通辽市图书馆知识资源细颗粒度建设和标签标引技术参数

一、建设内容

（一）知识资源细颗粒度建设和标签标引

根据项目要求，计划对图书、期刊、报纸、古籍、音视频等多种类型数字资源进行细颗粒度建设和标签标引，共计完成10000条细颗粒度建设和标签标引工作。

(二)知识资源细颗粒度平台

本项目系统平台利用语义网、知识图谱、大数据、智能计算等技术辅助，主要通过人工编辑加工实现知识资源细颗粒度的各项标引。系统平台包含图书管理、分类管理、用户管理、工作量管理和前端页面五个部分。图书管理系统包括结构单元管理和图表管理两部分，根据需要对图书封面、序、目录、篇章、段落和图表等结构单元进行知识组织，通过人工对相关对象数据进行编辑、标引，并进行本地存档。分类管理包括人物、机构、事件、地理和专题等子系统，通过各个子系统实现各类知识单元的人工标引。用户管理包括前台用户管理和后台用户管理功能。工作量管理用来对标引工作量进行统计。前端页面提供项目可视化呈现和知识图谱功能。

需在通辽市图书馆本地安装系统平台，实现统一检索。需协助通辽市图书馆做好验收工作，并确保符合省级和国家级的相关验收要求，通过终检。

二、文献数字资源的精细化标引

（一）加工原则

1.采用自动化抽取的方式开展精细化标引工作，加强自动化 抽取规范和方法的研究应用。

2.综合分析加工对象的文献形态、内容结构和服务需求，确 定知识资源加工粒度和著录标引对象。

3.根据文献实际情况，科学合理确定著录与标引内容，参照 文献著录规则开展著录与标引。文献所包含的各类插图和表格一 般应作为图表进行著录。

4.现代期刊和报纸类资源已建有商业数据库，建议不再重复建设。

（二）标引单位

数字资源精细化标引一般以文献组成要素单元为加工单位。对图书、期刊等类型数字资源，封面、前言、目录、正文篇章等 每个析出部分作为著录单元；对古籍数字资源，书衣、封面（内 封）、牌记、序、凡例、目录、正文卷目、插图、跋、签条、夹纸、校勘、附录、封底等每个析出部分作为著录单元；对报纸数 字资源，正文篇章、广告等每个析出部分作为著录单元。

知识抽取数据是在本批精细化标引的基础文献范畴内，以文 献中的人物、机构、地理名称、事件或其他具有标目意义的关键 词为标引单元。每个从文献中抽取出来的知识条目生成一条知识 抽取数据。

(三) 建设内容

1.基础资源著录

对图书、期刊、报纸、古籍等类型的精细化标引数据的基础 文献进行元数据著录，一般以文献“种”作为著录粒度。其中， 记录标识号为必备字段，记录标识号编制方法见附件，其他著录字段和要求参照数字图书馆推广工程联合建设项目相关标准规范。

2.细粒度文献著录

对基础文献析出的各个组成要素进行元数据著录，一般以篇章（包括封面、 目录等）、片段作为著录粒度。

图书文献组成要素一般包括：封面（封一、封二、书脊）、题词页、书名页、 出版说明、版权页、序、前言、凡例、 目次、 正文各篇章、图表、参考文献、附录、索引、插页、后记（跋）、 封底（封三、封四）等。各部分定义参照新闻出版行业标准《学术出版规范 图书版式》(CY/T 120-2015)。参照国家标准《期刊编排格式》(GB/T 3179-2009)。 报纸一般以单篇文章（单个广告）作为著录单元。古籍文献参考图书文献以及其他相关文献加工规则确定著录单元。

以下涉及的各类记录标识号编制方法参见附件。

(1) 图书细粒度著录

表 1 图书细粒度著录内容

著录内容	XML 标签	属性	说明
记录标识号	identifier	必备，不可重复	指细粒度加工数据的记录标识号，是数据的唯一标识符，具体见附件。
基础文献记录标识号	sourceID	必备，不可重复	指析出著录对象的基础文献的记录标识号，字段值取自基础文献元数据 identifier 字段，具体见附件。
起始页文件名	startFileName	必备，不可重复	对象数据文件名
结束页文件名	endFileName	必备，不可重复	对象数据文件名。如果著录对象只有 1 页，则结束文件名同起始文件名。
对象文件路径	filePath	必备，不可重复	对象数据文件存储的相对路径
结构类型	type	必备，不可重复	著录对象的结构类型，如：封面、书名页、版权页、凡例、 目次、正文篇章等
语种	language	必备，可重复	著录对象的文字语种
正题名	title	必备，可重复	著录对象的主要题名，原书该部分无
			标题则著录时可自拟标题
并列正题名	parallelTitleProper	有则必备，可重复	正题名的另外一种语言和/或文字的题名
其他题名	otherVariantTitle	有则必备，可重复	从属于正题名或并列题名的副题名或其他题名说明文字
责任者	contributor	有则必备，可重复	对著录对象负有主要责任的责任者名称
责任方式	role	有则必备，可重复	责任者的责任方式，如著
创作时间	originDate	有则必备，可重复	著录对象显示的文献撰写时间
创作地点	originPlace	有则必备，可重复	著录对象显示的文献撰写地点
内容	content	有则必备，可重复	著录对象的全文文本 内容为非结构式的，全部文字录入同一字段。内容为结构式时，则录入子章节的内容，重复本字段。

附注	notes	有则必备，可重复	著录对象位于文末或脚注信息，如摘自或引自或原载于
页数	extent	必备，不可重复	著录对象总页数
页码	pageNumber	有则必备，可重复	著录对象所在页的起止页码或首页码 仅对正文篇章著录
摘要	abstract	必备，可重复	仅对正文篇章著录，著录内容为篇章提要或文摘
分类号	classification	必备，可重复	《中国图书馆分类法》分类号。仅对正文篇章著录
关键词	keyword	必备，可重复	仅对正文篇章著录，著录内容为反映著录对象内容、主题或时空范围的词语
人物名称	personalName	有则必备，可重复	著录对象所含的人物名称。属于本项目知识抽取数据的人物，在人物名称后将人物数据的记录标识号著录在[ ]中
机构名称	organizationName	有则必备，可重复	著录对象所含的机构名称。属于本项目知识抽取数据的机构，在机构名称后将机构数据的记录标识号著录在[ ]中
地理名称	geographicalName	有则必备，可重复	著录对象所含的地理名称。属于本项目知识抽取数据的地理名称，在地理名称后将地理数据的记录标识号著录在[ ]中
事件名称	eventName	有则必备，可重复	著录对象所含的事件名称。属于本项目知识抽取数据的事件，在事件名称后将事件数据的记录标识号著录在[ ]中
图表记录标识号	tableID	有则必备，可重复	指著录对象所含图表的记录标识号，字段值取自图表元数据 identifier 字段，具体见附件。

图表数量	tableNumber	有则必备，不可重复	本加工项目中著录的图表数量

(2) 期刊细粒度著录

表 2 期刊细粒度著录内容

著录内容	XML 标签	属性	说明
记录标识号	identifier	必备，不可重复	指细粒度加工数据的记录标识号，是数据的唯一标识符，具体见附件。
基础文献记录标识号	sourceID	必备，不可重复	指析出著录对象的基础文献的记录标识号，字段值取自基础文献元数据 identifier 字段，具体见附件。
年卷期	volume	必备，不可重复	著录对象基础文献的出版年和卷期号， 出版年用四位数字表示，卷期号用两位数字表示，例如：1938 年
			第 02 期
起始页文件名	startFileName	必备，不可重复	对象数据文件名
结束页文件名	endFileName	必备，不可重复	对象数据文件名。如果著录对象只有 1 页，则结束文件名同起始文件名。
对象文件路径	filePath	必备，不可重复	对象数据文件存储的相对路径
结构类型	type	必备，不可重复	如：封面、书名页、版权页、凡例、目次、正文篇章等
语种	language	必备，可重复	著录对象文字语种
正题名	title	必备，可重复	著录对象的主要题名，原刊该部分无标题则著录时可自拟标题
并列正题名	parallelTitleProper	有则必备，可重复	正题名的另外一种语言和/或文字的题名
其他题名	otherVariantTitle	有则必备，可重复	从属于正题名或并列题名的副题名或其他题名说明文字
栏目名称	column	有则必备，不可重复	著录对象从属的栏目

责任者	contributor	有则必备，可重复	对著录对象负有主要责任的责任者名称
责任方式	role	有则必备，可重复	责任者的责任方式，如著
责任者单位	institution	有则必备，可重复	责任者工作单位
责任者简介	contributorDescription	有则必备，可重复	责任者简要介绍
创作时间	originDate	有则必备，可重复	著录对象显示的文献撰写时间或投稿时间
创作地点	originPlace	有则必备，可重复	著录对象显示的文献撰写地点
内容	content	有则必备，可重复	著录对象的全文文本 内容为非结构式的，全部文字录入同一字段。 内容为结构式时，则录入子章节的内容，重复本字段。
附注	notes	有则必备，可重复	著录对象位于文末或脚注信息，如

			摘自或引自或原载于、课题信息、鸣谢等
页数	extent	必备，不可重复	著录对象总页数
页码	pageNumber	有则必备，可重复	著录对象所在页的起止页码或首页码 仅对正文篇章著录
摘要	abstract	必备，可重复	仅对正文篇章著录，著录内容为篇章提要或文摘
分类号	classification	必备，可重复	《中国图书馆分类法》分类号。仅对正文篇章著录
关键词	keyword	必备，可重复	仅对正文篇章著录，著录内容为反映著录对象内容、主题或时空范围的词语
人物名称	personalName	有则必备，可重复	著录对象所含的人物名称。属于本项目知识抽取数据的人物，在人物名称后将人物数据的记录标识号著录在[ ] 中

机构名称	organizationName	有则必备，可重复	著录对象所含的机构名称。属于本项目知识抽取数据的机构，在机构名称后将机构数据的记录标识号著录在[ ] 中
地理名称	geographicalName	有则必备，可重复	著录对象所含的地理名称。属于本项目知识抽取数据的地理名称，在地理名称后将地理数据的记录标识号著录在[ ] 中
事件名称	eventName	有则必备，可重复	著录对象所含的事件名称。属于本项目知识抽取数据的事件，在事件名称后将事件数据的记录标识号著录在[ ] 中
参考文献	reference	有则必备，可重复	正文篇章明确著录的参考文献信息
图表记录标识号	tableID	有则必备，可重复	指著录对象所含图表的记录标识
			号，字段值取自图表元数据 identifier 字段，具体见附件。
图表数量	tableNumber	有则必备，不可重复	本加工项目中著录的图表数量

(3) 报纸细粒度著录

表 3 报纸细粒度著录内容

著录内容	XML 标签	属性	说明
记录标识号	identifier	必备，不可重复	指细粒度加工数据的记录标识号，是数据的唯一标识符，具体见附件。
基础文献记录标识号	sourceID	必备，不可重复	指析出著录对象的基础文献的记录标识号，字段值取自基础文献元数据 identifier 字段，具体见附件。
出版日期	issuedDate	必备，不可重复	著录对象基础文献的出版日期，格式为 YYYY-MM-DD
起始页文件名	startFileName	必备，不可重复	对象数据文件名
结束页文件名	endFileName	必备，不可重复	对象数据文件名。如果著录对象只有 1 页，则结束文件名同起始文件名。

对象文件路径	filePath	必备，不可重复	对象数据文件存储的相对路径
结构类型	type	必备，不可重复	如：正文、广告等
语种	language	必备，可重复	著录对象文字语种
正题名	title	必备，可重复	著录对象的主要题名，原报该部分无标题则著录时可自拟标题
并列正题名	parallelTitleProper	有则必备，可重复	正题名的另外一种语言和/或文字的题名
其他题名	otherVariantTitle	有则必备，可重复	从属于正题名或并列题名的副题名或其他题名说明文字
栏目名称	column	有则必备，不可重复	著录对象从属的栏目
责任者	contributor	有则必备，可重复 刊	对著录对象负有主要责任的责任者名称，个人名称或通讯社团体名称
责任方式	role	有则必备，可重复	责任者的责任方式，如著
责任者单位	institution	有则必备，可重复	责任者工作单位
内容	content	有则必备，可重复	著录对象的全文文本 内容为非结构式的，全部文字录入同一字段。 内容为结构式时，则录入子章节的内容，重复本字段。
附注	notes	有则必备，可重复	著录对象位于文末或脚注信息，如摘自或引自或原载于等
版次	spacenum	有则必备，可重复	著录对象所在页的版次信息，包括转版。仅对正文篇章著录
摘要	abstract	必备，可重复	仅对正文篇章著录，著录内容为篇章提要或文摘
分类号	classification	必备，可重复	《中国图书馆分类法》分类号。仅对正文篇章著录
关键词	keyword	必备，可重复	仅对正文篇章著录，著录内容为反映著录对象内容、主题或时空范围的词语

人物名称	personalName	有则必备，可重复	著录对象所含的人物名称。属于本项目知识抽取数据的人物，在人物名称后将人物数据的记录标识号著录在[ ]中
机构名称	organizationName	有则必备，可重复	著录对象所含的机构名称。属于本项目知识抽取数据的机构，在机构名称后将机构数据的记录标识号著录在[ ]中
地理名称	geographicalName	有则必备，可重复	著录对象所含的地理名称。属于本项目知识抽取数据的地理名称，在地理名称后将地理数据的记录标识号著录在[ ]中

事件名称	eventName	有则必备，可重复	著录对象所含的事件名称。属于本项目知识抽取数据的事件，在事件名称后将事件数据的记录标识号著录在[ ]中
图表记录标识号	tableID	有则必备，可重复	指著录对象所含图表的记录标识号，字段值取自图表元数据 identifier 字段，具体见附件。
图表数量	tableNumber	有则必备，不可重复	本加工项目中著录的图表数量

(4) 图表细粒度著录

表 4 图表细粒度著录内容

著录内容	XML 标签	属性	说明
记录标识号	identifier	必备，不可重复	指细粒度加工数据的记录标识号，是数据的唯一标识符，具体见附件。
基础文献记录标识号	sourceID	必备，不可重复	指析出图表的基础文献的记录标识号，字段值取自基础文献元数据 identifier 字段，具体见附件。
起始页文件名	startFileName	必备，不可重复	对象数据文件名

结束页文件名	endFileName	必备，不可重复	对象数据文件名，如果图表只有 1 页，则结束文件名同起始文件名。
对象文件路径	filePath	必备，不可重复	图表对象数据文件存储的相对路径
图表类型	type	必备，不可重复	用词语标识图表类型 通用图表类型包括：插图、地图、照片、示意图、统计表、乐谱、谱系表、工程图等。 古籍图表类型包括：插图、 肖像、地图、景物图、器物图、谱系表、航海图、工程图、故事图、 山石鸟兽图、神怪图、宗教图、乐谱等。
语种	language	必备，可重复	图表文字的语种
正题名	title	必备，可重复	图表的主要标题，如无标题则自拟
并列正题名	parallelTitleProper	有则必备，可重复	图表正题名的另外一种语言和/或文字的题名
其他题名	otherVariantTitle	有则必备，可重复	从属于正题名或并列题名的副标题或其他题名说明文字
责任者	contributor	有则必备，可重复	图表的主要创建者名称
责任方式	role	有则必备，可重复	图表主要创建者的责任方式
创作时间	originDate	有则必备，可重复	图表的撰写时间
创作地点	originPlace	有则必备，可重复	图表的创作地点
内容	content	有则必备，可重复	图表的非结构化全文文本 内容为表格的，录入全部文字 内容为图片时，录入图片中有内容含义的文字
附注	notes	有则必备，可重复	位于图表文末或脚注信息，如摘自或引 自或原载于
页数	fileNumber	必备，不可重复	图表页数
分类号	classification	必备，可重复	《中国图书馆分类法》分类号
关键词	keyword	必备，可重复	仅对正文篇章著录，著录内容为反映著录对象内容、主题或时空范围的词语

人物名称	personalName	有则必备，可重复	著录对象所含的人物名称。属于本项目知识抽取数据的人物，在人物名称后将人物数据的记录标识号著录在[ ]中
机构名称	organizationName	有则必备，可重复	著录对象所含的机构名称。属于本项目知识抽取数据的机构，在机构名称后将机构数据的记录标识号著录在[ ]中
地理名称	geographicalName	有则必备，可重复	著录对象所含的地理名称。属于本项目知识抽取数据的地理名称，在地理名称后将地理数据的记录标识号著录在[ ]中
事件名称	eventName	有则必备，可重复	著录对象所含的事件名称。属于本项目知识抽取数据的事件，在事件名称后将事件数据的记录标识号著录在[ ]中

3.知识内容抽取

充分利用自动化手段分析文献内容，建立知识抽取模型，确 定知识抽取方法，从文献中抽取人物、机构、事件、地理名称以 及其他具有标目意义的专题、实物等内容，开展知识标引工作， 以形成基于文献知识内容的语料库。

从同一基础文献、不同内容位置抽取的同一个人物、机构、地理名称、事件、专题等信息，原则上应合并为一条数据。

（1） 人物标引

表 5 人物知识内容标引

著录内容	XML 标签	属性	说明
记录标识号	identifier	必备，不可重复	指知识抽取数据的记录标识号，是数据的唯一标识符，具体见附件。

基础文献记录标识号	sourceID	必备，可重复	指本条数据的信息基础，字段值取自基础文献元数据 identifier 字段，具体见附件。
人物通用名称	personalName	必备，不可重复	
人物异名	variantPersonalName	有则必备，可重复	别名、字号、笔名等。
性别	gender	有则必备，不可重复	
时代	period	有则必备，不可重复	
出生年	birthDate	有则必备，可重复	公元纪年
卒年	deathDate	有则必备，可重复	公元纪年
国别	nationality	有则必备，可重复	
籍贯	nativePlace	有则必备，可重复	
民族	ethnicGroup	有则必备，不可重复	
亲属关系类别	kinship	有则必备，可重复	
亲属关系人物	kinshipPerson	有则必备，可重复	人名。属于本项目标引条目的人物，可在人物名称后将人物数据的记录标识号著录在[ ] 中
非亲属关系类别	nonKinship	有则必备，可重复	人名。属于本项目标引条目的人物，可在人物名称后将人物数据的记录标识号著录在[ ] 中
非亲属关系人物	nonKinshipPerson	有则必备，可重复	
传略	biography	必备，不可重复	可直接摘录原文
任职机构	institution	有则必备，可重复	
职务名称	position	有则必备，可重复	
任职时间段	employTime	有则必备，可重复	
著述	writings	有则必备，可重复	著述名称、时间、 出版信息等
附注	notes	有则必备，可重复	

(2) 机构标引

表 6 机构知识内容标引

著录内容	XML 标签	属性	说明

记录标识号	identifier	必备，不可重复	指知识抽取数据的记录标识号，是数据的唯一标识符，具体见附

			件
基础文献记录标识号	sourceID	必备，可重复	指本条数据的信息基础，字段值取自基础文献元数据 identifier 字段，具体见附件
机构中文全称	chiOrganizationName	必备，不可重复	
机构英文全称	engOrganizationName	有则必备，不可重复	
机构简称	abbreviatedOrganizationName	有则必备，可重复	包括机构别称
地址	address	有则必备，可重复	
前置机构	previousOrganization	有则必备，可重复	
后置机构	nextOrganization	有则必备，可重复	
存续起始时间	startTime	有则必备，可重复	
存续结束时间	endTime	有则必备，可重复	
行业类型	type	必备，可重复	
机构描述	description	必备，可重复	可直接摘录原文
重要人物名称	personalName	有则必备，可重复	通用名称或规范名称。属于本项目知识抽取数据条目的人物，可在人物名称后将人物数据的记录标识号著录在[ ]中
重要人物事迹	personalDescription	有则必备，可重复	可直接摘录原文
重要事件	event	有则必备，可重复	可直接摘录原文。属于本项目知识抽取数据条目的事件，可在机构名称后将事件数据的记录标识号著录在[ ] 中
重要成果	achievement	有则必备，可重复	著述成果以及文艺作品、建筑作品等各类型作品

(3) 事件标引

表 7　事件知识内容标引

著录内容	XML 标签	属性	说明
记录标识号	identifier	必备，不可重复	指知识抽取数据的记录标识号，是数据的唯一标识符，具体见附件。
基础文献记录标识号	sourceID	必备，可重复	指本条数据的信息来源，字段值取自基础文献元数据 identifier 字段，具体见附件
事件中文全称	chiEventName	必备，不可重复	
事件英文全称	engEventName	有则必备，不可重复	
事件简称	abbreviatedEventName	有则必备，可重复	
事件起始时间	startTime	有则必备，可重复	
事件结束时间	endTime	有则必备，可重复	
地点	place	有则必备，可重复	
事件类型	type	必备，可重复	
事件描述	description	必备，可重复	可直接摘录原文
重要人物名称	personalName	有则必备，可重复	通用名称或规范名称。属于本项目知识抽取数据条目的人物，可在人物名称后将人物数据的记录标识号著录在[ ]中
重要人物事迹	personalDescription	有则必备，可重复	可直接摘录原文
重要成果	achievement	有则必备，可重复	产生的著述成果以及文艺作品、建筑作品等各类型作品

(4) 地理名称标引

表 8　地理名称知识内容标引

著录内容	XML 标签	属性	说明
记录标识号	identifier	必备，不可重复	指知识抽取数据的记录标识号，
			是数据的唯一标识符，具体见附件

基础文献记录标识号	sourceID	必备，可重复	指本条数据的信息来源，字段值取自基础文献元数据 identifier 字段，具体见附件
地名专名	geographicalName	必备，不可重复	
地名简称	abbreviatedGeographicalName	有则必备，可重复	
异名	variantGeographicalName	有则必备，可重复	地名别名、惯用地名、历史地名等
行政层级	administrativeLevel	必备，不可重复	省、市、县、 乡、村分别为一级至五级；古代地名根据当时区划建立行政层级对应表，并给定行政层级。
起始年代	startTime	有则必备，可重复	地名建制时间
结束年代	endTime	有则必备，可重复	地名撤销时间
沿革事件类型	evolutionEvent	有则必备，可重复	分为地名设立、改名、行政层级调整、隶属调整、地理坐标调整、注销、重设等类型。
时间	evolutionTime	有则必备，可重复	沿革事件发生的时间
说明	notes	有则必备，可重复	沿革事件说明，可直接摘录原文
规范性文件	authorityDocument	有则必备，可重复	确定沿革事件的规范性文件名称
隶属	underJurisdiction	有则必备，可重复	该地名上一级行政单位名称
辖区	jurisdiction	有则必备，可重复	该地名下一级行政单位名称
经纬度	coordinate	有则必备，可重复	
参考方位	azimuth	有则必备，可重复	

(5) 专题标引

根据某一特定专题，从挖掘知识内涵明确标引内容，开展特 色突出、 内容丰富的专题标引。

表 9 专题知识内容标引示例

著录内容	XML 标签	属性	说明

记录标识号	identifier	必备，不可重复	指知识抽取数据的记录标识号，是数据的唯一标识符，具体见附件。
基础文献记录标识号	sourceID	必备，可重复	指本条数据的信息来源，字段值取自基础文献元数据 identifier 字段，具体见附件
物产名称	productName	必备，不可重复	
物产类型	type	必备，可重复	
产地	originPlace	必备，不可重复	属于本项目标引条目的地名，可在产地名称后将地理数据的记录标识号著录在[ ] 中
物产描述	description	有则必备，可重复	可直接摘录原文
产量	yield	有则必备，可重复	可直接摘录原文

（四） 成果形式

成果文件命名规则和文件存储结构参见附件。

1.元数据

包括基础文献元数据、细粒度加工元数据、知识抽取数据，一般采用 XML 格式，遵照 XML1.0 规范，使用 UTF-8 编码方式、Unicode5.0 字符集。

2.对象数据

基础文献的全部对象数据,包括长期保存级、发布服务级等 所有加工级别的数据，例如：TIF 文件、完成数字化识别的 TXT 文件、双层 PDF 文件等。

3.证明文件

项目涉及的版权证明文件等。版权证明文件包括：说明本项 目加工文献的版权来源、授权范围、授权使用方式与对象、

使用 期限等内容的整体版权说明，各权利人或各资源的具体授权文 件。

4.数据说明文件

项目提交各类数据的总体说明文件。总体说明文件内容包括：项目名称、提交单位名称、各类型资源数量、记录标识号号 段、存储介质情况以及特殊情况说明。数据加工过程中引用的词 表、规范库等情况，也应在数据说明文件中进行说明。

三、音视频资源的精细化标引

（一） 加工原则

参照前述第二部分“文献数字资源的精细化标引”中的加工 原则。

（二） 标引单位

本指南所述音视频资源指常见的图书馆音视频资源，例如：公开课、讲座、纪录片以及微视频等。公开课著录以讲授某一专 题的课程为著录对象，每个课程为一个著录单元。讲座、纪录片、 微视频等其他类型音视频资源参考公开课著录，本指南统称为 “音视频基础资源”，统一以“部”为单位。每部音视频基础资 源可以包含一个或多个音视频，每个音视频统称为“小节”。对 音视频资源精细化标引，一般以小节、责任者作为著录单元。每 个著录单元的著录信息生成一条析出元数据。

音视频资源知识内容抽取是在精细化标引的音视频基础资 源范畴内，以资源中的人物、机构、地理名称、事件、作品、实 物或其他具有标目意义的关键词为标引单元。根据音视频资源的 内容，抽取小节音视频资源中内容完整、较好体现观点、释义、 主要内容的一段连续的音视频，由一个以上相互关联的场景构成 知识内容片段。

（三） 建设内容

1.音视频基础资源著录

表 10　音视频基础资源著录内容

著录内容	XML 标签	属性	说明
音视频基础资源记录标识号	sourceID	必备，不可重复	音视频基础资源的唯一标识号，具体见附件。
原始音视频记录标识号	oldID	有则必备，不可重复	以往项目已建成音视频资源的记录标识号。如图书馆公开课的课程标识号。
音视频名称	title	必备，可重复	资源的对外正式公开名称。
其他名称	otherVariantTitle	有则必备，可重复	除“音视频名称”外的其他名称，包括资源出现的其他语种名称、简称、别称等
责任者	contributor	有则必备，可重复	据实著录。对于讲座、公开课资源，责任者一般是主讲人、翻译等；对于纪录片资源，责任者一般包含制片人、导演、监制、策划、主持人、播音、摄像、翻译等。
责任者 ID	contributorID	有则必备，可重复	作为音视频基础资源责任者对应的唯一标识号，有则必备。责任者 ID 应唯一，同一责任者对应唯一的责任者 ID。不同任务年建设应注意查重。
分类号	classification	必备，可重复	取值自中图法。多值时使用半角分号分隔

主题	subject	必备，可重复	采用受控词表为主题取值，如中分表。多值时使用半角分号分隔。
关键词	keyword	必备，可重复	用于描述该资源的自然语言词组。多值时使用半角分号分隔。一般来说，动词、形容词、副词不建议作为关键词。
简介	abstract	必备，可重复	对完整资源的内容进行描述，使用自然语言，200 字左右。
文件格式	format	必备，可重复	音视频资源的格式，英文大写著录。
时长	duration	必备，不可重复	音视频资源的播放时长。采用“HH:MM:SS”格式著录。如 00:20:00。
声道语种	vocalLanguage	有则必备，可重复	音视频资源的声道语种。采用国际标准 ISO639-2（或其等同标准 GB/T 4880.2-2000）著录 3 位语种代码。例：chi。
字幕语种	subtitleLanguage	有则必备，可重复	音视频资源的字幕语种。采用国际标准 ISO639-2（或其等同标准 GB/T 4880.2-2000）著录 3 位语种代码。例：chi。
创建日期	created	有则必备，可重复	音视频资源的录制 日期。采用 W3C-DTF 表示年月 日 (YYYY-MM-DD)或只有年 (YYYY) 。
版权	copyright	必备，可重复	对音视频资源版权归属的详细说明。
馆藏位置	location	有则必备，可重复	标记实体馆藏排架号等情况。
发布地址	URI	有则必备，可重复	标记资源在收藏单位的在线服务地址。
来源	source	有则必备，可重复	对于数字化音视频资源，著录数字化来源的音视频资源信息。

2.细粒度资源著录

对基础音视频资源析出的各个组成要素进行元数据著录，一 般以小节、责任者作为著录粒度。

(1) 小节细粒度著录

表 11　小节细粒度著录内容

著录内容	XML 标签	属性	说明
小节记录标识号	sectionID	必备，不可重复	音视频小节唯一标识号，具体见附件。
音视频基础资源记录标识号	sourceID	必备，不可重复	指析出著录对象所在的音视频基础资源的记录标识号。
对象数据文件名	fileName	必备，不可重复	列出该小节数据对应的所有对象数据文件名。
小节标题	sectionTitle	必备，可重复	作为被描述音视频小节的标题。需要使用能够说明概括本小节内容的自然语言。不能简单标记为“音视频名称+数字序号”。
分类号	classification	必备，可重复	取值自中图法。多值时使用半角分号分隔。
主题	subject	必备，可重复	采用受控词表为主题取值，如中分表。多值时使用半角分号分隔。
关键词	keyword	必备，可重复	用于描述该小节内容的、精炼化的自然语言词组。多值时使用半角分号分隔。一般来说，动词、形容词、副词不建议作为关键词。
简介	abstract	必备，可重复	对资源小节内容进行描述，使用自然语言，200 字左右。
时长	duration	必备，不可重复	音视频小节的播放时长。采用“HH:MM:SS”格式著录。
责任者	contributor	有则必备，可重复	对小节内容负有责任的责任者。对于讲座、公开课资源，责任者一般是主讲人、翻译等；对于纪录片资源，责任者一般包含制片人、导演、监制、策划、主持人、播音、摄像、翻译等。应据实著录。
责任者 ID	contributorID	有则必备，可重复	作为该小节所关联的责任者唯一标识号。

(2) 责任者细粒度著录

表 12　音视频责任者细粒度著录

著录内容	XML 标签	属性	说明
责任者记录标识号	contributorID	必备，不可重复	作为责任者的唯一标识号，具体见附件。
音视频基础资源记录标识号	sourceID	必备，不可重复	指析出著录对象所在音视频基础资源的记录标识号。
小节记录标识号	sectionID	必备，不可重复	指析出著录对象所在小节的记录标识号。
责任者	contributor	必备，可重复	被描述责任者最为人熟知的名称形式。
其他名称	variantContributorName	有则必备，可重复	标记除“责任者”外的其他名称，如音译名称、真名（当责任者名称为为人熟知的网名、笔名等时）等。
性别	gender	必备，不可重复	被描述责任者的性别。
出生日期	birthDate	有则必备，不可重复	被描述责任者出生日期。采用 W3C-DTF 表示年月 日 (YYYY-MM-DD) 或只有年 (YYYY) 。
籍贯	nativePlace	有则必备，不可重复	被描述责任者的祖居地。
职称	professionalTitle	有则必备，可重复	被描述责任者取得的职称。
学习经历	learnExperience	有则必备，可重复	被描述责任者就读过的院校。若有多个，按时间先后顺序著录。
工作经历	workExperience	有则必备，可重复	被描述责任者的工作经历，可按时间线著录。
研究领域	researchField	有则必备，可重复	被描述责任者的主要研究方向或专业领域。
研究成果	achievement	有则必备，可重复	被描述责任者的主要研究成果。

3.知识内容抽取

音视频资源知识内容抽取包含两部分内容：一是充分利用自 动化手段分析音视频内容，建立知识抽取模型，确定知识抽取方 法，从音视频中抽取与本部音视频资源内容或主题紧密相关的人 物、机构组织、事件、地理名称、作品、实物以及其他具有标目 意义的内容，开展知识内容（关键词）标引，以形成基于知识内 容的语料库；二是根据音视频资源的内容和主题，抽取小节音视

频资源中内容完整、较好体现观点、释义、主要内容的一段连续 的音视频，形成知识内容片段。从同一部音视频资源、不同内容 位置抽取的同一个人物、机构、地理名称、作品、实物等信息， 原则上应合并为一条数据。

（1） 人物标引

标引内容参考前述第二部分 文献数字资源的精细化标引中 的“表 5 人物知识内容标引”。(“基础文献记录标识号”字段改 为“音视频基础资源记录标识号”，著录“析出著录对象所在音 视频基础资源的记录标识号”。)

（2） 机构标引

标引内容参考前述第二部分 文献数字资源的精细化标引中 的“表 6 机构知识内容标引”。(“基础文献记录标识号”字段 改为“音视频基础资源记录标识号”，著录“析出著录对象所在 音视频基础资源的记录标识号”。)

（3） 事件标引

标引内容参考前述第二部分 文献数字资源的精细化标引中 的“表 7 事件知识内容标引”。(“基础文献记录标识号”字段 改为“音视频基础资源记录标识号”，著录“析出著录对象所在 音视频基础资源的记录标识号”。)

(4) 地理名称标引

标引内容参考前述第二部分 文献数字资源的精细化标引中 的“表 8 地理名称知识内容标引”。(“基础文献记录标识号” 字段改为“音视频基础资源记录标识号”，著录“析出著录对象 所在音视频基础资源的记录标识号”。)

(5) 作品

作品类型的知识内容包含文学作品、艺术作品、音乐作品、 建筑、器物等人造事物等。

表 13 作品知识内容标引

著录内容	XML 标签	属性	说明
作品记录标识号	workID	必备，不可重复	作品的唯一标识号，具体见附件。
音视频基础资源记录标识号	sourceID	必备，不可重复	指析出著录对象所在音视频基础资源的记录标识号。
作品名称	workTitle	必备，可重复	用于作品对外正式公开名称。
其他名称	otherVariantTitle	有则必备，可重复	除“作品名称”外的其他名称，包括其他语种名称、简称、别称等。
责任者	contributor	有则必备，可重复	对作品作出贡献的责任实体。
责任者 ID	contributorID	有则必备，可重复	作品责任者对应的唯一标识号。
分类号	classification	必备，可重复	取值自中图法。多值时使用半角分号分隔。
主题	subject	必备，可重复	采用受控词表为主题分类取值，如中分表。多值时使用半角分号分隔。

关键词	keyword	必备，可重复	用于描述该作品内容的、精炼化的自然语言词组。多值时使用半角分号分隔。一般来说，动词、形容词、副词不建议作为关键词。
简介	abstract	必备，可重复	对作品内容进行描述，使用自然语言，200 字左右。
创建日期	created	有则必备，可重复	作品创建的 日期。采用 W3C-DTF。

(6) 实物

实物类型的知识内容包括自然形成的山川、河流、动物、植 物等实物。

表 14 实物知识内容标引

著录内容	XML 标签	属性	说明
实物记录标识号	objectID	必备，不可重复	作为实物知识内容的唯一标识号，具体见附件。
音视频基础资源记录标识号	sourceID	必备，不可重复	指析出著录对象所在音视频基础资源的记录标识号。
实物名称	objectName	必备，可重复	用于自然实物对外正式公开名称。
其他名称	otherVariantName	有则必备，可重复	除“实物名称”外的其他名称，包括其他语种名称、简称、别称等。
分类号	classification	必备，可重复	取值自中图法。多值时使用半角分号分隔。
主题	subject	必备，可重复	采用受控词表为主题分类取值，如中分表。多值时使用半角分号分隔。
关键词	keyword	必备，可重复	用于描述该实物内容的、精炼化的自然语言词组。多值时使用半角分号分隔。一般来说，动词、形容词、副词不建议作为关键词。
简介	abstract	必备，可重复	对实物内容进行描述，使用自然语言，200 字左右。

(7) 知识内容片段

表 15　音视频知识内容片段内容标引

著录内容	XML 标签	属性	说明
片段记录标识号	fragmentID	必备，不可重复	知识内容片段的唯一标识号，具体见附件。
小节记录标识号	sectionID	必备，不可重复	指知识片段所在小节的记录标识号。
知识片段名称	title	必备，可重复	作为被描述音视频片段的名称。需要使用能够说明概括本片段内容的自然语言。
起止时间	period	必备，不可重复	片段音视频的起讫时间。
主题	subject	必备，可重复	采用受控词表为主题取值，如中分表。多值时使用半角分号分隔。
人物名称	personalName	有则必备，可重复	著录知识内容片段中所含的人物名称。属于本项目知识抽取数据的人物，在人物名称后将人物数据的记录标识号著录在[ ] 中
机构名称	organizationName	有则必备，可重复	著录知识内容片段中所含的机构名称。属于本项目知识抽取数据的机构，在机构名称后将机构数据的记录标识号著录在[ ] 中
地理名称	geographicalName	有则必备，可重复	著录知识内容片段中所含的地理名称。属于本项目知识抽取数据的地理名称，在地理名称后将地理数据的记录标识号著录在[ ] 中
事件名称	eventName	有则必备，可重复	著录知识内容片段中所含的事件名称。属于本项目知识抽取数据的事件，在事件名称后将事件数据的记录标识号著录在[ ] 中
作品名称	workTitle	有则必备，可重复	著录知识内容片段中所含的作品名称。属于本项目知识抽取数据的作品，在作品名称后将作品数据的记录标识

			号著录在[ ] 中
实物名称	objectName	有则必备，可重复	著录知识内容片段中所含的实物名称。属于本项目知识抽取数据的实物，在实物名称后将实物数据的记录标识号著录在[ ] 中
分类号	classification	必备，可重复	取值自中图法。多值时使用半角分号分隔。
文字提取	content	必备，不可重复	提取音视频资源片段的文字。
简介	abstract	必备，可重复	对片段内容进行描述，使用自然语言，200 字左右。

（四）成果形式

成果文件命名规则和文件存储结构参见附件。

1.元数据

包括音视频基础资源元数据、细粒度加工元数据（包括小节元数据、责任者元数据）、知识抽取数据（人物、机构、事件、地理名称、作品、实物、知识内容片段），一般采用XML 格式，遵照 XML1.0 规范，使用 UTF-8 编码方式、Unicode5.0 字符集，信息尽可能完整、正确。

2.对象数据

各类型对象文件，包含：各种格式的音视频文件、责任者头像文件、字幕文件等。如已建设完成音视频资源的封面文件、背景图文件、富文本文件、可下载附件文件等应一并提供。

3.证明文件

项目涉及的版权证明文件等。版权证明文件包括：说明本项 目音视频资源的版权来源、授权范围、授权使用方式与对象、使用期限等内容的整体版权说明，各权利人或各资源的具体授权文件。

4.数据说明文件

项目提交各类数据的总体说明文件。总体说明文件内容包括：项目名称、提交单位名称、音视频基础资源记录标识号起止 号、音视频基础资源总数量、小节记录标识号起止号、小节总数 量、责任者记录标识号起止号、责任者总数量、知识内容记录标 识号起止号、知识内容总数量、总存储量信息、存储介质情况以 及特殊情况说明。

明细说明表中，应描述完整的音视频基础资源名称、音视频 基础资源标识号、小节数量、责任者、各种格式的分辨率及封装 格式信息。

数据加工过程中引用的词表、规范库等情况，也应在数据说 明文件中进行说明。

四、知识组织与专题服务

知识组织与专题服务工作的内容是对资源进行精细化揭示，实现资源的知识化、专题化服务。主要利用知识（KnowledgeGraph）技术实现。

（一）建设内容

基于知识图谱技术，以结构化的形式描述客观世界中概念、实体及其关系，从结构化、半结构化、非结构化数据中获取知识， 并基于知识推理获得更多的知识，形成一个迭代的相互增强过 程，构建能够支持知识迭代的语义网络。最终目的是构建结构化 语义知识库，以实现知识导航、语义检索、智能推荐等智慧化服务。

所建知识图谱的知识内容应综合考虑本馆馆藏建设情况、用 户需求、本区域社会经济发展需要各方面要素。推荐建设方向为：特色馆藏；历史文化；红色文化；地方特色； 图情知识等。

所建知识图谱的逻辑结构包括模式层与数据层，模式层在数 据层之上，是知识图谱的核心，模式层存储的是经过提炼的知识， 通常采用本体库来管理。数据层主要是由一系列的事实组成，而知识将以事实为单位进行存储。

所建知识图谱的类型应按照项目具体需求确定，可建设以常识性知识为主的通用知识图谱，也可建设面向特定领域的领域知识图谱，还可建设通用与领域有机结合的知识图谱。在知识表示 层面上，通用知识图谱涵盖的范围更广，对领域知识图谱有着重 要的支撑作用，可以提供领域的Schema 表示，而领域知识图谱在概念图谱的层级体系上通常更具有深度，并涵盖更细粒度的知识。

（二）成果形式

1.项目建设成果

较为完整的知识图谱、能够支撑知识图谱应用的知识库、本 地的发布服务。

2.成果内容

(1) 说明文件

包括项目建设方案、知识模型的描述文档、元数据标准（含术语表、著录规则等)。

(2) 所有版权明晰的对象数据。

(3) 所有元数据或关联数据集。

(4) 自建受控词表或概念框架的语义描述。

包括分类法、普通主题词、人名主题词、团体名称主题词、家族名称主题词、会议名称主题词、地理名称主题词、统一题名主题词等。可选用的语义表示的标准包括：SKOS (Simple Knowledge Organization System，W3C 推荐标准）、OWL (Web Ontology Language，W3C 推荐标准）。成果文件命名规则和文件 存储结构参见附件。

(三) 建设流程

知识图谱主要技术包括知识获取、知识表示、知识存储、知 识建模、知识融合、知识理解、知识运维等多个方面。下面，本方案对项目建设中较为重要的环节进行具体说明。

1.知识获取

从不同来源、不同结构（结构化、半结构化和非结构化）的信息资源中通过知识抽取技术提取出计算机可理解和计算的结构化数据，形成结构化的知识并存入到知识图谱中。是知识图谱构建的第一步。

主要包括：实体抽取、关系抽取、属性抽取、事件抽取。

实体抽取的目的是从文本中识别出命名实体，包括：人物、地点、机构、日期等命名或专有实体。实体抽取的准确性直接影 响知识获取的质量和效率，因此，实体抽取是知识图谱构建和知 识获取的基础和关键。实体识别出来的实体名可能是有歧义的， 需要对实体进行消歧。实体的消歧应充分利用图书馆专业领域的 知识组织工具，包括规范档、分类法、叙词表，建设方式包括三 种：使用已有的受控词表，如：中国图书馆分类法、汉语主题词表等；在已有的受控词表上按照需求进行扩展；完全自建。

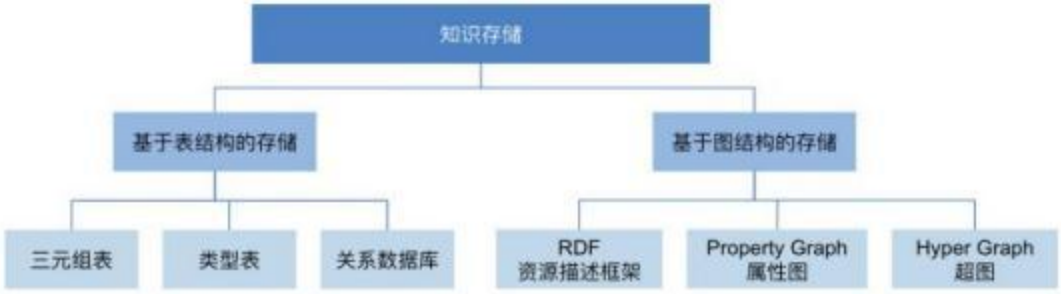
在实体抽取的基础上，从文本中自动挖掘出实体之间的关系。此外，还要对实体的自身属性信息进行采集，信息可来自于不同的信息源。

除了对实体、概念及其之间各种语义关系的提取外，还可从 自然语言中抽取得到用户感兴趣的事件信息，包括事件发生的时 间、地点、参与者、因果关系等属性信息，事件能描述粒度更大 的、动态的、结构化的知识，是现有知识资源的重要补充。

知识获取的实现主要采用规则和监督学习相结合的方法、半监督方法以及深度学习方法等，未来发展方向之一是大规模面向 开放领域的知识抽取技术。在技术的应用过程中应加强人工的干预， 以保证建成高质量的知识图谱。

2.知识存储

对各类知识的存储主要应考虑支持对大规模图数据的有效管理和计算。知识存储方式直接影响到知识图谱中知识查询、知 识计算及知识更新的效率。存储方式可选择基于表结构的存储和基于图结构的存储。



知识图谱的存储不依赖特定的底层结构，可以基于现有关系数据库或NoSQL 数据库进行构建。关系型数据库仍是市场的主流，近年来，图数据库随着机器学习、人工智能的发展而出现，与传统关系型数据库相比，能够快速解决复杂的关系问题项目建设馆可按项目需求选择合适的存储方式。

3.知识建模

知识建模的过程是知识图谱构建的基础，应是项目建设不可 或缺的一环，知识图谱应基于较为完整的知识模型实现。

知识建模应满足以下质量标准：所定义术语应给出明确的、 客观的语义定义；定义能完整表达领域内术语的含义；由术语得 出的推论与术语本身含义不会产生矛盾；具有可扩展性，支持添 加新的术语；本体约定最小，对建模对象尽可能少的约束；对用 户友好，具有易用性。

建模可选择自顶向下或自底向上的途径，也可结合使用。自 顶向下是指构建知识图谱时首先定义数据模型，可从其他高质量 的数据源中提取信息，或通过领域专家人工编制。自底向上则相 反，从实体层开始，借助于实体对齐和实体链接等技术手段，对现有实体进行归纳组织，形成底层概念，再逐步形成上层概念。

不同领域的知识具有不同的数据特点，可分别构建不同的本 体模型。所建模型主要可包括以下几个方向：

(1) 资源整合方向：揭示馆藏资源的逻辑结构、类型分布、共性特点等，实现资源有序融合。

(2) 知识组织方向：定义概念及概念间的关系，形成知识 网络。对资源中的知识进行挖掘，依托知识网络，形成能够揭示 更细粒度、更多维度、更复杂关系的知识图谱。

(3) 数字人文方向：通过文本分析、社会网络分析、知识 挖掘等技术方法，揭示资源集合中内在结构特征，发现海量数字 化对象中隐藏的知识脉络与演化规律，实现知识创新。

知识模型可基于国际上成熟通用的本体模型进行扩展，也可 按具体知识图谱的资源情况与服务需求独立构建。可复用的元数据词表或本体模型包括： 都柏林核心术语集 (DCMI Metadata Terms)，资源描述与检索 (RDA)，图书馆参考模型 (LRM)，书 目 框架 (BIBFRAME)，书 目本体 (BIBO)，资源描述框架规范 (RDFs)， 简单知识组织系统 (SKOS)， 网络本体语言 (OWL)，FOAF 本体， 事件本体(Event)、组织机构本体(The Organization Ontology)， 地名本体 (GeoNames)，地理位置 (Geo)，OWL 时间本体 (Time Ontology in OWL)，溯源本体 (PROV)，文化遗产信息交换参考 本体 (CIDOC-CRM)，欧洲数字图书馆数据模型 (EDM)，等等。

4.知识发布与服务

(1) 关联数据发布

推荐使用关联数据技术来进行知识图谱建设与服务。关联数 据提供了在Web 上发布和访问结构化数据的一种新方式，建立数据之间的链接以形成数据关系网。知识图谱与关联数据相集成，有助于形成富含语义、互联互通、计算机可理解的知识网络。

采用关联数据的形式发布和关联各种数据时应遵循以下标准：使用 URI 标识符命名任何事物 (包括信息资源和非信息资 源)，发布后能够通过 HTTP 访问，URIs 地址应稳定、持久；使 用 RDF 三元组数据模型，RDF 可以采用 RDF/XML、

N-Triples、 Turtle、RDFa、JSON-LD 等几种方式序列化；尽量多地在关联数 据中使用指向其他 URI 地址的链接，使用户可以发现更多的资源。

(2) 可视化服务

从知识图谱的特征出发，以图的形式提供知识图谱的数据服务。可视化能够将知识及知识间的关系转化为可理解的视觉表达形式，特别是在浏览、研究大规模数据时，有助于发现隐藏的特征和规律。可视化方法包括：结构图、热力图、标签云、地图、时间线、网络图等。可结合数据特点与展示需要进行选择。

(3) 智慧服务

提供基于知识图谱的信息服务，主要包括：实现语义搜索，使 Web 从网页链接向概念链接转变，支持用户按主题而不是关键 词检索，真正实现检全性与检准性；支持智能问答系统，以准确的自然语言为用户提供问题的解答；基于知识图谱的知识体系、多源异构数据的融合、用户偏好的分析等，为用户做出精准的智 能推荐，使用户获得更有深度与广度的信息资源等等。

附件：

文件存储结构与命名规则

一、文件命名规则

(一) 记录标识号

记录标识号是资源加工过程中精细化标引数据(包括基础文 献、细粒度加工数据、知识抽取数据）、知识组织与专题服务数 据、新型数字资源的唯一标识，每条数据赋予一个记录标识号。

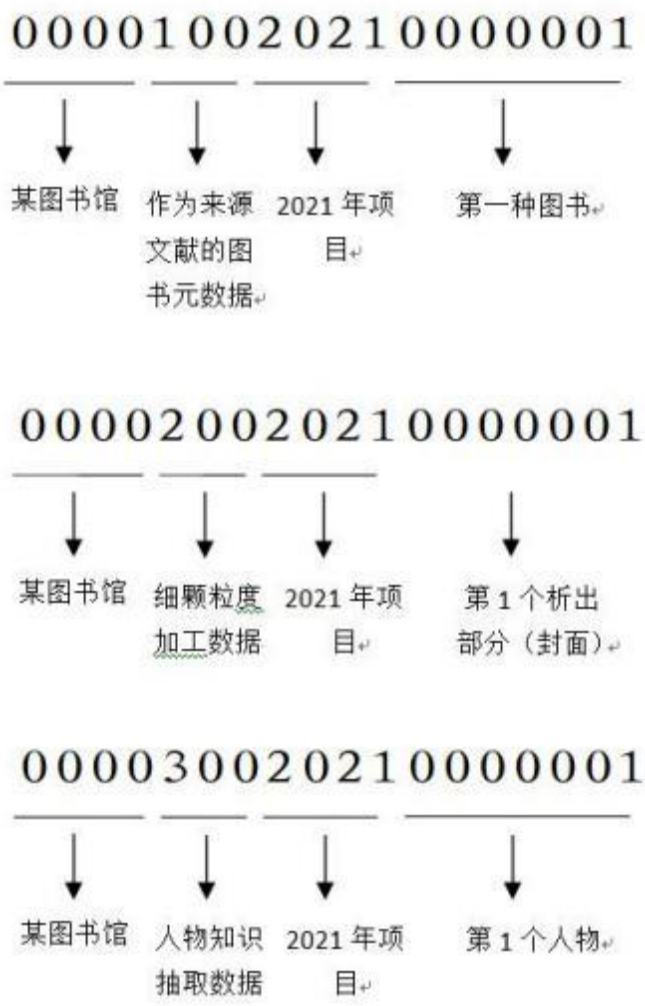
记录标识号共 18 位数字，由 4 段组成：机构代码-资源类型 -项目建设年-流水号，记录标识号各段之间不加任何连接符。其中：

- 机构代码：4 位。 同数字图书馆推广工程联合建 设项目图书馆机构代码。
- 资源类型代码：3 位。各类型资源代码见表 1。
- 项目建设年：4 位。
- 流水号： 7 位 。每条数据赋予一个流水号 ，从 0000001 起顺序排列，细粒度加工数据的流水号应按照标 引对象在基础文献中的先后顺序进行排列。

表 1 资源类型代码表

资源类型			**代码**
	基础文献	图书	100
		期刊	110
		报纸	120
		古籍	130
		音频	150
		视频	160

精细化标引数据	细粒度加工数据	析出资源（包括封面、目录、篇章以及音视频中析出的小节等，不包括图表和音视频资源的责任者）	200
		图表	210
		音视频责任者	220
	知识抽取数据	人物	300
		机构	310
		事件	320
		地理名称	330
		专题	340
		作品	350
		实物	360
		知识内容片段	370
知识组织与专题服务数据	元数据		400
	受控词表		410
新型数字资源	发布成品		500
	源文件		510



（二） 元数据文件命名

元数据文件名由 4 段组成，共 13 位数字：机构代码-资源类型-项目建设年-项目顺序号，各段之间不加任何连接符。其中：

● 机构代码、资源类型、项目建设年使用规则见记 录标识号规则。

● 项目顺序号为 2 位数字，用于区分同一年度的不 同项目。同一单位在同一年度如果只提交 1 个项目，则 项目顺序号为 01；如果提交了多个项目，则顺序号从 01 开始顺序排列。

● 同一项目加工的同类型资源（即资源类型代码相 同），其全部元数据尽可能集合成一个元数据文件。由于 元数据数量较多、容量较大等特殊情况可能导致全部元 数据需要分为多个元数据文件存储的，元数据文件命名 可在 13 位数字后增加 3 位文件顺序号（从 001 开始顺序 排列），用下划线连接,例如：0000100202101_001

（三）对象数据文件命名

各类对象数据文件名可根据建设单位和建设项目具体情况 而确定，一般采用数字或者数字与英文字母组合的命名形式，命名的序号顺序应与基础文献页码、音视频资源及相关文件、新型数字资源的内容顺序一致。

● 图书、报纸、期刊、古籍等图像类资源每页建立 一个对象数据。

● 音视频类对象数据包括：音视频文件、字幕文件、责任者头像文件、音视频资源封面文件及与资源相关的 可下载文件。

● 新型数字资源的对象数据区分发布成品与源文 件。

（四）证明文件命名

版权证明文件命名由 6 段组成，共 18 位数字：机构代码 - 资源类型-项目建设年-项目顺序号-bq-3 位流水号，各段之间不 加任何连接符。其中：

● 项目顺序号与元数据文件命名项目顺序号保持一 致；

● 流水号从 001 开始顺序排列。

（五）数据说明文件命名

数据说明文件命名由 6 段组成，共 18 位数字：机构代码 - 资源类型-项目建设年-项目顺序号-sm-3 位流水号，各段之间不 加任何连接符。其中：

- 项目顺序号与元数据文件命名项目顺序号保持一 致；
- 流水号从 001 开始顺序排列。

二、文件存储结构

（一）元数据存储

1.基础文献元数据文件存储路径为：

根目录\机构代码\项目顺序号\metadata\

项目顺序号与元数据命名中的项目顺序号保持一致，下同。

（二）对象数据存储

对象数据文件存储路径为：

根目录\机构代码\项目顺序号\object\对象数据格式\基础 文献记录标识号\分册（集）号\

其中：

- 对象数据格式是指在这一层级按照对象数据的格 式建立文件夹， 以文件格式作为文件夹名称，如:TIF、 PDF、JPG、TXT、MPG、MP4、WAV、MP3、SRT、PPT、DOC、CAD、PSD 等，存储对应格式的对象数据。

- 精细化标引项目对象数据按种集中存放于相应的 基础文献记录标识号文件夹下。

● 分册（集）号为 3 位数字，是指图像类对象数据 的分册号、音视频资源和新型数字资源的分集号，从 001 开始顺序排列。

（三）证明文件存储

证明文件存储路径为：根目录\机构代码\项 目顺序号\zhengming\

（四）数据说明文件存储

数据说明文件存储路径为：根目录\机构代码\项目顺序号\shuoming\

五、验收要求

符合国家图书馆智慧图书馆体建设项目关于“知识资源细颗粒度建设和标签标引”建设标准

对前期项目建设内容，提交本馆，如果不满足国家验收标准，承建商需对建设内容进行调整，最终以国家相关标准为准。

项目进度严格按照自治区、国家的要求进度执行。

六、售后服务要求

根据项目的技术要求，提供切实可行、响应及时的售后服务方案和质量保证措施。

根据项目成果的内容特色，提供技术培训及资源建设成果推广服务等工作。

售后服务承诺，遵照招标文件的采购需求、技术要求、版权要求、推广服务、技术培训、验收要求等，提供真实有效、内容详尽的售后服务承诺。为严格按照交付期提交建设成果，需附详细可行的交付进度表。

通辽市图书馆

2023 年 9 月