

中文 DOI 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1.0版。	<doi_batch version = "1.0.0">
2	头标区	head	标识头标区的开始/结束。	必须。	<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。	<doi_batch_id>123456</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。	<timestamp>19990628123304</timestamp>
5		depositor	存储者, 发送批处理文件的组织的相关信息。	必须。	<depositor><name>Name of the Depositor</name>
6		name	进行数据存储的组织名称。	必须。	<email_address>xyz@authorized_depositor.com
7		email_address	组织中接收有关反馈信息的电子信箱地址。	必须。	</email_address></depositor>
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。	<registrant>Name of Registrant</registrant>
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, 如一个或多个期刊的DOI注册元数据。	</head>
10	期刊 DOI 注册数据	journal	标识一个期刊的全部 DOI 注册数据的开始/结束。	必须。包含一个期刊的全部 DOI 注册数据。	<journal> <journal_metadata>...</journal_metadata> <journal_issue>...</journal_issue> <journal_article publication_type="full_text">...</journal_article> </journal>
11	刊元数据	journal_metadata	期刊的元数据。	必须, 每个期刊只能包含一组期刊元数据。	<journal_metadata> <journal_id>bjdxxb_zr</journal_id> <full_title>北京大学学报自然科学版</full_title> <full_title>ACTA SCIENTIARUM UNIVERSITATIS OEKINENSIS</full_title> <abbrev_title/> <issn media_type="print">0479-8023</issn> <cn media_type="print">11-2442/N</cn> </journal_metadata>

序号	分类	数据元素	描述	格式说明	例子
12		journal_id	由出版商提供的期刊唯一标识, 该标识相同的期刊作为同一种刊处理。	必须, 每个期刊一个。	
13		full_title	期刊全称, 不得超过256个字符。	必须, 每个期刊最多可以有10个。	
14		abbrev_title	期刊简称, 不得超过150个字符。	可选。每个期刊最多可以有10个。	
15		issn	期刊的ISSN号, 必须为8位数字, 或由连字号“-”分割的两组四位数字(末位可以是'X')。该元素中不能包含字符"ISSN"。	可选, 如果期刊有ISSN号则必须填写。每个期刊最多可有6个ISSN号。每个ISSN号必须填写属性media_type, 值为: “ print ”—印刷品ISSN号, “ electronic ” —电子版的ISSN号, 缺省值为“ print ”。	
16		cn	期刊的CN号。	可选, 如果期刊有CN号则必须填写。每个期刊最多可有6个CN号。每个CN号必须填写属性media_type, 值为: “ print ” —印刷品CN号, “ electronic ”—电子版的CN号, 缺省值为“ print ”。	
17	DOI数据	doi_data	DOI数据。	期刊文章的DOI数据是必须的, 刊、卷、期的DOI数据可选。	<doi_data> <doi>10.3321/j.issn:0479-8023.1999.06.bjdxxb990607</doi> <timestamp>20070513</timestamp> <resource><![CDATA[http://www.wanfangdata.com.cn/Search/PeriodicalArticle.aspx?qcode=bjdxxb199906007]]></resource> </doi_data>
18		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。	
19		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>

序号	分类	数据元素	描述	格式说明	例子
20		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。	
21	期元数据	journal_issue	(期刊中)某一期的元数据。	必须, 每个 journal 一个或多个。	<pre><journal_issue> <publication_date media_type="print"> <year>1999</year> </publication_date> <journal_volume> <volume>74</volume> </journal_volume> <issue>5</issue> <special_numbering>Suppl 1</special_numbering> <doi_data> <doi>10.3321/j.issn:0479-8023.1999.06.bjdxxb990607</doi> <timestamp>20070513</timestamp> <resource><![CDATA[http://www.wanfangdata.com.cn/Search/PeriodicalArticle.aspx?qcode=bjdxxb199906007]]></resource> </doi_data> </journal_issue></pre>
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”。	<pre><date date_type="print"> <year>1999</year> <month>06</month> <day>15</day> </date></pre>
23		year	年, 4位数字。	必须, 每个 publication_date 一个。	或: <pre><date date_type="print"> <year>1999</year> </date></pre>
24		month	月, 2位数字。	<p>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”)。</p> <p>当一期期刊同时有期号和季节性名称时, 则 issue 中只列出期号即可。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:</p> <p>春夏秋冬季 21-24 一二三四季度 31-34</p> <p>出现一期覆盖多个月份的情况时, 如“三月-四月”等, 只填写月份范围中的第一个月的数字值。</p>	

序号	分类	数据元素	描述	格式说明	例子
25		day	日, 2位数字。	如果不足两位, 须在前面补“0”(即, 填写"08"而不是"8")。	
26		journal_volume	期刊的卷。	可选, 每个 journal_issue 一个。通过在 journal_volume 中包含 doi_data, 可为整个一卷期刊注册一个 DOI 号。	<pre><journal_volume> <volume>74</volume> <doi_data> <doi>10.3321/j.issn:0479-8023.1999.06.bjdxxb990607</doi> <timestamp>20070513</timestamp> <resource><![CDATA[http://www.wanfangdata.com.cn/Search/PeriodicalArticle.aspx?qcode=bjdxxb199906007]]></resource> </doi_data> </journal_volume></pre>
27		volume	卷号, 不得超过15个字符。	可选, 每个 journal_volume 一个。不能包含“Volume”或“Vol.”、“卷”字符。可以是[a-z][A-Z][0-9]或罗马数字的组合。	
28		doi_data	卷的DOI数据, 结构和要求同本表17—20行。	可选, 每个 journal_volume 一个。	
29		issue	期号, 不得超过15个字符。	必须, 每个 journal_issue 一个。不能包含"issue", "No"或 "number"、“第”、“期”等字符。可以是[a-z][A-Z][0-9]的组合。	<pre>volume 74, issue 3 Spring 1999 <journal_issue> <publication_date media_type="print"> <month>2</month><year>1999</year> </publication_date> <journal_volume><volume>74</volume></journal_volume> <issue>3</issue> </journal_issue></pre>
30		special_numbering	特刊号或增刊号, 不得超过15个字符。	可选, 每个 journal_issue 一个。定义特刊类型的文字(如"suppl")必须与号码一起填写到该元素中。	<pre><special_numbering>Suppl 1</special_numbering></pre>
31	期刊文章元数据	journal_article	期刊文章元数据。	可选, 每个 journal 多个。	<pre><journal_article > <titles>...</titles> <contributors>...</contributors> <publication_date>...</publication_date> <pages>...</pages> <publisher_item>...</publisher_item> <doi_data>...</doi_data> </journal_article></pre>
32		titles	文章题名信息。	必须, 每个 journal_article 最多可有 20 个。	<pre><titles></pre>
33		title	文章题名, 不得超过256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。	<pre><title>构树去木质部后TIBA对组织再生的影响及其间过氧化物酶和酯酶同工酶的变化</title> <subtitle> <title>Effect of TIBA on the Regeneration of the Tissues and Changes of Peroxidase and Esterase Isozyme after Removal of Xylem in Broussonetia papyrifera</title> <subtitle/> </titles></pre>

序号	分类	数据元素	描述	格式说明	例子
34		subtitle	副标题, 不得超过 256 个字符。	可选, 每个 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>
35		contributor	责任者信息。	可选, 每个可含 1-255 个人名 (person_name) 或机构名称 (organization)。	<contributor> <person_name sequence="first" contributor_role="author">徐承和</person_name> <person_name sequence="additional" contributor_role="author">龚中麟</person_name> <person_name sequence="additional" contributor_role="author">饶亲江</person_name> <organization sequence="additional" contributor_role="author">北京大学电子学系</organization> </contributor>
36		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。	
37		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。	
38		publication_date	出版日期, 结构和要求同本表 22—25 行。		
39		pages	页码信息。	可选, 每个 journal_article 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>15</first_page> <last_page>26</last_page> <other_pages>29-35,41-45,49</other_pages> </pages>
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>2268</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>15</first_page> <last_page>26</last_page> </pages>

序号	分类	数据元素	描述	格式说明	例子
42		other_pages	其它页,当文章有不连续的页码范围时,描述附加的页码信息。不得超过 100 个字符。	可选,每个 pages 一个。应该包含冒号或连字符表示离散的页码或连续的页码范围。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,但不能包含空格。	<other_pages>29-35,41-45,49</other_pages>
43		publisher_item	出版商分配的文章标识。	可选,每个 journal_article 一个。当一篇文章的 DOI 元数据不足以将其与另一篇文章进行区分时,publisher_item 中的 item_number 可以起到区分文章的作用。通过将属性值设为<item_number item_number_type="sequence_number">并且为<item_number>提供一个唯一值,系统将能够区分两篇文章。	<publisher_item> <item_number>tm3001</item_number> </publisher_item>
44		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 publisher_item 1-3 个。属性 item_number_type 标识 item_number 的类型:如果 item_number 中只包含了出版商的流水号,则没必要提供该属性。如果 item_number 中还包含了其他数据,则该属性可被用于定义其内容。例如,如果期刊是在线出版的(即没有页码),并且目录中的每篇文章被标以流水号,则该文章编号可被放到 item_number 中,并且 item_number_type 属性可被设为"article_number"。	<item_number>tm3001</item_number> 或 <item_number item_number_type="article_number">1</item_number>
45		abstract	文章的摘要,字符	可选,每个 journal_article 最多 2 个	<abstract>地震发生在活动断层上,而断层活动不一定都发震,为进行地震危险性分析, ... </abstract> <abstract> Tectonic earthquakes are closely related to active faults.A large active fault,...</abstract>
46		keywords	文章的关键词,字符	可选,每个 journal_article 最多 2 个	<keywords>活动断层;地震;地震破裂段</keywords> <keywords>active faults; earthquake; earthquake rupture segment</keywords>
47		doi_data	文章的DOI数据,结构和要求同本表17-20行。	必须,每个 journal_article 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。