ICS
CCS 点击此处添加 CCS 号

T/CADPA

中 国 音 像 与 数 字 出 版 协 会 团 体 标 准

T/CADPA XXXX—XXXX

电子图书出口海外数据要求

Requirements on processing the data of e-book for export purpose

（征求意见稿）

（本草案完成时间：2023-09-28）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中国音像与数字出版协会　　发 布

目　　次

前言 III
1　范围 1
2　规范性引用文件 1
3　术语和定义 1
4　成品数据构成 2
4.1　文件结构 2
4.2　文件存储结构 2
4.3　文件命名 2
4.3.1　编号规则 2
4.3.2　命名方法 3
5　元数据要求 3
5.1　加工规格 3
5.2　基本元数据 3
5.2.1　标识数据组 3
5.2.2　名称数据组 4
5.2.3　责任者数据组 4
5.2.4　基本特征数据组 4
5.2.5　主题数据组 5
5.2.6　出版信息数据组 5
5.3　授权元数据 6
5.3.1　授权销售数据组 6
5.3.2　授权使用范围数据组 6
5.4　元数据质量要求 7
6　封面图像要求 7
6.1　封面图像加工规格 7
6.2　封面图像质量要求 7
7　全文要求 7
7.1　矢量化版式文件 7
7.1.1　加工范围 7
7.1.2　加工规格 7
7.1.3　质量要求 7
7.2　流式文件 7
7.2.1　加工范围 7
7.2.2　加工规格 8
7.2.3　质量要求 8
7.3　图像类对象数据 8
7.3.1　加工规格 8
7.3.2　质量要求 8

7.4 结构化对象数据 8
7.4.1 加工规格 8
7.4.2 质量要求 8
7.5 矢量化文字对象数据 8
7.5.1 加工规格 8
7.5.2 质量要求 8
附录A（资料性） 信息采集相关代码表 9
A.1 受众对象范围代码表 9
A.2 主题分类标识代码表 9
A.3 货币代码表 9
参考文献 10

前　　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由中国音像与数字出版协会出版融合工作委员会提出。

本文件由中国音像与数字出版协会归口。

本文件起草单位：

本文件主要起草人：

电子图书出口海外数据要求

1 范围

本文件规定了电子图书出口海外的成品数据构成、元数据、封面图像和全文的加工要求。
本文件适用于电子图书出口海外的数据采集、加工与制作。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 5271.17—2010 信息技术 词汇 第17部分：数据库
GB/T 30330—2013 中国出版物在线信息交换图书产品信息格式规范
GB/T 38548.3—2020 内容资源数字化加工 第3部分:加工规格
GB 18030-2022 信息技术 中文编码字符集
GB/T 4880.2-2000 语种名称代码 第2部分:3字母代码
GB/T 2659—2000 世界各国和地区名称代码
CY/T 96—2013 电子书内容术语
CY/T 101—2014 新闻出版内容资源加工规范

3 术语和定义

GB/T 5271.17、CY/T 96、CY/T 101界定的以及下列术语和定义适用于本文件。

3.1
电子图书 e-book
通过相关设备直接呈现文字、图像、音频、视频等内容，具有相当篇幅的数字专题出版产品。
[来源：CY/T 96—2013，4.2.5]

3.2
成品数据 finished data
已经完成全部数据处理过程且符合预先设定的规格和质量要求，可实现交付的最终数据形态。
[来源：CY/T 101.1—2014，7.1.1]

3.3
元数据 metadata
关于数据或数据元素的数据（可能包括其数据的描述），以及关于数据拥有权、存取路径、访问权和数据易变性的数据。
[来源：GB/T 5271.17—2010, 17.06.05]

3.4
数据组 data group
一组相同业务属性的元素和复合元素。
[来源：GB/T 30330—2013, 3.3]

3.5
流式文件 reflowing document
按照内容逻辑顺序，内容呈现可适应终端设备屏幕或窗口变化的一种文件。
[来源：CY/T 101.1-2014，6.3.12]

3.6
双层矢量化版式文件 vectorized dual-layer fixed-layout document

通过格式转换，生成与原加工对象版式完全一致的图像文件集合，且具有图像层对应的透明字体模式的文字层，可支持选取、拷贝和查找的版式文件。

[来源：CY/T 101.1—2014，6.3.13]

3.7

单层矢量化版式文件 vectorized fixed-layout document

按照原加工对象的文字位置，文字采用矢量字，修饰性图片、艺术字、底纹、线条、图表和公式等以图像形式显示的版式文件。

[来源：CY/T 101.1—2014，6.3.14]

4 成品数据构成

4.1 文件结构

电子图书成品数据由元数据文件、封面图像文件、矢量化版式文件和流式文件构成。矢量化版式文件必须包含单层矢量或者双层矢量中的一种文件。

4.2 文件存储结构

以单册图书为独立的文件夹，文件夹包括元数据文件、封面图像文件、全文文件，不对文件夹进行压缩，所有文档存储于其中，存储结构见图1。

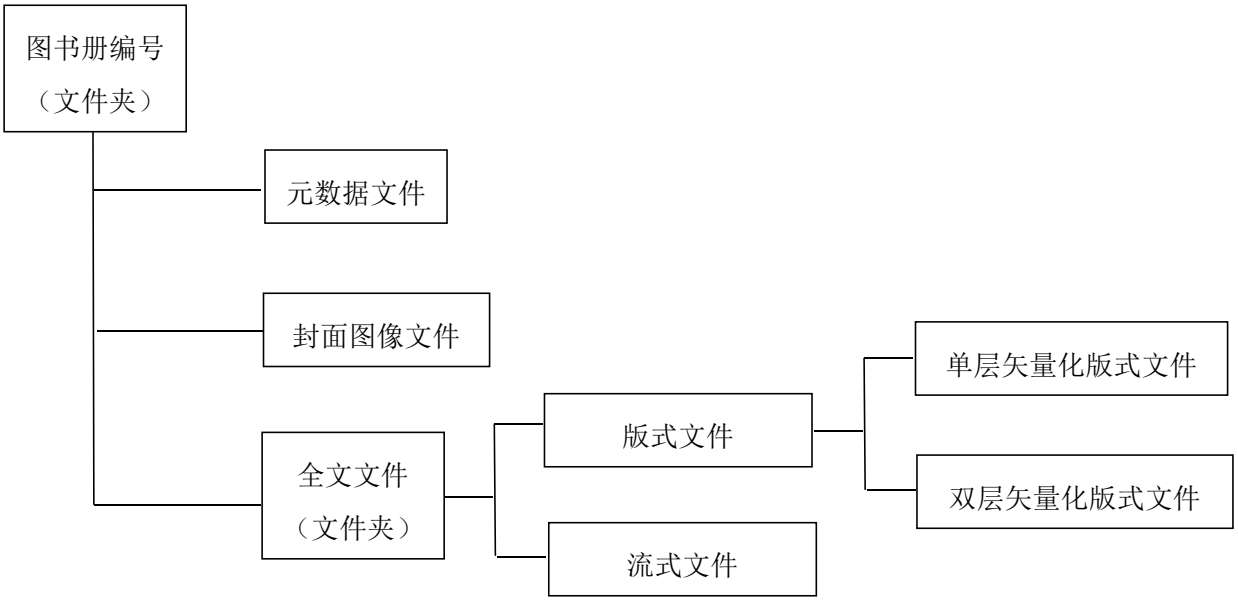

图1 文件存储结构

4.3 文件命名

4.3.1 编号规则

图书加工成品的命名遵循种、册独立编号的原则，不宜在命名中增加各种分类信息或者其他辅助标记信息，见表1。

表1 基本编号规则

类别	编号规则
图书种类编号	13位数字组成，ISBN

图书册编号	3位数字组成，每册从001记起，每册记1，丛书编号按照上中下的顺序编号

4.3.2 命名方法

图书名出现英文输入法状态不显示的字符：< > / \ | : ″ * ?时替换为_。

4.3.2.1 文件夹命名方法

文件夹命名方法：图书种类编号+图书册编号+图书名
示例：X:\9787516826447-001-中国历史1

4.3.2.2 元数据文件命名方法

元数据文件命名方法：图书种类编号+图书册编号+图书名
示例：9787516826447-001-中国历史1.xml

4.3.2.3 封面图像文件命名方法

封面图像文件命名方法：图书种类编号+图书册编号+图书名
示例：9787516826447-001-中国历史1.jpg

4.3.2.4 单层矢量版式文件命名方法

单层矢量版式文件命名方法：图书种类编号+图书册编号+图书名+D
示例：9787516826447-001-中国历史1_D.pdf

4.3.2.5 双层矢量版式文件命名方法

双层矢量版式文件命名方法：图书种类编号+图书册编号+图书名+S
示例：9787516826447-001-中国历史1_S.pdf

4.3.2.6 流式文件命名方法

流式文件EPUB命名方法：图书种类编号+图书册编号+图书名
示例：9787516826447-001-中国历史1.epub

5 元数据要求

5.1 加工规格

以XML1.0及以上版本格式存储，其中字符宜采用GB 18030字符集，编码方式宜采用UTF-8，元数据名称标签兼容GB/T 30330—2013的规定。

5.2 基本元数据

5.2.1 标识数据组

用于描述电子图书标识数据元的数据组，见表2。

表 2 标识数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
标识数据组	<Identifier>	用于描述电子图书标识数据元的数据组	字符	必选	不重		
国际标准书号	<IDValue>	图书的13位ISBN号	数字	有则必备	不重		
记录唯一标识	<RecordReference>	用于表示每种电子图书记录的唯一标识，并作为每次发送更新信息时	字符	必选	不重	见4.3.2.1	

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
		的永久标识					

5.2.2 名称数据组

用于描述电子图书名称数据元的数据组，见表3。

表 3 名称数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
名称数据组	<Title>	用于描述电子图书名称数据元的数据组	字符	必选	不重		
套书名	<TitleDetail>	用于描述电子图书的正题名	字符	必选	不重		
书名	<TitleText>	用于描述电子图书的正题名	字符	必选	不重		
副书名	<Subtitle>	对正题名作进一步解释说明而不独立使用的名称	字符	可选	不重		
全文格式类型	<Format>	电子图书的全文格式	字符	必选	可重	1=单层矢量化版式文件 2=双层矢量化版式文件 3=流式文件 4=其他	
全文文件名称	<FileName>	电子图书的全文文件名称	字符	必选	不重	见 4.3.2.4、4.3.2.5、4.3.2.6	

5.2.3 责任者数据组

用于描述电子图书责任者数据元的数据组，见表4。

表 4 责任者数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
责任者数据组	<Contributor>	用于描述电子图书责任者数据元的数据组	字符	必选	不重		
责任者角色	<ContributorRole>	用于表示个人或团体在电子图书创作时所起作用的作者、编者、译者、摄影者、绘画者等	字符	必选	不重	见GB/T 30330—2013附录表B.11	
责任者名称	<PersonName>	用于描述电子图书创作人或者团体以正序方式显示的名称，通常填写作者	字符	必选	可重		
责任者介绍	<BiographicalNote>	用于描述电子图书提供者个人简介或对团体的介绍	字符	有则必选	不重		

5.2.4 基本特征数据组

用于描述电子图书基本特征数据元的数据组，见表5。

表 5　基本特征数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
基本特征数据组	< Feature>	用于描述电子图书基本特征数据元的数据组	字符	必选	不重		
语种	<LanguageCode>	用于描述语种的代码	字符	必选	可重	见GB/T 30330—2013附录表B.60	
页数	<NumberOfPages>	用于描述有页码电子图书中的文本项的页数	数值	可选	不重		
版次号	<EditionNumber>	用于描述版本次序的编号	数值	可选	不重		
受众对象	<AudienceCode>	用于描述电子图书面向的受众对象	数字	可选	可重	见附录 A.1	
推荐语	<Text>	用于描述编辑对电子图书加以点评的内容	字符	有则必选	可重		
获奖名称	<PrizeName>	用于描述电子图书获奖的奖项名称	字符	有则必选	可重		

5.2.5　主题数据组

用于描述电子图书主题数据元的数据组，见表6。

表 6　主题数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
主题数据组	<Subject>	描述电子图书主题数据元的数据组	字符	必选	不重		
主题类型	<SubjectSchemeIdentifier>	用于表示采用的主题表方案	数字	必选	不重	见附录A.2	
分类号	<SubjectCode>	用于描述主题表标识规定方案的代码	字符	必选	可重		
关键词	<SubjectHeadingText>	描述电子图书内容的主要词语	字符	可选	可重		
内容简介	<Text>	用于描述电子图书的介绍信息	字符	必选	不重		

5.2.6　出版信息数据组

用于描述电子图书出版信息数据元的数据组，见表7。

表 7　出版信息数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
出版信息数	< Publisher>	用于描述电	字符	必选	不重		

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
据组		子图书出版信息数据元的数据组					
出版社	<PublisherName>	于描述电子图书出版相关机构的名称	字符	必选	可重		
出版国	<CountryOfPublication>	用于表示电子图书出版所在国家的代码	字符	可选	不重	见 GB/T 2659—2000	
出版地	<CityOfPublication>	用于描述与出版说明或出版者相关的城市名称	字符	可选	不重		
出版日期	<PublishingDate>	用于描述电子图书出版的日期	数字	必选	不重		样式为yyyymmdd的八位数字

5.3 授权元数据

5.3.1 授权销售数据组

用于描述电子图书授权销售数据元的数据组，见表8。

表 8 授权销售数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
授权销售数据组	<SalesRights>	用于描述电子图书授权销售数据元的数据组	字符	必选	不重		
授权可销售国家	<CountriesIncluded>	用于描述可授权销售的国家	字符	必选	可重	见 GB/T 2659—2000	
授权不可销售国家	<CountriesExcluded>	用于描述不可授权销售的国家	字符	必选	可重	见 GB/T 2659—2000	
授权起始日期	<StartDate>	用于描述电子图书授权的开始日期	数字	必选	不重		样式为yyyymmdd 的八位数字
授权截止日期	<EndDate>	用于描述电子图书授权的结束日期	数字	必选	不重		样式为yyyymmdd 的八位数字
价格类型	<PriceType>	用于描述价格(<Price>)复合元素中价格金额元素采用的价格的ONIX 代码	数字	必选	不重	02=电子书销售价格 42=纸书定价	
价格	<PriceAmount>	用于描述出版者赋予纸质图书或电子图书的价格	数字	有则必选	不重		
币制	<CurrencyCode>	用于描述价格所采用的 GB/T 12406 货币代码	字符	必选	不重	见附录 A.3	

5.3.2 授权使用范围数据组

用于描述电子图书授权使用范围数据元的数据组，见表9。

表 9 授权使用范围数据组

中文名称	英文标签	描述	数据类型	必备性	可重复性	取值	备注
授权使用范围数据组	<DRM>	用于描述电子图书授权使用范围数据元的数据组	字符	必选	不重		
试读范围	<PreviewAllowed>	用于描述未经购买即可预览的内容篇幅	数值	必选	不重		样式为百分比
下载条款	<DownloadAllowed>	用于描述购买后允许下载的内容篇幅	数值	必选	不重		样式为百分比
打印范围	<PrintAllowed>	用于描述购买后允许打印的内容篇幅	数值	必选	不重		样式为百分比

5.4 元数据质量要求

元数据的规范性、有效性、准确性应符合CY/T 101.4—2014中4.4、4.5、4.6的规定。

6 封面图像要求

6.1 封面图像加工规格

封面图像加工规则要求见表10。

表 10 封面图像加工规格

格式	精度	尺寸	大小
JPEG	不低于 150DPI	宽度固定 600 像素点，高度按原比例缩放	不超过 200KB

6.2 封面图像质量要求

按CY/T 101.4—2014中5.5.2的规定，图像质量的差错率要求为千分之一以下。

7 全文要求

7.1 矢量化版式文件

7.1.1 加工范围

从封面到封底所有页面。

7.1.2 加工规格

宜采用PDF1.6及以上版本，规格见CY/T 101.3—2014的表2。

7.1.3 质量要求

矢量化版式文件质量要求如下：

a) 文字质量要求应符合CY/T 101.4—2014中5.5.1质量管理的规定；
b) 按照CY/T 101.4—2014中5.5.5规定，版式文件质量的差错率要求为万分之三以下；
c) 按照CY/T 101.4—2014中5.5.2规定，图像质量的差错率要求为千分之一以下；
d) 按照CY/T 101.4—2014中5.5.4规定，关联关系质量的差错率要求为千分之三以下。

7.2 流式文件

7.2.1 加工范围

从封面到封底所有页面。

7.2.2 加工规格

宜采用EPUB2.0及以上版本，规格见CY/T 101.8—2014的表14。

7.2.3 质量要求

流式文件质量要求如下：

a) 文字质量要求应符合CY/T 101.4—2014中5.5.1质量管理的规定；
b) 遵循CY/T 101.4—2014中5.5.6的相关规定，流式文件质量的差错率要求为万分之三以下；
c) 遵循CY/T 101.4—2014中5.5.2的相关规定，图像质量的差错率要求为千分之一以下；
d) 遵循CY/T 101.4—2014中5.5.4的相关规定，关联关系质量的差错率要求为千分之三以下。

7.3 图像类对象数据

7.3.1 加工规格

图像类对象采用JPEG等格式存储,其中生僻字图片采用GIF格式存储,其他图像类对象数据采用JPEG等格式存储，规格见GB/T 38548.3—2020的表1。

7.3.2 质量要求

按CY/T 101.4—2014中5.5.2的规定，图像类对象数据质量的差错率要求为千分之一以下。

7.4 结构化对象数据

7.4.1 加工规格

表格、公式加工按照GB/T 38548.3—2020，6.1的规定：

a) 公式、表格宜采用有损压缩 JEPG 文件类型为存储格式；
b) 结构化数学公式,宜采用 MathML文件类型为存储格式；
c) 结构化化学式，宜采用 CML文件类型为存储格式；
d) 结构化表格,宜采用 HTML文件类型为存储格式。HTML文件应遵循HTML5.0以上版本规范。

7.4.2 质量要求

按CY/T 101.4—2014中5.5.3的规定，结构化对象数据质量的差错率要求为万分之三以下。

7.5 矢量化文字对象数据

7.5.1 加工规格

按CY/T 101.8—2014中4.4.3.3的规定，特殊字符集采用TTF(TrueTypeFont）格式存储。

7.5.2 质量要求

按CY/T 101.4—2014中5.5.1的规定，矢量化文字对象数据质量的差错率要求为万分之一以下。

附 录 A
（资料性）
信息采集相关代码表

A.1 受众对象范围代码表

受众对象范围代码表取自于ONIX 3.0 List28见表A.2。

表 A.1 受众对象范围代码表

代码	代码名称	代码说明
01	普通大众	非专业的成年观众
02	儿童/青少年	对青少年观众来说，不是专门为了任何教育目的
03	青年人	对青年观众来说，不是专门为了任何教育目的
04	小学和中学/小学和高中	幼儿园、学前、小学/小学或中学/高中教育
05	大学/高等教育	继续教育和高等教育的大学和学院
06	专业和学术	面向成年专家观众，包括专业发展和学术研究

A.2 主题分类标识代码表

主题分类标识代码表取自于ONIX 3.0 List27见表A.3。

表 A.2 主题分类标识代码表

代码	代码名称	代码说明
10	BISAC主题标目	美国图书业协会推出，在北美市场用于根据主题内容对书籍进行分类
88	中图分类法	中国图书馆分类法

A.3 货币代码表

货币代码表取自于ONIX 3.0 List96见表A.1。

表 A.3 货币代码表

代码	代码名称	代码说明
CNY	人民币	
USD	美元	
EUR	欧元	
GBP	英镑	
CAD	加拿大元	
AUD	澳大利亚元	
NZD	新西兰元	
HKD	港元	
MXN	墨西哥比索	
INR	印度卢比	
RUB	卢布	
BRL	巴西里亚尔	

参 考 文 献

[1] CY/T 101.1—2014 新闻出版内容资源加工规范 第1部分：加工专业术语
[2] CY/T 101.8—2014 新闻出版内容资源加工规范 第8部分：图书加工
[3] CY/T 96—2013 电子书内容术语
[4] GB/T 7408—2005 数据元和交换格式 信息交换 日期和时间表示法