

附件 1

推广工程数字资源联合建设
报备表
(2014)

数字图书馆推广工程

2014 年 9 月

说 明

(一) 填写报备表前, 请认真阅读《关于开展数字图书馆推广工程 2014 年度数字资源联合建设工作的通知》的有关规定。报备各项内容, 要实事求是, 逐条认真填写。表达要明确、严谨, 字迹要清晰易辨。第一次出现的缩写词, 须标注全称。

(二) 经费金额以万元为单位, 用阿拉伯数字表示。

(三) 报备表用 A4 纸打印, 于左侧装订成册。各栏空格不够时, 请自行加页。

(四) 省馆和市级承建馆均须填写承建项目报备表(表 1—表 6), 签署意见并盖章; 省馆同时负责本省资源报备情况汇总工作(表 7), 将本省区域内承建馆报备表与汇总表统一寄送至国家图书馆, 同时提交电子版。

1 报备表原件寄送地址:

北京市海淀区中关村南大街 33 号国家图书馆数字资源部

邮编: 100081 收件人: 韩 萌 汪 静

联系电话: (010) 88544243 (010) 88544639

2 报备表电子版发送至:

E-mail: szzylj@nlc.gov.cn

表1

推广工程数字资源联合建设项目报备表

单位名称				
一、项目负责人及主要参与人员情况				
1、项目负责人				
姓名	职务	职称	电话	邮箱
2、分项目负责人				
项目名称	姓名	部门及职务	电话	
元数据仓储				
唯一标识符注册与维护				
地方文献数字化				
政府信息公开				
网事典藏				
二、项目建设内容				
项目名称	建设量	情况说明		
元数据仓储	(条)			
唯一标识符注册与维护	(条)			
地方文献数字化	(页)			
政府信息公开	(条)			
网事典藏	(个)			
注：1、唯一标识符注册与维护只需唯一标识符系统部署馆填写；地方特色文献数字化项目只需省馆填写；网事典藏项目相关参与试验馆填写。 2、若本馆元数据仓储数量未达要求，予以情况说明，并用政府信息公开或地方文献数字化抵充（2条元数据=1条政府信息/地方文献数字化）。				

三、经费使用情况说明	
支出科目	金额（万元）
元数据仓储	
唯一标识符注册与维护	
地方特色文献数字化	
政府信息公开	
网事典藏	
项目质检费	
其他费用（请标注）	
四、报备单位声明	
<p>我单位保证上述填报内容的真实性。我单位将严格遵守《关于开展数字图书馆推广工程 2014 年度数字资源联合建设工作的通知》有关规定，遵循该项目采用的标准规范，按照《数字图书馆推广工程数字资源联合建设报备表（2014）》，认真开展资源建设工作，完成本馆建设任务。</p> <p style="text-align: right;">单位负责人：（签章）</p> <p style="text-align: right;">年 月 日</p>	

表 2

推广工程数字资源联合建设元数据仓储项目报备表

单位名称				
提交数量（条）				
一、项目负责人及主要参与人员情况				
1、负责人				
姓名	职务	职称	电话	邮箱
2、参与人				
姓名	职务	职称	电话	邮箱
二、项目建设内容				
1、本馆自建资源建设情况				
自建资源容量		自建资源登记 完成情况		
条数：	容量（TB）：	完成数量（条）：	完成比例（%）：	
2、2014年建设内容				
填表说明：本项描述2014年元数据仓储项目的主要数据来源、资源类型等。				

三、建设方式	
<input type="checkbox"/> 自建	<input type="checkbox"/> 外包
四、情况说明	
填表说明：若元数据仓储数据量无法满足要求则填写，否则不填。	
五、建设周期与进度安排	
六、经费预算（万元）	

表3

推广工程数字资源联合建设唯一标识符注册与维护项目
报备表

单位名称				
一、唯一标识符系统部署情况				
<input type="checkbox"/> 已部署		<input type="checkbox"/> 未部署		
二、唯一标识符系统预计部署时间（未部署馆填写）				
三、项目负责人及主要参与人员情况（已部署馆填写）				
1、负责人				
姓名	职务	职称	电话	邮箱
2、参与人				
姓名	职务	职称	电话	邮箱
四、项目建设内容（已部署馆填写）				
填表说明：本项描述2014年唯一标识符注册与维护项目的主要数据来源、资源类型等。				

五、建设方式（已部署馆填写）	
<input type="checkbox"/> 自建	<input type="checkbox"/> 外包
六、建设周期与进度安排	
<div></div>	
七、经费预算（万元）	
<div></div>	

表4

推广工程数字资源联合建设地方文献数字化项目申报表
(省馆填写)

单位名称				
一、项目负责人及主要参与人员情况				
1、负责人				
姓名	职务	职称	电话	邮箱
2、参与人				
姓名	职务	职称	电话	邮箱
二、项目建设内容				
1、文献内容				
<p>填表说明：本项描述2014年地方文献数字化项目的主要资源来源及文献内容，具体文献信息请填入附表。</p>				

附表

推广工程数字资源联合建设地方文献数字化项目
文献详细信息表

文献名称	责任者	出版社	出版时间	ISBN	页数	尺寸	分类	版权状态	备注
------	-----	-----	------	------	----	----	----	------	----

注：此表为Excel表头，提交时请将文献元数据基本信息填写完整，提交Excel表格。

备注项填写文献的状态信息，如是否缺页等；

表5

推广工程数字资源联合建设政府信息公开项目报备表

单位名称				
提交数量（条）				
一、项目负责人及主要参与人员情况				
1、负责人				
姓名	职务	职称	电话	邮箱
2、参与人				
姓名	职务	职称	电话	邮箱
二、项目建设内容				
填表说明：本项描述2014年政府信息公开的主要数据来源等。				
三、建设方式				
<input type="checkbox"/> 自建		<input type="checkbox"/> 外包		
四、情况说明				
填表说明：若政府信息无法满足要求则填写，否则不填。				

五、建设周期与进度安排
六、经费预算（万元）

表6

推广工程数字资源联合建设网事典藏项目报备表
(相关参与试验馆填写)

单位名称				
提交数量(个)				
一、项目负责人及主要参与人员情况				
1、负责人				
姓名	职务	职称	电话	邮箱
2、参与人				
姓名	职务	职称	电话	邮箱
二、项目建设内容				
<p>填表说明：本项描述2014年推广工程数字资源联合建设网事典藏项目主要数据来源等，具体网站采集地址请以Excel方式提交采集列表。</p>				

三、建设周期与进度安排
四、经费预算（万元）

表7

推广工程数字资源联合建设项目报备汇总表

一、 承建机构

省份				
省馆				
市馆				

二、 项目人员信息(省市馆项目负责人信息)

姓名	单位	部门	职务	联系方式		
				座机	手机	邮箱

三、 资源建设量

项目名称	建设量 (全省)	建设量 (省馆)	建设量 (具体市馆名称)						
元数据仓储	(条)								
唯一标识符注册与维护	(条)								
地方文献数字化	(页)								
政府信息公开	(条)								
网事典藏	(个)								

四、 省馆报备声明

我单位保证上述填报内容的真实性。我单位将严格遵守《关于开展数字图书馆推广工程 2014 年度数字资源联合建设工作的通知》有关规定，统筹本省市级承建馆，遵循该项目采用的标准规范，按照《数字图书馆推广工程数字资源联合建设报备表（2014）》各项报备指标，认真开展资源建设工作，完成本省建设任务。

单位负责人：（签章）

年 月 日

附件 2

推广工程数字资源联合建设 元数据仓储项目建设方案（2014）

为加强我国公共图书馆对数字资源的组织揭示能力、实现资源的共知共享，2014 年国家图书馆将继续联合省市级公共图书馆，开展元数据仓储建设。

一、建设内容

元数据仓储收录对象为分布在全国各公共图书馆已发布的且拥有对象数据的自建数字资源。

省市级图书馆每馆提交数据不少于 10000 条。本项目不收录无对象数据的数字资源、未发布的数字资源、推广工程前期及本项目已建设资源、已纳入国家图书馆数字资源征集项目及登记项目的资源、联合编目系统中的资源。

二、工作流程

1、提交元数据样例

省市级图书馆依据资源类型，对应《元数据简表》，编写元数据样例。市馆样例先提交到省馆，由省馆规范后统一提交至国家图书馆。国家图书馆反馈样例数据修改意见。

2、元数据制作

各级图书馆根据国家图书馆的反馈意见进行元数据制作。新

制作的元数据需要先在一标识符系统中注册唯一标识符，原已在元数据仓储中的数据需要补充唯一标识符字段的内容。

元数据可以直接在元数据仓储的平台中制作；也可在系统中选择“从 EXCEL 创建记录”，将字段填写完成后，统一导入系统平台。

3、元数据维护

各级图书馆负责在元数据仓储的管理平台上对本机构制作的元数据进行长期维护，当资源状态发生变动时，要及时修改元数据，以保证元数据的正确有效。

三、成果提交

- 1、按本项目规则要求制作的元数据。
- 2、第三方质检报告。

附录: 元数据仓储著录规则

附录

元数据仓储著录规则

一、著录对象

资源对象为分布在全国各公共图书馆已发布的自建数字资源，类型包括但不限于数据库、图书、连续性资源、文章、图像、音频、视频、网页等。

二、著录粒度

数字资源的著录粒度，是指数字资源的记录单元，一般以具有独立名称，并可独立使用的一个资源为著录单位，如一个数据库、一种图书、一本期刊、一幅图像、一个网页等，以及数据库中的子库、会议论文集或期刊中的一篇论文、一张光盘中的一个曲目等。

三、元数据简表

序号	元素	修饰词	必备性	注释
1	名称	正题名	必备	著录资源的正题名，根据在资源中出现的形式著录，内容为可以概括资源内容的词、词组、符号等，不能是无意义的名称。各语种的名称著录要严格按照各语种的语言习惯。
		其他题名	有则必备	著录资源除正题名以外的题名信息，包括并列题名、交替题名等各类题名信息。
2	责任者	责任者名称	有则必备	创建资源内容的主要责任者或对资源内容做出贡献的其他实体。
		责任方式	有则必备	责任者与资源之间的责任关系。
3	主题	中图分类	非古籍类 必备	描述资源主题内容的中图分类法类号。依照 CLC（《中国图书馆分类法》第五版）进行编著。
		四库分类 （数据库、	古籍类 必备	古籍数据库、图书主题内容的四库分类法类号。

		图书)		
		主题词或 关键词	必备	描述资源主题内容的受控或非受控的词汇。受控主题词依据 CCT (中国分类主题词表) 著录。
4	描述	摘要	必备	对资源内容的说明描述。
		目录	有则必备	资源内容的目录, 如组图中各幅图片的名称等。连续性资源、文章及网页类资源可不著录此字段。
		图像 (数据库)	必备	数据库的图像, 可以是发布后的页面截图, 上传图像为 JPG 格式, 大小为 800×600-1024×768 。
		附注	有则必备	关于资源内容的其他说明。
5	出版者	出版者名称	有则必备	使资源可以获得和利用的责任实体, 例如资源的出版者或颁布者。
		出版地	有则必备	著录出版者的所在地, 一般著录到城市, 如无法著录到城市名称, 则著录合适的州、省或国家。
6	日期	开始日期 (连续性资源)	有则必备	记录连续性资源最早一期的正式出版时间。
		终止日期 (连续性资源)	有则必备	结束出版日期, 记录资源结束出版的日期。
		出版日期	有则必备	用于注明电子资源出版或颁布的日期。具体著录时, 对派生电子资源著录印刷版原件的出版日期。
		创建日期	有则必备	著录电子资源的创建日期, 用于原生电子资源。例如照片的拍摄日期, 绘画的创作日期等。
		发布日期	有则必备	著录发布资源的日期。
7	更新频率 (数据库)		必备	元素的取值为: **更新、停止更新、不定期更新、不详。

8	类型	内容形式	必备	著录资源的内容形式词及内容限定词, 取值参见国标《信息资源的内容形式和媒体类型标识(GB/T 3469-2013)》。 例: 图像(动态)
		媒体类型	必备	著录资源的媒体类型, 取值参见国标《信息资源的内容形式和媒体类型标识(GB/T 3469-2013)》。 例: 视频
		学科类型 (数据库)	必备	著录数据库的学科分类, 分类表见表后《学科分类表》。
9	原始标识符		有则必备	著录资源在原始数据库中的唯一标识符。
10	格式	文件大小	有则必备	资源文件的大小, 单位为 KB、MB 或 GB 等。连续性资源、数据库可不著录此字段。
		文件格式	必备	资源对象数据的格式。 数据库可不著录此字段。
		颜色 (图像、视频)	必备	资源的颜色, 主要用于图像和视频 图像的取值为: 彩色、黑白、灰度。 视频的取值为: 为彩色、黑白。
		分辨率 (图像、视频)	必备	资源的分辨率, 主要用于图像和视频 如 640×480, 其中“×”为乘号, 不能用“*”或英文字母“X”。
		声音 (视频)	必备	著录资源是否有声音, 主要用于视频。 视频取值为有声、无声。
		比特率 (音频)	必备	音频资源的比特率, 如 192 KBIT。
		时长 (音频、视	必备	著录资源持续的时间, 主要用于音频和视频。如 00:55:03。

		频)		
		页码 (图书)	必备	著录图书总页数。
		运行环境 (数据库)	有则必备	著录数据库运行所必须的系统环境。
11	语种	文字语种	数据库、 图书、连 续性资 源、文 章、网页 必备，图 像有则必 备	著录资源的文字语种。著录 3 位语种 编码。例：chi、eng。
		字幕语种 (视频)	有则必备	著录视频资源的字幕语种。例： chi。
		声道语种 (音频、视 频)	有则必备	著录音视频资源的声道语种。例： chi。
12	关联	包含于	有则必备	说明资源是另一种资源的物理或逻辑 组成部分。
		所属数据库	有则必备	著录资源的所属数据库名称。
13	版本 (图书)		有则必备	著录图书的版本或版次信息以及其他 附加说明。例：影印版
14	权限	访问方式	必备	资源可以提供服务的范围。取值为： 互联网访问、局域网访问。
		访问范围	局域网访 问必备	该字段著录在什么范围内的局域网 能够访问，规范为“XX 局域网”，如 “国家图书馆局域网”。
15	发布地址		必备	资源在因特网或局域网的发布地址。
16	适用对象		必备	资源适宜使用的人群，取值为：普通 用户、少年儿童、残障人士。
17	出版频率 (连续性资源)		必备	著录资源出版的时间间隔。 例：半月刊、月刊、季刊、年刊。

18	馆藏范围 (连续性资源)		必备	著录馆藏电子连续出版物的收录范围。著录方式请参照《通用著录规则》。
19	标识符	ISSN	有则必备	著录在国际标准连续出版物编号中确认资源的唯一标识。
		ISRC	有则必备	著录在国际标准音频制品编码中确认音像资源的唯一标识。
		ISBN	有则必备	著录在国际标准书号中确认资源的唯一标识。
		统一书号	有则必备	著录在统一书号中确认资源的唯一标识。
		CN	有则必备	著录在国内统一刊号中确认资源的唯一标识。
		DOI	有则必备	著录在 DOI 中确认资源的唯一标识。
		CDOI	必备	著录在 CDOI 中确认资源唯一标识。
20	出处 (文章)		必备	著录文章所在的期刊、报纸或论文集的年卷期版等信息。
21	数据提交单位		必备	描述提交该资源的单位,如国家图书馆、山东省图书馆等。
22	数据提交日期		必备	描述提交该资源的日期,遵循《通用著录规则》日期格式,如 2014-07-16。
23	其他字段		有则必备	资源元数据中其他字段,直接保留。

学科分类表

人文与社会科学	语言/文学/艺术 文化/科学/教育/体育 历史/地理 经济/管理 政治/法律 军事 社会科学总论
工程与技术科学	航空航天/交通运输 水利工程/建筑学 环境科学/安全科学 计算机/自动化 动力/机械/能源 材料科学/其它工业技术 航天/交通运输 水利工程/建筑学 环境科学/安全科学 计算机/自动化 动力/机械/能源 材料科学/其它工业技术
农业科学与医药科学	医学 农业科学
自然科学	数学/物理/化学 生物学 天文学/地球科学

四、通用著录规则

序号	字段	标准格式
1	日期	年: YYYY (eg 1997) 年和月: YYYY-MM (eg 1997-07) 年月日: YYYY-MM-DD (eg 1997-07-16) 如果不确定的日期, 可填“不详”, 在附注中用[]备注日期。如果不确定具体年份, 可用? 代替。如 20 世纪 90 年代, 可填为 199?。
2	时长	HH: MM: SS (例 01: 33: 23)
3	责任者名称	同种责任方式对应的不同责任者名称间分隔用 半角分号 。 如果责任者有朝代或国家限制, 则在责任者前用() 半角括号 著录责任者朝代或国家。 例: (清)曹雪芹
4	责任方式	不同责任方式对应的责任者名称, 应分别著录。 例: 责任者名称 1: (美)拉斯·克兰德; 责任方式 1: 著; 责任者名称 2: 陈永剑; 责任方式 2: 译
5	主题词或关键词	主题词或关键词之间的分隔用 半角分号 。
6	摘要	著录摘要字段, 描述文字不少于 100 字。
7	语种	采用 3 位语种代码, 可参考《新版中国机读目录格式使用手册》。常见代码如下: 中: chi 英: eng 藏: tib 蒙: mon 维吾尔: uig
8	文件格式	文件格式需用英文大写著录。例: TXT
9	文件大小	文件大小的著录格式为英文大写。 例如: 192MB

10	目录	<p>各组成单元间以半角分号“;”分隔。在各个组成单元中,可依次著录正题名、其他题名信息以及责任者等信息,并在其他题名信息前冠半角冒号“:”,责任者前冠斜线“/”。</p> <p>例:第一篇,修辞学/亚里士多德著;第二篇,论出版自由/弥尔顿著;</p>
11	出版地	要写“省”或“市”,如吉林省。
12	馆藏范围	<p>著录电子连续性出版物的馆藏范围。</p> <p>如果某一时间段内的电子资源全部存在,可以以时间段著录。例:1990年第7期至2014年第7期。</p> <p>如果电子资源不具有连续性,中间存在缺刊,则需分段著录,中间用半角分号分隔。</p> <p>例:1990年第1期至2004年第5期;2005年第2期至2014年第7期。</p>
13	其他字段	<p>其他字段自由行文前应加前置导语,并用半角冒号分隔;自由行文。</p> <p>各组成单元间需用半角分号分隔。</p> <p>例:印制地:北京;印制者:商务印刷馆</p>

附件 3

推广工程数字资源联合建设 唯一标识符注册与维护项目建设方案（2014）

为加强我国公共图书馆数字资源的管理能力，促进全国公共图书馆数字资源的共建共享，推广工程联合全国各省、市级公共图书馆共同开展全国数字资源唯一标识符注册与维护工作，共同建设公共图书馆数字资源唯一标识符体系。

一、建设内容

公共图书馆数字资源唯一标识符注册与维护项目主要围绕全国公共图书馆已发布的馆藏自建资源进行注册。资源类型包括但不限于文本文档、数据集、图书、照片、连续出版物、音频资源、视频资源、软件、抽象作品、艺术作品等。本项目将其归纳为八类，分别是图书、期刊、论文、报纸、图片、音频、视频，以及网页采集数据。

二、工作流程

1. 数据准备

各图书馆应按照附件“唯一标识符注册数据著录规则”提交 excel、xml 或 access 格式的唯一标识符注册数据，注册数据模板可以在唯一标识符系统【元数据注册（批量）】界面下载。

2. 唯一标识符注册

唯一标识符注册包括两种方式，分别为手工单条注册和批量导入注册。手工单条注册需要打开唯一标识符系统【元数据注册】界面，填写注册数据，提交后即可生成一条唯一标识符；批量导入注册需要打开唯一标识符系统【元数据注册（批量）】界面，根据“批量元数据维护”模板，整理唯一标识符注册数据，提交系统后即可批量生成唯一标识符数据。

3. 对象数据 URL 地址管理

打开唯一标识符系统【URL 维护（批量）】界面，根据“URL 维护”模板，整理成功注册唯一标识符的对象数据 URL 地址，提交系统后即可完成对象数据 URL 地址的回写。

4. 数据维护

对本机构所注册唯一标识符的元数据和发布地址进行长期维护，保障数据准确无误并及时更新。

三、成果提交

- 1、在唯一标识符平台上存储并注册唯一标识符的数据。
- 2、第三方数据质检报告。

附录：唯一标识符注册数据著录规则

附录

唯一标识符注册数据著录规则

一、数据质量总体要求

1. 唯一标识符注册数据应完整、准确、命名规范、不得出现错别字，数据组织合理；
2. 无访问地址的唯一标识符数据需要在本项目规定时间范围内补充完整；
3. 进行全部数据的检查，错误率不超过万分之三。

二、唯一标识符注册数据标准

1. 注册粒度

唯一标识符的标识对象粒度包括多种形式，如：单册图书和一套丛书；图书的章节；期刊；期刊中的一期；一期中的目次、一篇论文；一篇论文的几页、单页；单页中的一幅图片、公式；多媒体资源集合；一首乐曲；乐曲中的一段音乐等。本项目中，资源的标识粒度以“种、卷（册/首/幅）”为原则。以下注册元素将依照此原则进行说明。

本规范包含唯一标识符注册数据的 23 个元素，其中元数据注册字段 19 项，URL 地址注册字段 4 项，分别描述如下。

2. 元数据注册字段

注册文件模板请参见唯一标识符系统【元数据注册】菜单中【批量元数据注册】子菜单的“模板下载”项。元数据注册字段共有 19 项，具体描述如下：

表 1 元数据注册字段说明表

序号	数据元素	注册文件模板的 对应字段	字段注释	必备性	可重复性
1	编目系统系统号	系统号	使用资源对应的编目系统的加工流水号，若无，则使用资源来源系统流水号。	必备	不可重复
2	MARC 记录标识号	MARC 记录 001 标识号	MARC 记录 001 字段	有则必备	不可重复

3	名称	题名(资源名称)	标识对象最通常被使用的名字, 如题名。	必备	可重复
4	作者	作者	创建资源内容的主要责任者或者对资源内容做出贡献的其他实体。	有则必备	可重复
5	ISBN	ISBN	国际标准书号	有则必备	不可重复
6	ISSN	ISSN	国际标准连续出版物编号	有则必备	不可重复
7	出版者	出版者	使资源成为可以获得并可用的责任者, 包括个体, 组织, 或服务。	有则必备	可重复
8	出版时间	出版时间	用于注明电子资源出版或颁布的日期。建议采用的日期格式应符合 ISO 8601 [W3CDTF] 规范	有则必备	不可重复
9	标识对象格式	格式编号	如 PDF、TIFF、JPG 等, 具体参见表 3	必备	不可重复
10	标识对象类型	资源种类编号	著录资源的资源类型, 元素的取值包括但不限于图书、期刊、论文、报纸、图片、音频、视频, 以及网页采集数据, 具体参见表 4。	必备	不可重复
11	标识粒度	颗粒度 K1	标识对象的一级粒度, 如种、卷(册/首/幅)	有则必备	不可重复
12	粒度值	K1 值	描述一级粒度的值	有则必备	不可重复
13	关联关系	关联	描述该资源与其他资源的关联关系	有则必备	不可重复
14	语种	语种	著录资源的语种, 包括文字、声道的语种	有则必备	可重复
15	数据提交单位	来源	描述该资源来自单位, 如国家图书馆、山东省图书馆等	有则必备	可重复

16	描述信息	描述信息	资源内容的说明。描述可以包括但不限于以下内容：文摘、目录、对以图形来揭示内容的资源而言的文字说明、或者一个有关资源内容的自由文本描述。	有则必备	可重复
17	原始资源库标识号	扩展字段 1	原始资源库流水号。	有则必备	不可重复
18	与编目系统的关联信息	扩展字段 2	描述文献资源与编目系统中元数据的关联关系。	有则必备	可重复
19	所属数据库	扩展字段 3	著录资源内容的类型，如：抗日战争图片库、徽州家谱数据库、长白山动植菌物数据库等。	有则必备	不可重复
20	ISRC	扩展字段 4	音像制品的国际标准编码	有则必备	不可重复

3. URL 地址注册字段

URL 地址管理模板参见唯一标识符系统【元数据维护】菜单中【URL 维护（批量）】子菜单的“模板下载”项。URL 地址管理涉及增、删、改三种操作方法，具体模板填写规则如下：

表 2 URL 地址注册字段说明

序号	数据元素	操作方法 1 (新增 URL 地址)	操作方法 2 (修改 URL 地址)	操作方法 3 (删除 URL 地址)
1	操作类型 (ADD/MOD/DEL)	填写“ADD”	填写“MOD”	填写“DEL”
2	唯一标识符	必备	必备	必备
3	需替换的 url	空	必备 填写原始发布地址	必备 填写原始发布地址
4	替换后的 url	必备 填写新发布地址	必备 填写新发布地址	空

4. 注意事项说明

(1) 出版时间

建议参照 W3CDTF、period、中国时代表；日期取值自编码体系修饰词，一般采用 YYYY-MM-DD 的表达方式。

(2) 文件格式说明

在资源注册时对于不同的资源格式进行不同的编号,用于区分资源格式,如下表所示:

表 3 资源格式编码表

编号	格式名称
F1	ARC
F2	ARJ
F3	ASP
F4	AVI
F5	BMP
F6	CAB
F7	DBF
F8	DOC
F9	FLV
F10	GIF
F11	ICO
F12	ISO-Disk
F13	ISO-Marc
F14	JPG
F15	LZH
F16	LZW
F17	MDB
F18	MID
F19	MOV
F20	MP3

F21	MPEG2
F22	MPEG4
F23	PCD
F24	PDF
F25	PIC
F26	PNG
F27	PPT
F28	PSD
F29	RAR
F30	TAR
F31	TIF
F32	TXT
F33	WAV
F34	XLS
F35	XML
F36	YUV
F37	ZIP
F38	JP2
F39	WMV
F40	HTML
F41	PDG
F42	CEBX
F43	MPG
F44	SWF
F45	WARC

（3）资源类型说明

数字资源分为八大类，每种类型的编号如下表所示：

表 4 资源类型编码表

编号	资源类型
T1	图书
T2	期刊
T3	论文
T4	报纸
T5	图片
T6	音频
T7	视频
T8	网页采集

（4）注册数据颗粒度描述

针对多卷（册）数据，需要对数字资源进行颗粒度信息的标注。例如，对于图书类资源，需要标注卷（册）信息；对于期刊类资源，需要标注出版年、期刊号；对于报纸类资源，需要标注出版年、期刊号、版次信息等。

具体规则描述如下：

表 5 注册数据颗粒度描述规则

资源类型	颗粒度 1	颗粒度 1 (Key)	颗粒度 1 (可输入值)
图书 (T1)	单卷/册	无输入	无输入
	多卷/册	T1K1V2	输入 1
期刊 (T2)	全年	T2K1V1	输入 2
	期刊	T2K1V2	输入 3
	增刊	T2K1V3	输入 4
	合订本	T2K1V4	输入 2
论文 (T3)	无输入	无输入	无输入
报纸 (T4)	全年	T4K1V1	输入 2
	期刊	T4K1V2	输入 3
图片 (T5)	图号	T5K1V1	输入 1
音频 (T6)	单曲号	无输入	无输入

	多曲号	T6K1V2	输入 1
视频 (T7)	单场次	无输入	无输入
	多场次	T7K1V2	输入 1
网页采集 (T8)	页号	T8K1V1	输入 1

期刊和报纸涉及多级粒度的描述，具体说明如下：

A. 期刊二级粒度（输入 1）描述说明

- ① 前封(含封一、封二): T2K2V1
- ② 目录页之前的前附页: T2K2V2
- ③ 目录页之后的前附页: T2K2V3
- ④ 目录页: T2K2V4
- ⑤ 正文: T2K2V5
- ⑥ 后附页: T2K2V6
- ⑦ 后封: T2K2V7

B. 报纸多级粒度描述说明

➤ 二级粒度（输入 1）:

- ① 版面: T4K2V1

➤ 三级粒度（输入 1）:

- ① 插图: T4K3V1
- ② 文章: T4K3V2

颗粒度取值范围描述如下：

输入名称	输入数据	数据类型	数据解释
输入 1	a	数字	第 a 卷/册/页
	b-c	起始数字	第 b-c 卷/册/页
	a, b-c	数字、起始数字任意组合(逗号分割)	
输入 2	YYYY	年份(4 位年)	
	YYYY-YYYY	起始年份-结束年份	某年至某年

	YYYY, YYYY-YYYY	年份、起始年份-结束年份任意组合(逗号分割)	
输入 3	YYYYim	起始年份(4 位年)+字母 i+m 期 (最少 1 位数字)	某年的 m 期
	YYYYia-YYYYib	起始年份(4 位年)+字母 i+a 期 (最少 1 位数字)-结束年份(4 位年)+字母 i+b 期(最少 1 位数字)	某年的 a 期至某年的 b 期
输入 4	YYYYsm	起始年份(4 位年)+字母 s+第 m 增刊(最少 1 位数字)	某年的第 m 增刊
	YYYYsa-YYYYsb	起始年份(4 位年)+字母 s+第 a 增刊(最少 1 位数字)-结束年份(4 位年)+字母 s+第 b 增刊(最少 1 位数字)	某年的第 a 增刊至某年的第 b 增刊

附件 4

推广工程数字资源联合建设 地方文献数字化项目建设方案（2014）

文献数字化工作是实现数字资源广泛传播和共享利用的基础。为全面展现我国各地区的社会、历史、政治、经济文化全貌，推广工程将联合全国各省级公共图书馆依托馆藏资源开展地方文献数字化工作，建设具有鲜明地域特色、较高文献价值和历史价值的地方文献数字化资源。

一、建设内容

本项目建设对象为 1949 年以来的地方文献，包括但不限于行业志、地方志等资源。申报文献不限于正式出版物，已进入公有领域或者自有版权的地方文献优先。2014 年面向省馆开展，每个省馆数字化加工不少于 10000 页。

二、工作流程

1、文献目录确认

各省馆根据项目要求，整理资源清单并报备至国家图书馆，国家图书馆对各省馆提交的申报内容进行审核查重，确定资源加工清单，并向各省馆返回任务确认书。

2、数字化加工

各省馆按照“地方文献数字化加工规范”（附录 A）对文献进

行数字化加工，对象数据命名和保存均要按加工规范的要求进行，同时需要制作相应的数据库和说明文件。

3、数据验收

按照联建方案规定的项目进度，各省馆在规定日期之前，向国家图书馆提交已由第三方机构初检合格的全部数字化资源。经国家图书馆终验合格后，提交成品数据。

三、成果提交

1、需要提交的内容包括：TIFF 图像(单页)、双层 PDF (单版和合并版)、TXT 文件(单版和合并版)、对应的数据库、说明文件以及文献的书目数据(ISO 格式)。

2、验收数据提交时以上内容各上交一份，以硬盘方式提交至国家图书馆(验收完成后硬盘返还提交馆)，同时提交“地方文献验收数据提交单”(附录 B)和第三方质检报告。

3、成品数据以硬盘方式提交，同时提交“地方文献成品数据提交单”(附录 C)，国家图书馆清点内容和数量后，签字接受成品(硬盘返还提交馆)。

附录 A: 《地方文献数字化加工规范》

附录 B: 《地方文献验收数据提交单》

附录 C: 《地方文献成品数据提交单》

附录 D: 《地方文献数字化机构代码》

附录 A

地方文献数字化加工规范

1 术语和定义

1.1 档案典藏级

用于档案数据保存及必要时出版印刷用，不应用于网络。可作格式转换和复制的母本。

对不同类型的对象，其色彩、扫描精度、位深等参数的具体要求也不同。TIFF 格式。

1.2 双层 PDF

将标准资料通过扫描仪快速录入后，经过去污、纠偏和 OCR 识别，直接生成的可检索的 PDF 文件。该文件是双层的，上层是原始图像，下层是识别结果。

1.3 单版 TXT 文件

由图像数据一一对应进行文本转换得到的 TXT 文件。一个单版 TXT 文件对应一个图像文件。

1.4 合并版 TXT 文件

将单版 TXT 文件按顺序合并得到一册书的完整 TXT 文件。

2 图像数字化和命名规则

2.1 图像数字标准

2.1.1 扫描标准

(1) 黑白页和灰度页用灰度方式扫描

色彩位深：8 位

分辨率：300 dpi；小于 5 号字体用 400 dpi

档案典藏级格式：TIFF 不压缩

(2) 彩色页用彩色方式扫描

色彩位深：24 位

分辨率：300 dpi；小于 5 号字体用 400 dpi

档案典藏级格式：TIFF 不压缩

2.1.2 拍照标准

像素：不小于 300 万

档案典藏级格式：TIFF 不压缩

2.2 数字化要求

数字化环境注意防护光源，避免透光或反射光的影响。

数字化后的图像清晰，文件页码连续，没有重页、缺页，错页等情况（原书缺页、错页除外）。补扫的图像要与同册图像文件的大小一致，颜色接近。

- （1）以原文献的上边沿为基准，以中缝为中心线，保持原文献的天头、地脚的尺寸不变，左右两边的尺寸基本不变。
- （2）数字图像放大至实际尺寸 100%，图像不失真。
- （3）数字图像文件与文献原件颜色不一致，须先进行设备色彩校正，再重新进行扫描或拍照工作。

2.3 数字图像处理要求

数字图像处理是在未改变原扫描图像的色彩、分辨率、格式、压缩等情况下进行。数字图像文件处理内容及要求如下：

- （1）纠偏处理。对出现偏斜的图像进行纠偏处理，图像歪斜度不可以超过一度，对方向不正确的图像进行旋转还原，以符合阅读习惯。
- （2）图像剪裁。图像保留到文献的外边缘。
- （3）不能进行锐化或者图像增强处理，不能更改图像的颜色，尽量减少对图像文件的后期处理。

2.4 数字图像检查要求

根据本项目要求，数字化单位检查各级别图像数据的质量，建议扫描完成后立即进行质检。数据检查的内容和要求如下：

- （1）图像文件（各种格式）放大到 1:1 状态，逐页检查。检查文件是否有透光、透字、有彩点、彩线、太淡、太浓、黑边、污点、歪斜、模糊（马赛克等）或图像内容不完整等现象。若不符合图像质量要求应进行图像校正或重新扫描（拍照）。
- （2）发现文件漏扫时，应及时补扫并正确插入图像。
- （3）拼接图像接缝处无错位、无缝吻合，不应出现白边和内容缺失，没有明显的歪斜。

- (4) 检查是否符合扫描（拍照）规格要求和技术参数。
- (5) 所有文件保存位置正确，可以有效打开和显示。
- (6) 图像名称必需正确，同一数据流水号不得有跳号情况，按顺序排列命名；图像文件的排列顺序应与原文献一致。
- (7) 按照命名规则，检查目录、文件、数据库、文档、介质等名称是否正确。
- (8) 检查各类说明、统计、验收等文档是否齐全。

2.5 命名规则

2.5.1 加工编号（book-id）

文献数字化加工过程中一册文献的唯一标识，它由 11 位数字组成。

文献基本资料类型（1 位）、文献语种（1 位）、加工年（2 位）、机构代码（4 位）、单位内部流水号（3 位）。

本规范针对图书的文献基本资料类型为 0，文献语种为 1，加工年为公元年后两位数字（如 14），机构代码由国家图书馆统一分配（见附录 D），单位内部流水号由各单位自行分配，从 1 开始，不足 3 位以 0 补齐。

2.5.2 图像文件名

文件名后缀为小写字母

(1) 前封（含封一、封二）

扫描文件名为 Axxxxx-00，其中 xxxxx 为 5 位数字，按原书顺序依次排序

(2) 前附页

目录页之前的前附页扫描文件名为 Bxxxxx-00，其中 xxxxx 为 5 位数字，按原书顺序依次排序。

目录页之后的前附页扫描文件名为 Dxxxxx-00，其中 xxxxx 为 5 位数字，按原书顺序依次排序。

(3) 目录页

扫描文件名为 Cxxxxx-00，其中 xxxxx 为 5 位数字，按原书顺序依次排序。

(4) 正文

有页码的正文扫描文件名为 Txxxxx-00，其中 xxxxx 为 5 位数字，与原书页号一致，按原书顺序依次排序。

正文中插页扫描文件名为 Txxxxx-yy，其中 xxxxx 为 5 位数字，表示插页的前一页顺

序号,yy 为数字,表示插页,并按原书顺序依次排序。

(5) 后附页

扫描文件名为 Yxxxxx-00,其中 xxxxx 为 5 位数字,按原书顺序依次排序。

(6) 后封(含封三、封四)

扫描文件名为 Zxxxxx-00,其中 xxxxx 为 5 位数字,按原书顺序依次排序。

3 双层 PDF 标准

3.1 双层 PDF 输出

将处理完成后的 TIFF 图像进行 OCR 逐页识别校对,差错率达到千分之三以内。采用图在文上的模式进行双层 PDF 输出,包括单版 PDF 和合并版 PDF 两种,单版 PDF 命名与对应 TIFF 文件保持一致,合并版的命名同该文献的加工编号。

PDF 文件根据图像尺寸、颜色、数据存储量、按 JPEG2000 有损压缩,压缩因子适度动态调整,在确保图像清晰的情况下,尽量压缩图像文件所占空间至最小。

3.2 PDF 目录提取

将输出后的双层 PDF 进行目录标签提取,提取级别上限为三级。提取过程中应按照原书页面顺序进行提取,不可出现章节丢失或级别划分错误等情况。目录内容差错率应低于千分之一。

3.3 PDF 文件质量要求

- (1) 双层 PDF 数据需完整,避免缺页、重页、页码顺序颠倒等问题。
- (2) 双层 PDF 文件的图像层和文字层的文字对位准确,反显区域与文字区域相差 1 毫米以内。
- (3) PDF 文字层所使用的字体以“已嵌入子集”方式嵌入 PDF 文件。缺字对应文本统一为“■”。
- (4) 打开一本电子文献阅览并对文字放大时,保证在放大到百分之二百的时候,字迹清晰,笔画连续,无断裂、缺块的现象。
- (5) 每个双层 PDF 综合差错率小于万分之一。
- (6) 整本 PDF 必须制作书签。书签是电子书的目录,内容和纸质书的目录一致。书签的功能是用户点击书签的某一章节,电子书会自动显示为相应页。打开 PDF 文件时,自动显示书签,书签只展开到第一级目录。

4 图像文本转换要求

4.1 单版 TXT 和合并版 TXT

- (1) 对图像数据逐页进行版式分析，并对版式分析结果进行质检。
- (2) 对图像数据逐页进行文本转换，生成单版 TXT 文件，文件名与对应 TIFF 文件名一致。
- (3) 对单版 TXT 文件进行校对。
- (4) 经验收合格的单版 TXT 文件需合并生成合并版 TXT 文件，合并版文件的命名同加工编号。
- (5) 图像中的注释、下划线等特殊情况需进行专门处理，表格、插图、空白页等需进行描述，描述内容用 [=]进行标识（[=] 均用半角标识），并保存未转换文件对应的图像文件。

4.2 TXT 文件质量要求

- (1) 文本数据应如实反映原书内容、版面等所有原书相关信息。
- (2) 文本数据保存格式为 TXT 纯文本格式。
- (3) 文件命名无误，且在数量上与 TIFF 图像一致。
- (4) 文本数据内容与 TIFF 图像内容吻合，不存在乱码、转换错误等问题。
- (5) 文本数据应如实反映原文的章节、段落，不应出现与文章段落不符的字符、段落、硬回车、空格等。
- (6) 单版 TXT 文件与合并版 TXT 文件内容完全一致。

4.3 特殊情况处理办法

4.3.1 注释

注释分注释类和解说类，仅对有意义注释进行转换，无意义注释不做转换。

- (1) 注释出现在当页，且明确标出与正文对应位置的
将注释内容填入括号“（）”内，插回到原文注释所对应的位置。
- (2) 注释在一篇文章或章节结尾，以参考或引用等专项标题单独列出的
按原文版式转换，不插回原文。
- (3) 解说类注释出现在正文文字段落中间或左右两侧，在正文中无对应位置的将注释内容放在其出现的那段文字后，另起一段，段首标注“注释：”。

- (4) 解说类注释出现在黑框或深色底框内，且在正文文字段落中间的将注释内容放在【】内，原位置转换。

4.3.2 空白页

为保持内容完整性和页面连贯性，正文中空白页需保留，并按照命名规则正确命名，内容标注为“[=此处为空白页=]”。

4.3.3 插图与插图页

(1) 插图

仅对有意义插图进行转换和标注。在插图出现的段落后另起一段，标注为“[=此处为插图（图注）=]”。

(2) 插图页

均需保留，并按照命名规则正确命名，内容标注为：“[=此处为插图页（图注）=]”；对于包含多个插图的插图页，内容标注为：“[=此处为插图页：图一（图注），图二（图注）...=]”。

4.3.4 表格与表格页

表格只转换表格内文字及表注，任何形式表格边框不予转换。

(1) 表格

简单的单列或横向排版表格应予转换；多列复杂表格，内容为分类列举的文字按列予以转换；其他表格不予转换，在表格出现的位置，标注为“[=此处为表格（表格说明）=]”。

(2) 表格页

需保留，并按照命名规则正确命名，内容标注为“[=此处为表格页（表格说明）=]”。

对未进行转换的图像和表格，需在其所属单板 TXT 文件目录内建立“未转换文件对应图像”子目录，将此插图/表格所在的图像文件保存在该子目录内；多个插图/表格对应同一个图像时，仅保存一个图像文件。同时对未进行转换的插图/表格在《文献全文转换未转换文件记录表》中进行记录。

4.3.5 拼音文字混编

拼音文字混编分为对全文均做拼音标注和仅对个别文字做拼音标注两种情况。

(1) 全文均做拼音标注的（拼音一般标注在文字上方）

只转换文字，不转换拼音

(2) 个别文字做拼音标注的(拼音一般标注在文字后面)

需按照原文版式,同时转换文字和拼音。

4.3.6 下划线

(1) 下划线标记的为单个字符或字母的

将标注内容填入【】,放在有下划线的字符或字母后。

(2) 下划线标注的为一句话或一段字符或字母的

将有下划线的内容放在()内,标注内容填入【】,放在有下划线的内容后。

4.3.7 其他

无法录入的生僻字、公式、符号等内容用“■”表示。同时将“■”所对应图像文件保存在档案典藏级数据内,以“■”命名的文件夹内。文件夹建立方法:

(1) 子目录内应包含所有用“■”表示的图像文件;

(2) 图像文件删除其他信息,仅保留“无法录入的内容”所在行的信息。图像命名不变。

(3) 多个用“■”表示的内容对应同一个图像时,仅保存一个图像文件。

5. 数据库命名及加工

5.1 数据库命名

对象数据所对应数据库。数据库名由8位组成:

文献基本资料类型(1位)、文献语种(1位)、加工年(2位)、机构代码(4位)。

本规范针对中文图书其文献基本资料类型为0,文献语种为1,加工年为公元年后2位数字。

例:01140101.mdb

5.2 数据库制作

5.2.1 对marc数据进行加工处理,制作文献基本信息表。

序号	中文名称	字段名称	对应书目数据(MARC)内容
1	加工编号	book_id	
2	分类	cat_id	第一个690字段\$a
3	书名	book_name	200字段\$a,\$h,\$i,\$e
4	作者	author	200字段\$f
5	出版社	pub_house	210字段\$c

6	出版时间	pub_date	210 字段\$d
7	ISBN 号	isbn	010 字段\$a
8	001	record_id	001 字段
9	条码号	barcode	
10	唯一标识符	cdoi	

注：文献基本信息表（book 表）除加工编号和条码号外，各字段内容原则上均取自书目数据（MARC）。对于相同“001”、不同“条码号”的图书，要进一步标注“书名”字段，标注内容用“（）”括起来。如：（上册）、（下册）。如果没有 MARC 数据，可人工添加。

5.2.2 目录信息著录，生成文献目录信息表。

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	序号	serial_num	
3	章节号	chapter_num	
4	章节名	chapter_name	
5	作者	author	
6	页码	page_num	客观著录，如实反映目录页原貌（可为空）
7	绝对页码	ppage_num	文件名数字部分
8	页位置	page_place	文件名字母部分
9	属性	page_prop	1) “目录” 属性为 “1”； 2) “无目录” 属性为 “2”； 3) 每册文献除第一条目录外，其余记录的属性默认为 “0”。

5.2.3 文献版权页著录，生成文献版权信息表。

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	书名	book_name	
3	作者	author	
4	001	record_id	

5	版权页位置	copyright_place	记录版权页文件名
---	-------	-----------------	----------

5.2.4 不带页号插图信息著录，生成文献插页信息表。

序号	中文名称	字段名称	备注
1	加工编号	book-id	
2	插页前正文页号	prior-text-page	文献印刷页码
3	插页数量	inset-num	

5.2.5 文献缺页信息著录，生成文献缺页信息表。

序号	中文名称	字段名称	备注
1	加工编号	book-id	
2	缺页前正文页号	start-text-page	文献印刷页码
3	缺页数	lostpage-num	

5.2.6 文献结构信息著录，生成文献结构信息表。

序号	中文名称	字段名称	备注
1	加工编号	book-id	
2	封面页数	fore-cover-num	
3	目录前，前附页数	preface1-num	
4	目录前，前附页起始页号	preface1-start-page	
5	目录页数	content-num	
6	目录起始页号	content-start-page	
7	目录后，前附页数	preface2-num	
8	目录后，前附页起始页号	preface2-start-page	
9	正文页数	text-num	
10	正文起始页号	text-start-page	
11	后附页数	appendix-num	
12	后附页起始页号	appendix-start-page	
13	封底页数	back-cover-num	

注：起始页号均为文献印刷页码

5.2.7 记录扫描分辨率、压缩因子、文件数量、存储量等信息，生成文献加工信息表。

序号	中文名称	字段名称	备注
1	加工编号	book_id	
2	书名	book_name	
3	扫描分辨率	dpi	
4	压缩因子	comp_factor	
5	灰度页数量	grey_num	
6	彩色页数量	col_num	
7	TIFF 数量	tiff_num	
8	PDF 数量	pdf_num	
9	TXT 数量	txt_num	
10	TIFF 存储量	tiff_mb	存储单位: MB
11	PDF 存储量	pdf_mb	存储单位: MB
12	TXT 存储量	txt _mb	
13	典藏级硬盘位置	hdA_place	典藏级硬盘号
14	双层 pdf 硬盘位置	hdB_place	
15	TXT 硬盘位置	hdC_place	

5.3 数据库制作要求

数据库的制作符合下列要求。

- (1) 文件格式为 mdb 格式。
- (2) 使用字符集为 UTF-8。
- (3) 著录信息应严格按照文献实际内容进行描述, 所著内容与对象文件应正确链接, 文字处理错误率不超过 0.3%。
- (4) 数据库字段、说明文件、各类表格等内容严格按照附件规定和样例版式。

6. 数据备份

内容包括 TIFF 图像、双层 PDF (单版和合并版)、TXT 文件 (单版和合并版)、对应数据库、说明文件、介质说明文件、文献的书目数据, 各一份。

各类型对象数据按加工编号递增顺序依次放置在硬盘内, 保存目录及文件名由四级组成:

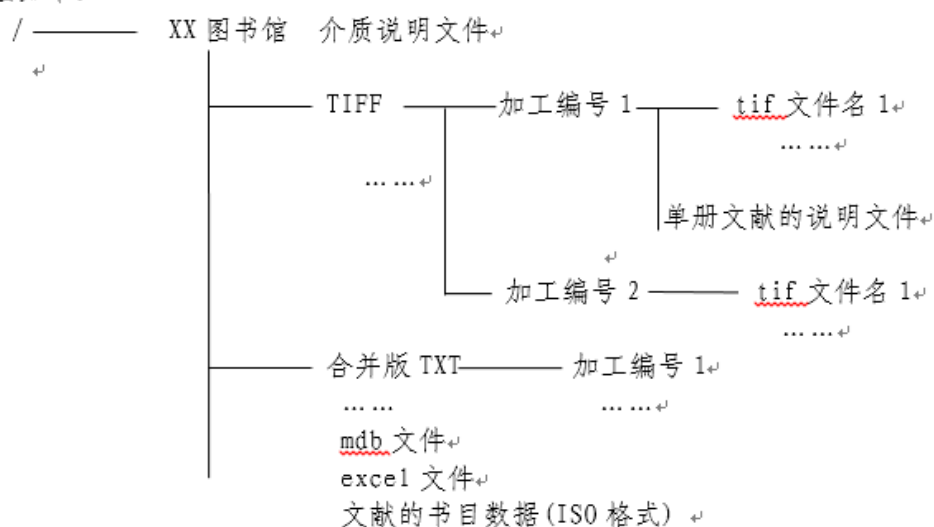
第一级目录为加工单位名称, 如首都图书馆, 同时将介质说明文件并行保存;

第二级目录为 TIFF\单版 PDF\合并版 PDF\单版 TXT\合并版 TXT，同时将对应数据库（mdb 文件）、说明文件（Excel 文件）及文献的书目数据（ISO 格式）并行保存；

第三级目录为加工编号；

第四级目录为相应格式的对象数据文件，TIFF 和单页 PDF 还需要相应的单册文献的说明文件。

框架图如下：



7. 说明文件

7.1 制作文献的说明文件

说明文件格式：Microsoft Excel

说明文件命名：地方文献数据说明

说明文件内各表命名及结构说明：

1. 数据总体说明

序号	名称	说明
1	项目名称	项目名称
2	承建馆名称	承建馆名称全名
3	质检单位	项目第三方质检单位全名
4	项目时间	项目完成年份
5	数据加工编号	以起始～终止表示，如 01140101001～01140101030

6	完成文献总册数	项目实际完成文献册数
7	完成文献总页数	项目实际完成图像页数
8	序号	以 1 开始，顺次编号
9	存储内容	存储的数据内容，如扫描图像、数据库、说明文件等
10	数据格式	以大写英文字母表示，如 TIFF、PDF、TXT、EXCEL、ACCESS
11	存储介质	数据存储的介质，如硬盘
12	介质编号	硬盘填写硬盘编号，如 01140101
13	介质数量	存储介质的数量
14	存储量	以 GB 为单位表示
15	备注	需要说明的特殊情况

2. 文献单册数据量统计表

序号	名称	说明
1	序号	以 1 开始，顺次编号
2	加工编号	文献加工编号
3	灰度页	本册文献中按灰度方式扫描的图像数量
4	彩色页	本册文献中按彩色方式扫描的图像数量
5	图像合计	灰度页与彩色页数量之和
6	目录条数	本册文献录信息表（catalog 表）条目数
7	合计	文献加工编号以起止号表示，合计灰度页数、彩色页数、图像总数、目录总条数

注：文献加工编号按递增顺序排列，填写保存级对象数据相关信息

3. 文献全文转换未转换文件记录表

序号	名称	说明
1	序号	以 1 开始，顺次编号

2	加工编号	文献加工编号
3	未转换图像文件名	
4	对应单版文本文件名	
5	未转换原因	
6	备注	

注：文献加工编号按递增顺序排列

4. 全文转换加工文字量统计表

序号	名称	说明
1	序号	以 1 开始，顺次编号
2	加工编号	文献加工编号
3	文献名称	
4	文字量	
5	转换后文字量（千字）	
6	备注	

7.2 制作每册文献和保存数据介质的说明文件

每册文献的说明文件（bookinfo.txt）

book_id=01140101-001

book_name=xxxxxx

author=xxxxxxxx

file_number=519

说明：

book-id: 文献加工编号

book-name: 书名

author: 作者

file-number: 总文件数

存储介质的说明文件 (readme.txt)

一、文献类型：例如，中文图书

二、制作时间：2014 年 月

三、承建馆：

四、加工单位：

五、存储介质信息

1、介质名称：

2、加工起止号：例如，01140101-001 ~ 01140101-099

3、文献数量：

4、文件数量：

5、存储容量：

六、技术参数

1、存储格式：例如，不压缩 tif 格式

2、加工设备：

3、加工软件：

4、参数指标

 扫描分辨率：例如，300dpi

 扫描方式：

8. 存储介质命名及要求

8.1 硬盘命名

由 8 位组成：

文献基本资料类型（1 位）、文献语种（1 位）、加工年（2 位）、机构代码（4 位）

8.2 存储介质要求

介质中不得存放与备份内容无关的文件、严禁携带病毒、严禁浪费介质空间。

附录 B

地方文献验收数据提交单

提交时间		
承建馆		
接收单位		
介质	类型 (编号)	
	数量	
文献加工 起止号		
内容	文献数量	
	文件数量	
	数据库	
	其它	
备注		

附录 C

地方文献成品数据移交单

移交时间		
承建馆		
提交人员签字		
接收单位		
接收人员签字		
介质	类型 (编号)	
	数量	
	存储量	
文献加工起止号		
内容	文献数量	
	文件数量	
	数据库	
	其它	
备注		

附录 D

地方文献数字化机构代码

序号	机构名称	代码
1	首都图书馆	0100
2	天津图书馆	0200
3	河北省图书馆	0300
4	山西省图书馆	0400
5	内蒙古图书馆	0500
6	辽宁省图书馆	0600
7	吉林省图书馆	0700
8	黑龙江省图书馆	0800
9	上海图书馆	0900
10	南京图书馆	1000
11	浙江图书馆	1100
12	安徽省图书馆	1200
13	福建省图书馆	1300
14	江西省图书馆	1400
15	山东省图书馆	1500
16	河南省图书馆	1600
17	湖北省图书馆	1700
18	湖南图书馆	1800
19	广东省立中山图书馆	1900
20	海南省图书馆	2000
21	广西壮族自治区图书馆	2100
22	广西壮族自治区桂林图书馆	2200
23	重庆图书馆	2300
24	四川省图书馆	2400
25	贵州省图书馆	2500
26	云南省图书馆	2600
27	西藏自治区图书馆	2700
28	陕西省图书馆	2800
29	甘肃省图书馆	2900
30	青海省图书馆	3000
31	宁夏回族自治区图书馆	3100
32	新疆维吾尔自治区图书馆	3200
33	新疆生产建设兵团文化中心	3300

附件 5

推广工程数字资源联合建设 政府信息公开项目建设方案（2014）

为加强我国公共图书馆的政府信息服务能力，促进全国公共图书馆政府信息服务工作的交流合作与资源共享，建立政府信息资源组织整理的相关业务规范和工作指南，推广工程联合全国各省市级公共图书馆共同开展全国各地政府信息的整合与服务。各级图书馆利用中国政府公开信息整合服务平台进行所在行政区域的政府信息加工与服务。

一、建设内容

政府信息公开项目整合的内容主要包括政府网站上的政府信息公开专栏、政府公报和其他栏目下的公文信息。

政府信息公开专栏：指各政府机构在其网站上专门设置的信息公开专栏下的信息。

政府公报：是指政府机关出版发行的以登载法令、方针、政策、宣言、声明、人事任免等各类政府文件为主要内容的连续出版物。一般政府门户网站中都设有“政府公报”专栏。

其他公文信息：政府网站其他栏目中的公文类信息，主要包括通知公告、法律法规、统计信息、财政信息、应急管理、总结规划等。机构新闻、调查研究、案例点评、文学创作等均不在收录之列。

二、工作流程

1、采集准备

按照政府信息公开项目建设内容，对符合收录要求的政府信息进行整理。将拟采集的政府信息网址整理成 Excel 文档，提交至国家图书馆，经审核确认后形成采集列表。

在采集数据前，要对拟采集的政府信息进行查重，本项目已建成的资源不再重复收录。

2、数据制作

数据制作流程根据数据采集方式不同分为两种，各机构可根据需求选择适合的数据制作方式。

（1）自动采集

采集：利用网络采集工具，按照采集列表和预先设定的采集规则，自动采集各政府网站的页面信息。

导入：将采集到的政府信息导入到本项目的后台管理系统。选择自动采集的机构需要事先通知国家图书馆，并提交样例数据，以便实现系统对接。

编目：按照“政府信息公开元数据著录规则”（见附录）对采集到的政府信息进行编目。如果系统在数据采集中实现了自动编目，则需对自动编目的内容进行校对和个别字段内容的补充。

（2）人工采集

基于政府信息公开后台管理系统，根据采集列表人工逐条录入政府信息。在信息录入的同时，要根据“著录规则”完成信息的编

目。

3、数据审校

对完成编目的数据要按照“著录规则”进行审校，保证各字段的准确、完整、命名规范，真实反映政府信息原貌。

4、数据发布

数据经审校无误后即可发布。“中国政府公开信息服务平台网站”是政府信息公开项目的门户网站，提供所建政府信息的浏览、检索和集中展示。

5、数据维护

对本机构负责制作及发布的元数据信息进行长期维护，保障数据准确无误、正常显示并及时更新。

三、成果提交

- 1、存储并发布于政府信息公开整合服务平台的数据。
- 2、第三方数据质检报告。

附录：政府信息公开元数据著录规则

附录

政府信息公开元数据著录规则

一、数据质量总体要求

- 1、所有数据需与政府信息相关，不相关则视为无效数据。
- 2、采集到的数据内容及版式与发布网站一致，图片快照正常显示，不出现乱码，完整采集附件，原文链接准确。
- 3、元数据完整、准确、命名规范、不得出现错别字，数据组织合理，严格按照元数据标准规范进行各元素的著录。著录部分应对资源进行客观描述，真实反映原貌；标引部分应完整正确地揭示资源内容。
- 4、错误率不超过 1%

二、元数据标准

本项目中资源的著录对象是单条政府信息，政府公报也以其中的单条信息为著录对象，与网站上其他的政府信息有所区别的是“政府公报”的信息需要在元数据中标明“出处”，著录刊载单条信息的公报期数。

元素		必备性	解释
题名		必备	政府信息的正题名。按政府信息中出现的形式著录。
	其他题名	有则必备	对正题名起补充解释作用的题名。
题材分类		必备	按照政府信息的类型及规范题材分类表进行标引。元素取值见题材分类表。
主题分类		必备	按照政府信息的内容及规范主题分类表进行标引。元素取值见主题分类表。
文种		公文类必备	著录公文的种类，非公文不必著录。元素取值见文种分类表。

关键词		必备	著录规范的电子政务主题词或关键词。
文号		有则必备	政府信息的文号。非公文不必著录。
日期	来源网站 发布日期	必备	政府信息在来源网站上的发布时间。
	发文日期	公文类 有则必备	政府信息自身的发文时间。著录信息源为正文。 非公文不必著录。
	实施日期	有则必备	政府信息的实施时间。著录信息源为正文。
	废止日期	有则必备	政府信息的废止时间。著录信息源通常为该政府信息变更后的新信息的正文。
原标识符		有则必备	政府信息的原始标识符。
发布机构		有则必备	发布政府信息的机构。著录机构全称。
发文机构		公文类 有则必备	政府信息自身的发布机构。著录机构全称。著录信息源为正文。非公文不必著录。
转载来源		有则必备	著录政府信息转载自某信息来源。
信息来源		必备	著录该政府信息的来源网站。
原文地址		必备	著录该政府信息的发布地址（URL）。
附件		有则必备	资源内容以文件形式独立存在，可以是信息内容的部分或全部。
快照		必备	快照两种格式，一种图片格式，一种网页格式。
正文		必备	著录该政府信息的全文。要求与原网站保持内容及版式完全一致。
公开方式		有则必备	该政府信息是否主动公开，通常为主动公开。
公开范围		有则必备	该政府信息的公开范围，通常为面向社会。
时效性		有则必备	该政府信息的有效期，通常为长期有效。
修改日期		必备	著录信息的最后修改日期。可由系统自动提取。
出处		政府公报必备	著录刊载单条信息的公报期数。

注释：公文是党政机关实施领导、履行职能、处理公务的具有特定效力和规范体式的文书，是传达贯彻党和国家的方针政策，公布法规和规章，指导、布置和商洽工作，请示和答复问题，报告、通报和交流情况等的重要工具。公文一般由份号、密级和保密期限、紧急程度、发文机关标志、发文字号、签发人、标题、主送机关、正文、附件说明、发文机

关署名、成文日期、印章、附注、附件、抄送机关、印发机关和印发日期、页码等组成。

摘自《党政机关公文处理工作条例》，网址为：

http://www.gov.cn/zwgk/2013-02/22/content_2337704.htm

三、分类表

1、题材分类

一级分类	二级分类
政策法规	法律
	行政法规
	地方法规
	规章
	司法解释
	其他
机构文件	发展规划
	财政预算
	工作报告
	计划总结
	招投标信息
	奖励处罚
	人事任免
	其他
政府公报	
工作动态	
统计信息	
行政职权	行政处罚
	行政许可
	行政确认
	行政强制

	行政给付
	其他
其他信息	

2、主题分类

主题分类参照国务院办公厅政府信息公开栏目所使用的主题分类设置（含 21 个一级分类及 110 个二级分类）。

序号	一级分类	二级分类
1	组织机构	
	综合政务	政务督查
		应急管理
		电子政务
		文秘工作
		保密工作
		信访
		参事、文史
		其他
3	国民经济管理、国有资产监管	宏观经济
		经济体制改革
		统计
		物价
		国有资产监管
		其他
		重大建设项目
4	财政、金融、审计	财政
		税务
		银行

		货币（含外汇）
		证券
		保险
		社会信用体系建设
		审计
		其他
5	国土资源、能源	土地
		矿产
		水资源
		海洋
		煤炭
		石油与天然气
		电力
		其他
6	农业、林业、水利	农业、畜牧业、渔业
		林业
		水利
		其他
7	工业、交通	机械制造与重工业
		轻工纺织
		化工
		国防工业
		航天、航空
		信息产业（含电信）
		公路
		水运
		铁路

		民航
		邮政
		其他
8	商贸、海关、旅游	对外经贸合作
		国内贸易（含供销）
		海关
		检验、检疫
		旅游
		其他
9	市场监管、安全生产监管	工商
		质量监督
		标准
		食品药品监管
		安全生产监管
		其他
10	城乡建设、环境保护	城市规划
		城乡建设（含住房）
		环境监测、保护与治理
		节能与资源综合利用
		气象、水文、测绘、地震
		其他
11	科技、教育	科技
		教育
		知识产权
		其他
12	文化、广电、新闻出版	文化
		文物

		新闻出版
		广播、电影、电视
		其他
13	卫生、体育	卫生
		医药管理
		体育
		其他
14	人口与计划生育、妇女儿童工作	人口与计划生育
		妇女儿童
		其他
15	劳动、人事、监察	劳动就业
		社会保障
		人事工作
		军转安置
		监察
		纠正行业不正之风
		其他
16	公安、安全、司法	公安
		国家安全
		司法
		其他
17	民政、扶贫、救灾	减灾救济
		优抚安置
		社会福利
		行政区划与地名
		社团管理
		扶贫

		其他
18	民族、宗教	民族事务
		宗教事务
19	对外事务	外交、外事
		国际条约、国际组织
		其他
20	港澳台侨工作	港澳工作
		对台工作
		侨务工作
21	国防	国防建设
		国防动员
22	其他	

来自中央政府网，网址为 <http://www.gov.cn/xxgk/pub/govpublic/>

3、文种分类表

文种分类按照中共中央办公厅、国务院办公厅颁发的《党政机关公文处理工作条例》设置。

决议	决定	命令（令）	公报	公告	通告	意见	通知
通报	报告	请示	批复	议案	函	纪要	

来自中央政府网，网址为：

http://www.gov.cn/zwgk/2013-02/22/content_2337704.htm

附件 6

推广工程数字资源联合建设 网事典藏项目建设方案（2014）

互联网上有大量有关人类社会、政治、历史、文化的信息，对于传承人类文化遗产和学术研究具有非常重要的作用，但是由于互联网资源具有易失性，每天都有海量有价值的信息在消亡。为保存具有重要价值的互联网资源，推广工程将开展网络资源采集与长期保存工作。

一、建设内容

网事典藏项目建设初期先以政府网站的采集和存档为重点，主要采集反映地方政治、经济、文化发展等信息的政府网站或重要的事业单位网站（包括但不限于.gov.cn 域名内的网站），并将采集到的网站进行编目和发布。各馆原则上只采集行政上从属于本地区的政府和事业单位网站。

二、工作流程

1、采集准备

按照网事典藏项目建设内容，对符合收录要求的政府及事业单位网站进行全面整理，确定采集范围。将需要采集的政府网站网址（URL 地址）整理成采集列表（excel 表格），提交国家图书馆审核，审核通过作为正式采集列表进行采集。

2、资源采集

根据采集列表，利用网络采集软件，对政府网站进行全面采集，要求所采集的文件包含采集列表中政府网站域名内的全部内容，但不包括论坛等需链接后台数据库的内容。所采集的文档格式遵循 WARC1.0 标准，不含病毒、垃圾文件及采集列表外的其他信息。每个网站单独采集，采集结果以文件夹形式存储，文件夹名称以对应的加工编号（加工编号见备注）来命名。

3、加工编目

按照“政府及相关机构存档网站元数据著录规则”对采集到的政府及事业网站进行编目，每个采集结果对应一条完整的编目信息。将编目信息制作成 excel 表（该表中访问地址一项需要在数据发布后才能取得补齐）。需要在唯一标识符系统中注册 CDOI。

4、数据预发布

将采集到的文档（WARC 文档）数据进行索引后预发布，取得所采集政府网站的访问地址，在编目数据中补齐访问地址字段的内容。

发布方式由各馆根据实际情况自行选择。

5、数据审校

（1）编目数据审校：对编目完整的数据按照“著录规则”进行审校，保证各字段的准确、完整。

（2）对象数据审校：在预发布页面通过点击的方式进行查验，保证页面内容都能正常打开，且与原网站保持一致。文件夹命名无误，与编目数据一一对应。

(3) 数据本馆审校合格后交第三方进行验收，验收不合格需要修改或重采，直到验收合格。

6、数据发布

验收合格后，将最终成果进行正式发布，并提供用户服务。同时将编目数据、对象数据、统计信息表（见下文）最终版统一提交至国家图书馆。

7、数据维护和长期保存

各馆负责对本机构制作及发布的信息及其发布网站进行长期维护，保障数据准确无误，显示正常，同时做好数据备份与长期保存工作。

三、成果提交

1、以移动硬盘方式（数据迁移后硬盘返还提交馆）提交的以下三个方面的数据：

(1) 编目数据：政府网站归档资源的编目数据以 excel 表格方式提交。

(2) 对象数据：政府网站归档资源的对象数据以文件夹（文件夹名以加工编号命名）的形式提交网络采集软件所采集的全部采集结果。

(3) 统计信息表：统计信息表包括所采集网站名称、网址、运行时间、url 数量、压缩前后数据量等信息，详见表一。

表一：网络采集统计信息表

加工 编号	网站 名称	网址	任务名	压缩前 大 小 (GB)	压缩后 大 小 (GB)	URL 数量	运行 时间	采集 日期

2、第三方质检报告。

附录: 政府及相关机构存档网站元数据著录规则

附录

政府及相关机构存档网站元数据著录规则

1. 著录对象

著录对象为存档政府和事业单位网站（简称政府网站）。以具体的存档政府网站对象为一个著录单位。

如果一个政府网站具有多个主页域名，著录时作为一个对象著录，多个域名作为子项著录。

2. 著录要求

对采集到的网站进行编目加工，要求参照相关元数据规范和著录规则进行编目，且准确无误。编目数据以 EXCEL 文件形式提交，对编目数据项的要求如下：

政府及相关机构存档网站基本字段及著录要求

元素	修饰词	元素的著录内容	元素必备性、可重复性
标识符	加工编号	给予资源的一个明确标识，可以唯一标识元数据的标识符。具体规则见备注	必备，不可重复
	CD01	唯一标识符号	必备，不可重复
网站名称	正题名	著录网站名称。信息取自网站页面首页源代码中的 title。例如：“北京市朝阳区人民政府网站”的 title 是“北京朝阳”，将“北京朝阳”著录为网站名称。若 title 为空，或不反映网站内容，可用 banner 或其他明显反应网站内容的名称。	必备，不可重复
	交替名称	按照统一要求著录“**网站”作为交替名称，对网站名称进行解释说明。例如：朝阳区人民政府的网站统一著录交替名称为：“北京市朝阳区人民政府网站”。	必备，可重复
描述		对网站内容的任何说明	必备，可重复

类型		所采集的政府网站类型：政府机构网站、事业单位网站	必备，不可重复
语种		资源所使用的语种	必备，可重复
格式	文件格式	所采集的网站资源存档格式。在本规范中为：WARC	必备，不可重复
	文件数量	著录WARC文件的数量	必备，不可重复
	文件大小	著录所有WARC文件的总容量	必备，不可重复
机构	机构名称	著录网站的所属机构的名称。例如：北京市朝阳区人民政府网站的所属机构是：北京市朝阳区人民政府	必备，不可重复
	交替名称	著录除正式机构名称以外的翻译名称、缩写名称、统一名称、并列名称等。例如：中华人民共和国外交部 其交替名称为：外交部	有则必备，可重复
域名		著录网站的主页域名，如果一个网站有多个域名，则全部著录。多个域名重复本元素。	必备，可重复
采集地址		政府网站的网址	必备，不可重复
日期	采集日期	资源被采集的日期	必备，不可重复
访问地址		所采集的政府网站存档访问网址	必备，不可重复
权限		与本资源有关的权利和许可的声明	有则必备，可重复
	版权信息	资源所有权信息	有则必备，可重复
数据提交单位			必备，不可重复
数据提交日期			必备，不可重复

备注：政府网站采集加工编号命名规则

1、政府网站加工编号由 17 位组成：GOV（3 位）+行政代码（6 位）+政府网站级别（2 位）+机构类型代码（2 位）+流水号（4 位 0001-9999）

2、说明:

(1) 行政代码: 各地方馆所属的行政代码, 例如河北省图书馆, 其所属河北省的行政代码为 130000

(2) 政府网站级别: 中央级(00)、省级(01)、地级(02)、市级(03)、县级(04)、乡级(05)。

(3) 机构类型代码: 政府机构(01)、事业单位(02)

附件 7

推广工程数字资源联合建设
项目质检报告

年 月 日

项目名称			
建设单位		建设数量	
质检单位		抽检数量	
提交时间		抽检时间	
质检结果			
<div>一、总体说明</div> <div>二、质检过程</div> <div>三、主要问题</div> <div>四、检查明细（另附文）</div> <div>质检单位</div> <div>（ 签章 ）</div>			