中文 DOI 电子书注册 XML 文件数据说明

序号	分类	数据元素	描述	格式说明	例子
1	批处理文件	doi_batch	批处理文件的开始/结束标志。	必须，属性version标识目前的版本为2.0.0版。	<doi_batch version = “2.0.0”>
2	头标区	head	标识头标区的开始/结束。	必须。	<head> <doi_batch_id>20120101</doi_batch_id> <timestamp>20120716614950</timestamp> <depositor> <name>Name of the Depositor</name> <email_address>xyz@authorized_depositor.com </email_address> </depositor> <registrant>Name of Registrant</registrant> </head>
3		doi_batch_id	批处理文件的唯一标识号,由出版商产生。	必须，格式及内容由出版商决定。出版商应该为提交的每个批处理文件产生唯一的doi_batch_id。	
4		timestamp，不得超过17个字符	时间戳，标识批处理数据产生的时间或版本。	必须，一般采用数据产生的日期和时间的整型表达式。格式可由出版商确定。高版本（时间戳较新）的数据可取代低版本（时间戳较旧）的数据。	
5		depositor	存储者，发送批处理文件的组织的相关信息。	必须。	
6		name	进行数据存储的组织的名称。	必须。	
7		email_address	组织中接收有关反馈信息的电子信箱地址。	必须。	
8		registrant，不得超过 130个字符	注册者，拥有注册内容的组织的信息。	必须。	
9	主体	body	标识全部注册数据的开始/结束。	必须。包含一个或多个DOI注册元数据集,如一个或多个电子书的DOI注册元数据。	
10	电子书DOI注册数据	book	标识一本电子书的全部DOI注册数据的开始/结束。	必须。包含一本电子书的全部DOI注册数据。	<book> <book_metadata>...</book_metadata> <content_item>...</content_item> <content_item>...</content_item> </book >
11	电子书元数据	book_metadata	电子书的元数据。	必须，每本电子书只能包含一组电子书元数据，并且可以在属性language中定义基于ISO-639标准的的语言类型	<book_metadata language="zh"><contributors>...</contributors> <titles><title>大众密码学</title></titles> <isbn media_type="electronic">978-7-04-017267-6</isbn> <publisher> <publisher_name>高等教育出版社</publisher_name> </publisher> <doi_data><doi>10.3868/b.isbn.978-7-04-017267-6</doi> <resource><![CDATA[http://bookonline.hep.com.cn/xpe/portal/24cb6c01-11c9-1000...]]></resource></doi_data> </book_metadata>

序号	分类	数据元素	描述	格式说明	例子
12		contributors	责任者信息。	可选，每个可含 1-255 个人名（person_name）或机构名称（organization）。	<contributors> <person_name sequence="first" contributor_role="author">毛明</person_name> <person_name sequence="additional" contributor_role="editor">关旭</person_name> </contributors>
13		person_name	人名，每个长度不得超过 450 字符。	可选。含两个属性。 属性 sequence 说明责任者的顺序：**•"first"**－第一责任者；**• "additional"**－除第一责任者以外的其他责任者。必须。 属性 contributor_role 说明责任者的角色：**•"author"**－作者；**•"editor"**－编者；**•"translator"** －译者。必须。	
14		organization	机构名称，每个长度不得超过 450 字符。	可选。含两个属性。属性的填写方法同 35 行。 如果多个组织创建了一个实体，每个组织应填写在一个单独的 organization 元素中。如果一个实体由多个从属于一个或多个组织中的个人创建，则 person_name 和 organization 可以在 contributors 中自由混杂排列。	
15		titles	文章题名信息。	必须，每个 book_metadata 一个。	<titles> <title>大众密码学</title> </titles>
16		title	文章题名，不得超过 900 个字符。	必须，每个 titles 一个。若一个题名包含副标题，最好将副标题部分填写在 subtitle 中。	
17		original_language_title	源语言标题，不得超过 512 个字符。	可选，每个 titles 一个。若一个源语言题名包含副标题，最好将副标题部分填写在 subtitle 中。	
18		subtitle	副标题，不得超过 900 个字符。	可选，每个 title 或 original_language_title 一个。如果可能，最好将题名和副标题分别录入不同的元素中。如果做不到，将二者都包含在题名元素中，并用标点符号（如冒号）分割也可以。当副标题紧随主标题时，不应输入主、副标题文字之间的空格和标点符号。	正确的、最佳的方式： <title>The Human Brain</title><subtitle>A Handbook</subtitle> 正确但非最佳方式： <title>The Human Brain: A Handbook</title> 不正确的方式 <title>The Human Brain: </title><subtitle>A Handbook</subtitle>
19		edition_number	版本号，不得超过 15 个字符	可选，每个 book_metadata 一个。	<edition_number>2.0</edition_number>
20		isbn	ISBN 号，最少 10 个字符，最多 17 个字符，符合标准的 ISBN 格式	必须，最多可有 6 个 ISBN 号。	<isbn media_type="print">978-7-04-017267-6</isbn>
21		item_number	出版商对注册内容的内部唯一标识，不得超过 32 个字符。	可选，每个 book_metadata 一个。	<item_number>dzmmx9787040172676</item_number>

序号	分类	数据元素	描述	格式说明	例子
22		publication_date	出版日期。	必须，最多可有10组数据。不同时间出版的各种版本的联机出版物和印刷品都允许有多个日期。属性**media_type**可选，包含：**•"print"**－印刷版出版日期；**•"online"**－电子版出版日期；**•"other"**－其它；缺省为**"print"**。	< publication_date media_type="print"> <year>1999</year> <month>06</month> <day>15</day> </ publication_date > 或: < publication_date media_type="print"> <year>1999</year> </ publication_date >
23		year	年，4位数字。	必须，每个publication_date一个。	
24		month	月，2位数字。	可选，每个 publication_date 一个。月份必须用数字格式表达，而不是拼写月份名称（即，填写"10"，而不是"October"或"十月"）。如果不足两位，须在前面补"0"（即，填写"05"而不是 "5"）。 如果不知道精确的出版月份而只有季节性名称时，month 中按照以下规则存放两位数字： 春夏秋冬季 21-24 一二三四季度 31-34	
25		day	日，2位数字。	可选，如果不足两位，须在前面补"0"（即，填写"08"而不是 "8"）。	
26		publisher	出版社	必须，每个 book_metadata 一个。	<publisher> <publisher_name>高等教育出版社</publisher_name> <publisher_place>北京市西城区德外大街4号</publisher_place> </publisher>
27		publisher_name	出版社名称，不得超过 255 个字符	必须，每个 publisher 一个。	
28		publisher_place	出版社地址，不得超过 255 个字符	可选，每个 publisher 一个。	
29	DOI数据	doi_data	DOI数据。	必须，每个book_metadata一个。	<doi_data> <doi>10.3868/b.isbn.978-7-04-017267-6</doi> <timestamp>20070513</timestamp> <resource><![CDATA[http://bookonline.hep.com.cn/xpe/portal/24cb6c01-11c9-1000-b3d6-85ca69804372?uuid=83f19fe5-29e0-450a-bd3a-894cde8ed32b]]></resource> </doi_data>
30		doi，不得超过256个字符，不能包含字符"#"、"?"、"&"、"<"、">"、"/"、"\"。	为内容分配的DOI号。	必须，每个doi_data中一个。	

序号	分类	数据元素	描述	格式说明	例子
31		timestamp，不得超过17位字符。	时间戳，用于区分DOI数据的版本。当DOI被一次或多次修改时，用时间戳来唯一标识批文件及DOI值。高版本（时间戳较新）的DOI数据可取代低版本（时间戳较旧）的DOI数据。	可选。不填时采用头标部分中的时间戳。一般采用数据产生的日期和时间的整型表达式。格式可由出版商确定。	<timestamp>19990628123304</timestamp>
32		resource，不得超过2048字符	与一个DOI相关的URI，一般为URL地址。	必须，每个 doi_data一个。	
33	内容元数据	content_item	电子书中某一篇、章、节的元数据	可选，每本电子书不限制个数。可以在component_type属性中定义本内容的类型，类型包括：chapter、section、part、track、reference_entry、other，并且可以在属性language中定义基于ISO-639标准的的语言类型	<content_item component_type="chapter" language="zh"> <contributors> …</contributors> <titles>… </titles> <doi_data> <doi>10.3868/b.isbn.978-7-04-017267-6.c03</doi> <resource><![CDATA[http://bookonline.hep.com.cn/xpe/portal/ff40f9c-11c3-1000-ad80-85ca69804372?uuid=83f19fe5-29e0-450a-bd3a-894cde8ed32b]]></resource> </doi_data> </content_item>
34		contributors	责任者信息，结构和要求同本表 12－14 行。		
35		titles	文章题名信息，结构和要求同本表15－18行。	可选，每个content_item一个。	
36		component_number	部件号，最多不超过15个字符	可选，每个content_item一个，可以对一本电子书中的内容部件进行编号。	<component_number>13</component_number>
37		publication_date	出版日期，结构和要求同本表22－25行。		
38		item_number	出版商对注册内容的内部唯一标识，结构和要求同本表21行。		
39		pages	页码信息。	可选，每个 content_item 一个。当一篇文章有不连续的页码范围时，应该将第一个页码范围填入 first_page 和 last_page 中，并将其它附加的页码信息填入 other_pages 中。	<pages> <first_page>33</first_page> <last_page>54</last_page> <other_pages>29-35,41-45,49</other_pages> </pages>

序号	分类	数据元素	描述	格式说明	例子
40		first_page	首页页码，不得超过 15 个字符。	必须，每个 pages 一个。允许大写或小写的罗马数字。可以是字母、汉字、数字或三者的混合，但不能包含标点符号。	<first_page>33</first_page>
41		last_page	末页页码，不得超过 15 个字符。	可选，每个 pages 一个。允许大写或小写的罗马数字。可以是字母、汉字、数字或三者的混合，但不能包含标点符号。如果文章有不连续的页码，在 last_page 中填入第一个页码范围最后一页页码，并将所有其他信息都填入 other_pages 中。	<pages> <first_page>33</first_page> <last_page>54</last_page> </pages>
42		doi_data	内容的DOI数据，结构和要求同本表19－32行。	必须，每个 content_item 一个。	

*注意：关于字符集的说明

用于DOI注册的XML文件中，所有非ASCII字符必须使用Unicode编码，可以采用16进制或10进制字符引用（character references）的方法，如将Müller转换为Müller或Müller等；并且XML文件必须被设置和存储为UTF-8编码格式。