

一、核心数据结构框架

文件类型：UTF-8 编码 XML 文件根节点：doi_batch（属性 version="2.0.0"）层级关系：
doi_batch → head(头标区) → body → dissertation(学位论文) → dissertation_metadata
(论文元数据) + dissertation_article(单篇论文内容)

二、头标区（必填项）

序号	分类	数据元素	描述	格式说明（依据文档 1）	示例（学位论文场景）
3	头标区	doi_batch_id	批处理唯一标识（机构自定义）	字符不限，需保证全局唯一（如：学位授予年份 + 流水号）	<doi_batch_id>PKU2025D001</doi_batch_id>
4		timestamp	数据生成时间戳（区分版本）	17 位整型（YYYYMMDDHHMMSSmmm），新时间戳覆盖旧数据	<timestamp>20251022143000000</timestamp>
6		depositor/name	提交机构名称（如高校图书馆）	必须，≤130 字符（中文 / 英文）	<name>北京大学图书馆</name>
7		depositor/email_address	反馈邮箱	必须，≤200 字符（如：lib@pk.edu.cn）	<email_address>lib@pk.edu.cn</email_address>
8		registrant	学位授予单位（论文版权所属）	必须，≤255 字符（如：北京大学）	<registrant>北京大学</registrant>

三、学位论文元数据（dissertation_metadata，必填项）

序号	数据元素	描述	格式说明（依据文档 1）	示例（博士论文）
12	journal_id	学位论文唯一标识（机构自定义）	必须，≤400 字符（建议：学科代码 + 导师缩写 + 年份）	<journal_id>PKU_D_2025_001</journal_id>

序号	数据元素	描述	格式说明 (依据文档 1)	示例（博士论文）
13	full_title	授予单位全称 (多语种)	必须，≤256 字符，需标注 language 属性 (zh/en)	<full_title language="zh">北京大学</full_title>
25	publication_date	论文提交 / 答辩日期 (印刷版)	必须，media_type="print"，包含 year (4 位)	<publication_date media_type="print"><year>2025</year></publication_date>

四、单篇学位论文内容（dissertation_article，必填项）

4.1 基础信息

序号	数据元素	描述	格式说明（依据文档 1）	示例
35	titles	论文标题 (多语种)	必须，≤20 组，标注 language（如：zh/en）	<titles language="zh"><title>AI 在音乐创作中的应用研究</title></titles>
38	contributors	作者及单位 (第一作者 + 机构)	必须，person_name 需标注 sequence="first"	<person_name sequence="first">张三</person_name><organization>北京大学艺术学院</organization>
20	subject	学科专业 (如：音乐与舞蹈学)	必须，≤450 字符（中文）	<subject>音乐与舞蹈学</subject>
21	degree	授予学位 (博士 / 硕士 / 学士)	必须，枚举值（文档 1 未明确，沿用行业规范）	<degree>博士</degree>

4.2 标识符与链接

序号	数据元素	描述	格式说明 (依据文档 1)	示例
18	doi	学位论文 DOI 号	必须，10. 开头，≤256 字符，不含汉字及特殊符号	<doi>10. 1234/PKU. D. 2025. 001</doi>

序号	数据元素	描述	格式说明 (依据文档 1)	示例
20	resource	论文全文链接 (如: 机构知识库)	必须, ≤2048 字符, XML 中用 <![CDATA[]]>包裹	<resource><![CDATA[http://kns.pku.edu.cn/handle/123456789/001]]>

4.3 补充信息 (可选)

序号	数据元素	描述	格式说明 (依据文档 1)	示例
48	abstract	论文摘要 (多语种)	可选, ≤2 组, 标注 language, ≤5000 字符	<abstract language="en">This study explores...</abstract>
49	keywords	关键词 (多语种)	可选, ≤2 组, 标注 language, 关键词用分号分隔	<keywords language="zh">人工智能; 音乐生成</keywords>
42	pages	页码范围 (如: 1-150)	可选, first_page 必填, last_page 可选	<pages><first_page>1</first_page> <last_page>150</last_page></pages>

五、五、XML 模板示例 (简化版)

```
<?xml version="1.0" encoding="UTF-8"?>
<doi_batch version="2.0.0">
  <head>
    <doi_batch_id>PKU2025D001</doi_batch_id>
    <timestamp>20251022143000000</timestamp>
    <depositor>
      <name>北京大学图书馆</name>
      <email_address>lib@pk.edu.cn</email_address>
    </depositor>
    <registrant>北京大学</registrant>
  </head>
  <body>
    <dissertation>
      <dissertation_metadata>
        <journal_id>PKU_D_2025_001</journal_id>
        <full_title language="zh">北京大学</full_title>
        <publication_date media_type="print"><year>2025</year></publication_date>
```

```
</dissertation_metadata>
<dissertation_article>
  <titles language="zh"><title>AI 在音乐创作中的应用研究</title></titles>
  <contributors>
    <person_name sequence="first">张三</person_name>
    <organization>北京大学艺术学院</organization>
  </contributors>
  <subject>音乐与舞蹈学</subject>
  <degree>博士</degree>
  <doi_data>
    <doi>10.1234/PKU.D.2025.001</doi>
    <resource><![CDATA[http://kns.pku.edu.cn/handle/123456789/001]]></resource>
  </doi_data>
</dissertation_article>
</dissertation>
</body>
</doi_batch>
```