

中文 DOI 科学数据注册 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2.1.0版。	<pre> <doi_batchversion="2.1.0"> </doi_batchversion> <head> <doi_batch_id>123456</doi_batch_id><timestamp>19990628123304</timestamp> < depositor><name>NameoftheDepositor</name> <email_address>xyz@authorized_depositor.com</email_address></ depositor> <registrant>NameofRegistrant</registrant> </head> </pre>
2	头标区	head	标识头标区的开始/结束。	必须。	
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	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。不得超过17个字符。	
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	注册者，拥有注册内容的组织的信息。	必须。不得超过130个字符。	
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,如一个或多个科学数据的DOI注册元数据。	
10	科学数据注册数据	science_data	标识全部科学数据DOI注册数据的开始/结束。	必须。包含一个（条）科学数据的全部DOI注册数据。	<pre> <science_data> <database>...</database> <datasetdataset_type="record">...</dataset> </science_data> </pre>
11	数据库注册数据	database	标识一个数据库(产品)的全部DOI注册数据的开始或结束	必须，一个数据库（数据产品）包含一个或者多个数据集（记录）	<pre> <database> <contributors>...</contributors> <titles><title>黑河综合遥感联合试验:冰沟飞行区机载微波辐射计K%20Ka波段数据集(2008年3月29日)</title></titles> <description>本数据集……能够直接使用的产品</description> <publisher><publisher_name>寒区旱区科学数据中心</publisher_name></publisher> <doi_data><doi>10.3972/water973.0237.db</doi><timestamp>200963101000</timestamp><resource><![CDATA[http://westdc.westgis.ac.cn/data/726fe99c-4423-4b73-94c4-8ed44990a6d0.xml]]></resource></doi_data></database> </pre>

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，每个可包含1-255个人名或者机构名称，必须包含至少一个人名或者机构名称。	<pre><contributors> <person_namessequence="first"contributor_role="author"> 毛明 </person_name> <person_namessequence="additional"contributor_role="editor"> 关旭 </person_name> </contributors></pre>
13		person_name	人名	可选。 每个长度不得超过 450 字符。 包含两个属性。 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	可选。 每个长度不得超过 450 字符。 含两个属性。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	数据库的名称信息	必须，每个数据库最多可以有20个titles。每个titles填写可选属性 language，表示语种，值按照按照 GB/T4880.1 要求填写。常用语种：中文即 language="zh"，英文即 language="en"。	<pre><titles language="zh"> <title>黑河综合遥感联合试验：临泽飞行区L&K波段机载微波辐射计数据集（2008年7月8日）</title> </titles></pre>
16		title	数据库名称	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。不得超过900个字符。	
17		subtitle	副标题	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。不得超过900个字符。	正确的方式： <pre><title>TheHumanBrain</title><subtitle>AHandData</subtitle></pre> 正确但非最佳方式： <pre><title>TheHumanBrain:AHandData</title></pre> 不正确的方式 <pre><title>TheHumanBrain:</title><subtitle>AHandData</subtitle></pre>
18		original_language_title	源语言标题	可选，每个 titles 一个。如果数据库的是从一个原作翻译而成的，则填写原语言的题名，并应填写属性 language，表示语种，值按照按照 GB/T4880.1 要求填写。常用语种：中文即 language="zh"，英文即 language="en"。	

序号	分类	数据元素	描述	格式说明	例子
19		description	数据库的描述信息	可选，每个description填写可选属性language，表示语种，值按照GB/T4880.1要求填写。常用语种：中文即language="zh"，英文即language="en"。	<description language="zh">本数据产品是在《中国生态系统定位观测与研究数据集》丛书《森林生态系统》卷《海南尖峰岭站（生物物种数据集）》分册的基础上进行数字化加工，通过完善元数据并与实体数据相关联所生产的数据产品。本数据集收集了自1958年中国林业科学研究院在尖峰岭地区开始的植物资源调查以来的积累资料，同时在1995年出版的《海南岛尖峰岭地区生物物种名录》的基础上，进行重新编辑、补充和校对，包括维管束植物名录、微生物和大型真菌名录、脊椎动物名录、昆虫名录、蜘蛛纲名录等。本数据产品还包含海南尖峰岭站介绍。</description>
20		publisher	数据库的发布组织名称	必须，每个database2个，每个publisher填写可选属性language，表示语种，值按照按照GB/T4880.1要求填写。常用语种：中文即language="zh"，英文即language="en"。	<publisher language="zh"> <publisher_name>寒区旱区科学数据中心</publisher_name> <publisher_place>甘肃省兰州市东岗西路320号</publisher_place> </publisher>
21		publisher_name	发布组织名称	必须，每个 publisher 一个。不得超过 255 个字符。	
22		publisher_place	发布组织地址	可选，每个 publisher 一个。不得超过 255 个字符。	
23	DOI数据	doi_data	DOI数据。	必须。	<doi_data> <doi>10.3972/water973.0237.db</doi><timestamp>200963101000</timestamp><resource><![CDATA[http://westdc.westgis.ac.cn/water/726fe99c-4423-4b73-94c4-8ed44990a6d0]]></resource></doi_data>
24		doi	为内容分配的DOI号。	必须，每个doi_data中一个。 后缀不得超过256个字符，不能包含字符“#”、“&”、“<”、“>”、“?”、“\”、“ ”、“+”、“’”、“空格”、“;”、“%”、“@”。 在DOI编码中应该只使用ASCII数字与基本拉丁字母表（41-78Hex），可以使用的标点符号包括连字符（“-”即2DHex）、句点（“.”即2EHex）以及下划线（“_”即5FHex）。请不要在DOI编码中使用其它ASCII或者多字节编码的字符（如中文等）。	
25		timestamp	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。不得超过17位字符。	<timestamp>19990628123304</timestamp>

序号	分类	数据元素	描述	格式说明	例子
26		resource	与一个DOI相关的URI, 一般为URL地址。	必须, 每个doi_data一个, 不得超过2048字符。	
27	数据集注册数据	dataset	标识一个数据集(记录)的全部DOI注册数据的开始或结束	必须。	<pre> <datasetdataset_type="record"> <contributors>...</contributors> <titles language="zh"><title>数据集-标题1</title></titles> <dataset_date> <creation_date><year>2001</year></creation_date> <publication_datemediatype="online"><year>2002</year></pub lication_date> <update_date><year>2003</year></update_date> </dataset_date> <item_number>science0001</item_number> <description>数据集-描述1</description> <formatMIME_type="image">text</format> <doi_data><doi>10.3779/water973.0237.ds1</doi> <timestamp>200963101000</timestamp> <resource><![CDATA[http://westdc.westgis.ac.cn/water/ds1]]></ resource></doi_data> </dataset> </pre>
28		contributors	责任者信息。	结构和要求同本表12—14行。	
29		titles	数据集的名称信息	1个dataset最多可以有6个titles, 结构和要求同本表15—18行。	
30		creation_date	创建日期。	必须。	<pre> <creation_date> <year>1999</year> <month>06</month> <day>15</day> </creation_date> 或: <creation_date><year>1999</year></publication_date> </pre>
31		year	年	必须, 每个creation_date一个。4位数字。	
32		month	月	可选, 每个 creation_date 一个。2 位数字。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")。	
33		day	日	可选, 2位数字。如果不足2位, 须在前面补“0”(即, 填写"08"而不是"8")。	
34		publication_date	出版日期	可选, 结构和要求同本表29—32行。	
35		update_date	更新日期	可选, 结构和要求同本表29—32行。	
36		item_number	发布组织对注册内容的内部唯一标识	可选, 每个 dataset 一个。不得超过 32 个字符。	<item_number>science0001</item_number>
37		description	数据集的描述信息	可选, 结构和要求同本表 18 行。	

序号	分类	数据元素	描述	格式说明	例子
38		format	格式，描述数据集的格式。	必须。	<formatMIME_type="image">text</format>
39		doi_data	数据集的DOI数据	结构和要求同本表22－25行。	

***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(characterreferences)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。