described below.

<external_char> inserts an *External character*. The viewer may display it according to the following methods.

a) Display the character set by the alt_set and alt_code attributes.

b) Display the image set by the alt_img or alt_vimg attributes.

c) Display the alternative letter set by the alt attribute.

It has the following attributes.

[Attributes]

alt_set: Together with the alt_code attribute, it allows designating the *External character* to be used. This alt_set attribute indicates the font name, while the alt_code attribute indicates the character code point within the font. The alt_set attribute is written in the following way:

alt_set = "font1,font2, ..."

The alt_set attribute may hold several font names, separated by "," (0x002C). In that case, the viewer should use the first font of the list that is available (either from the platform, or included in the contents data itself) to display the character.

alt_code: Selects a character code point in the font specified by the alt_set attribute. It may be written both as a decimal number, or a hexadecimal number, prefixed by "0x". In case several fonts have been defined in the alt_set attribute, the character code shall represent the same character in all of them. This attribute can be omitted.

alt_img: Defines an alternative character image. Written as a *Filename*. Before opening the file indicated by this attribute, the img_type attribute should be checked for authorized file types. Note that it may be used only when