



中国国家图书馆 · 中国国家数字图书馆
NATIONAL LIBRARY OF CHINA NATIONAL DIGITAL LIBRARY OF CHINA

推广工程数字资源云发布系统建设

国家图书馆 信息技术部

2018.10





一、建设背景

数字资源生命周期

生产

数字资源生产包括传统资源数字化、网络资源建设、特色资源库建设等

组织

将数字资源从无序变成有序，通过对资源整理、加工、引导、揭示、控制等一系列组织化过程及方法，实现知识组织的自动化。

发布

将组织整理好的元数据和对象数据，通过数据抽取、数据转换、数据封装、数据索引等处理手段，发布到网站上，提供读者服务。

保存

对图书馆所拥有和使用的数字形态的资源进行有效保存，能够长期存储和长期可获取。

数字图书馆推广工程资源联建项目启动以来，全国各级图书馆充分结合地方特色，依托联建项目建设了一批内容丰富、形式多样的优秀数字资源，内容丰富、形式多样。



云计算、大数据、关联数据以及移动互联网等新技术的快速发展和成功应用为国家数字图书馆特色资源云平台的建设提供了技术保障。



《国家“十三五”时期文化发展改革规划纲要》指出要“运用云计算、人工智能、物联网等科技成果，催生新型文化业态”；

《文化部“十三五”时期文化产业发展规划》也指出要“大力培育基于大数据、云计算、物联网、人工智能等新技术的新型文化业态，形成文化产业新的增长点”。



推广工程数字资源云发布系统实现对国家图书馆自建资源、征集资源以及推广工程联建资源等数据的统一管理及发布。未来，将建设图书、期刊、报纸、论文、地方志、图片、民国期刊、民国法律、音频、视频讲座等多个专题库面向读者提供服务。





二、架构设计

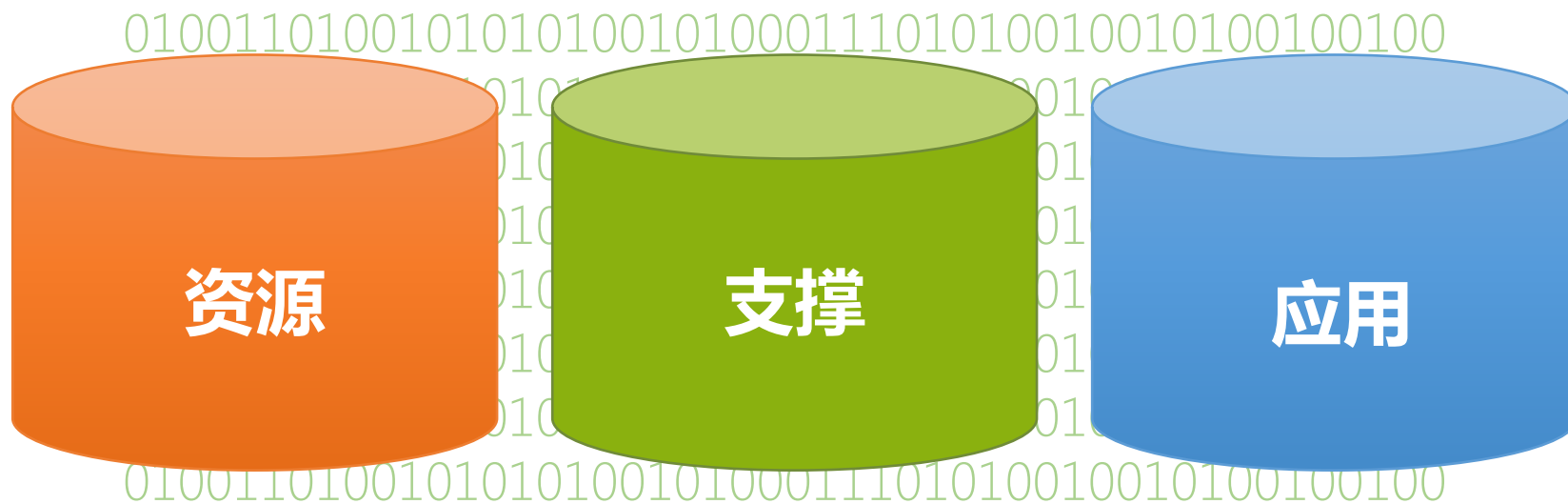
(一) 设计思路

- 支持数据异地发布
- 开放云搜索引擎
- 集成第三方开发控件（PDF和流媒体）
- 成熟的开发框架及基础功能
- 对接统一用户管理系统、唯一标识符系统

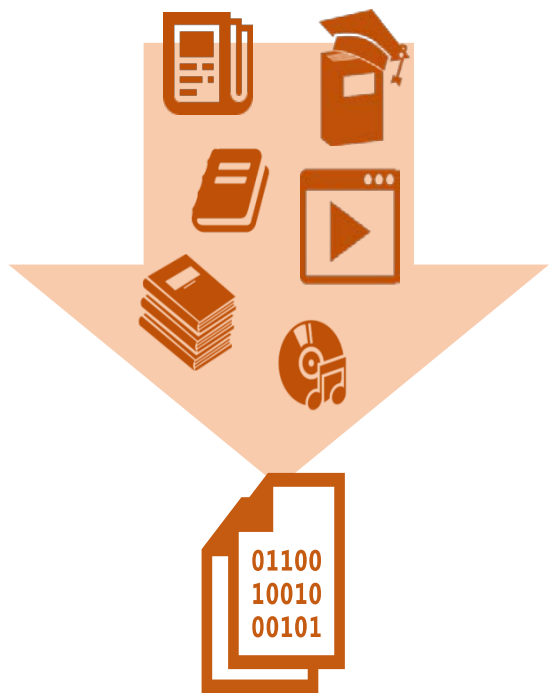


(二) 应用架构

突出文献服务和资源整合，将整个项目设计成“资源+支撑+应用”的体系结构。



资源



- ◆ 对推广工程各图书馆提供的数据进行查重、清洗、转换（格式、字符集）处理，最终形成统一格式元数据。

支撑



- ◆ 统一用户管理系统；唯一标识符系统；WebPDF插件、流媒体平台。

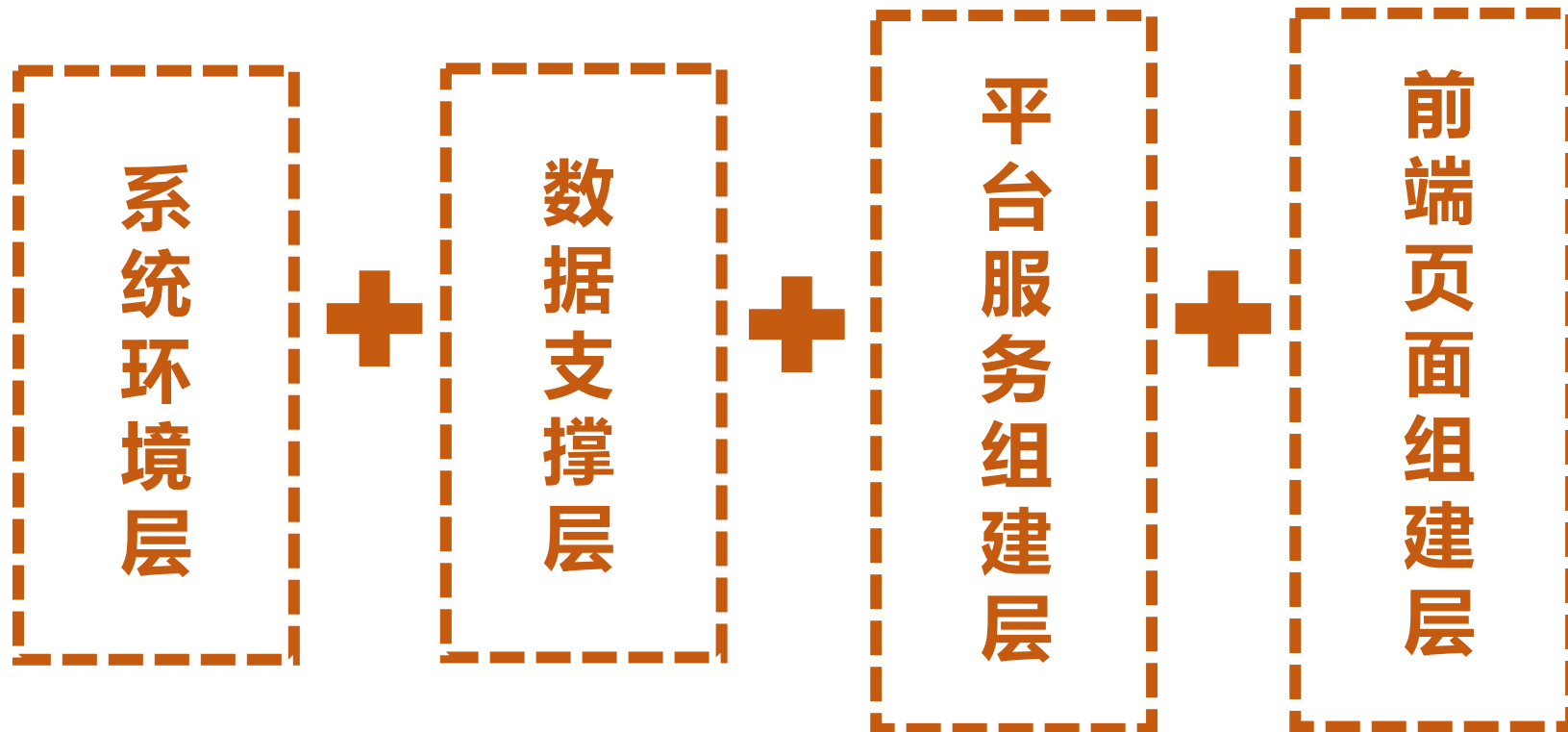
应用



- ◆ 提供用户的实际应用业务，具体包括数据发布管理模块和简化版前台门户展示模块。

(三) 技术架构

- 平台基于成熟的开发平台框架，在使用已有功能组件的基础上，开发实现业务功能组件。



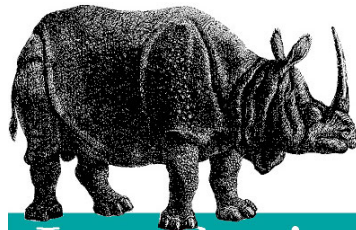
系统环境层



HTML



CSS



JavaScript

系统开发环境采用当前流行开发语言 Java，前端页面采用 Javascript 脚本语言 + CSS/HTML 表单样式。

数据支撑层



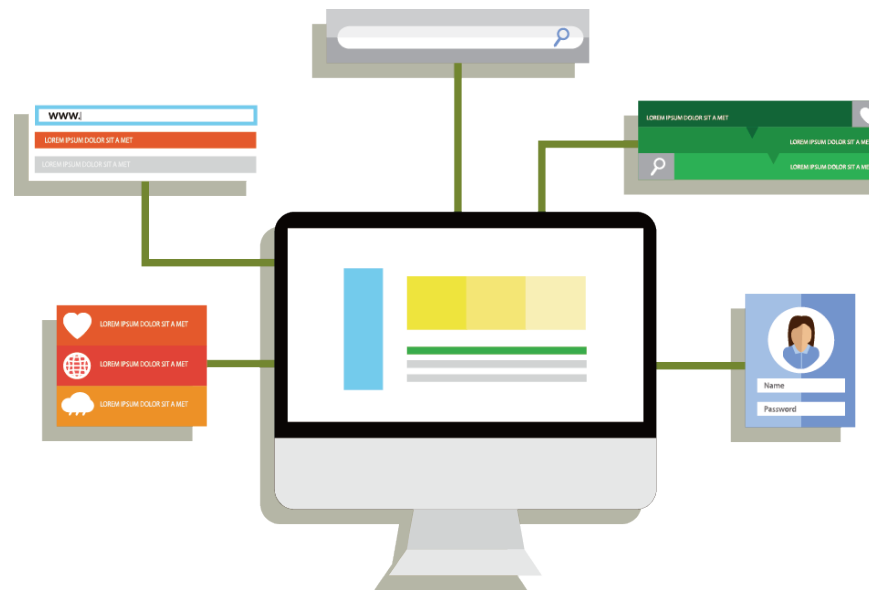
系统采用 MySQL 数据库，实现用户系统数据及业务数据的物理存储；使用 WEB 页面的数据采集工具实现数据采集。

平台服务组件层



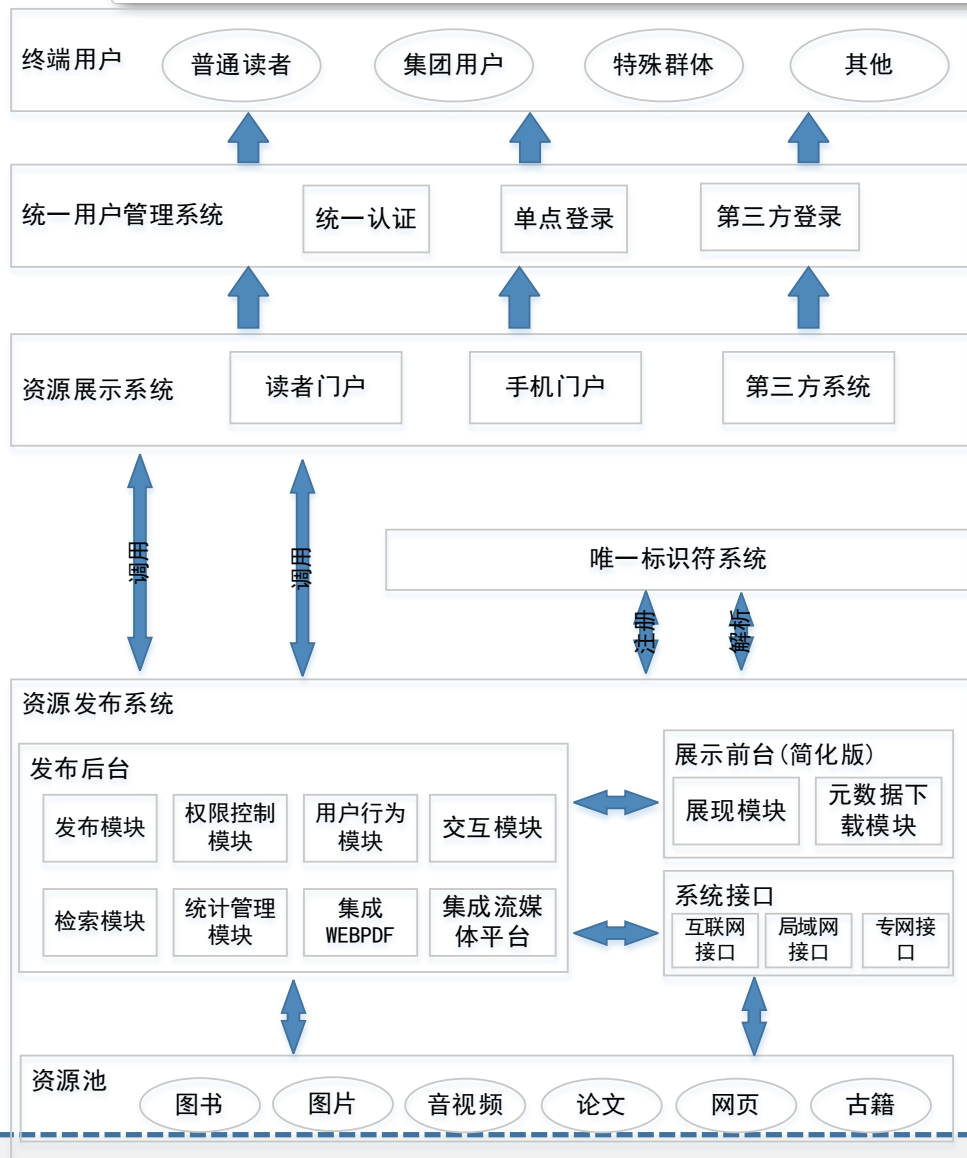
平台服务组件包括开发平台自身的通用服务组件和根据系统需求开发的业务服务组件，如：系统管理、用户管理、角色管理、内容管理、在线阅读（WEBPDF）等服务组件。

前端页面组件层



前端页面服务组件是系统提供给用户使用的系统功能，包括统一检索、资源服务、数据服务。

(四) 系统功能架构



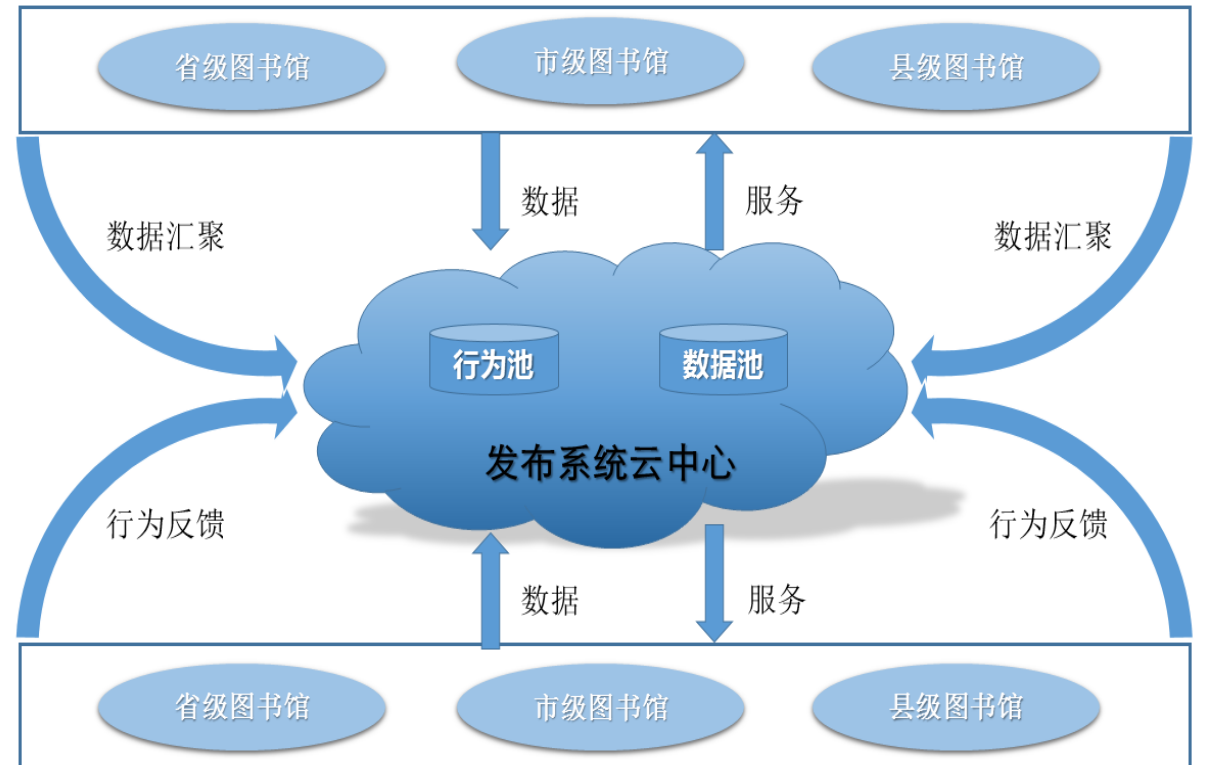
云发布系统立足于推广工程，为各省、市级图书馆提供数字资源发布支撑平台，提升公共图书馆数字资源的发布和服务能力。

- 1、建设推广工程数字资源池
- 2、建设推广工程用户行为池

(五) 服务模式

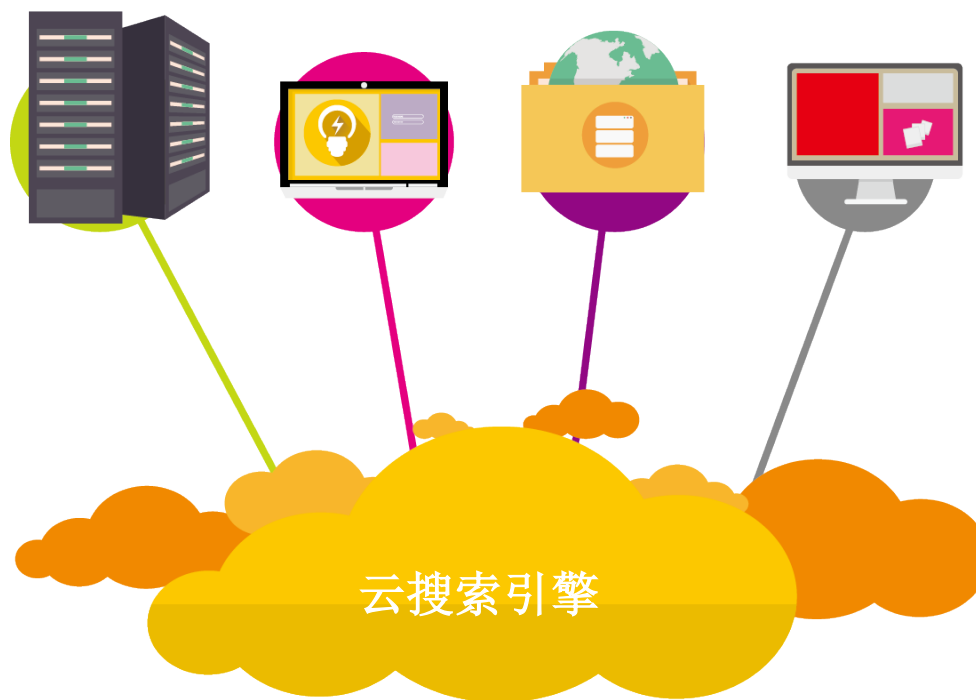
系统采用**中心--地方云架构模式**，各地方图书馆采用“软件租用”的方式实现数字资源云发布，发布系统云中心将数字资源以统一标准进行封装，对外提供统一的数据访问服务。

云发布系统降低了数字资源发布的技术门槛，各地方馆不需要进行系统建设，便可借助系统实现各自资源的云发布，从而更好地为广大读者提供服务。例如，各图书馆可以通过专网、互联网等多种渠道，采用FTP或HTTP协议，将元数据上传到云中心，中心对元数据进行统一管理，对象数据存在本地或者是上传到云发布系统中心，便可实现资源的发布。



(六) 部署架构

Web服务器、应用服务器、元数据库服务器、云搜索引擎等各个环节，均支持采用**集群服务方式**。



基于分布式多用户的全文搜索引擎,采用**RESTful web接口**来支撑云发布系统面向互联网大用户量、大并发量客户的检索和分析应用。整个系统具备高度的**开放性和可扩展性**，可以根据业务拓展和服务负载情况进行按需扩展。

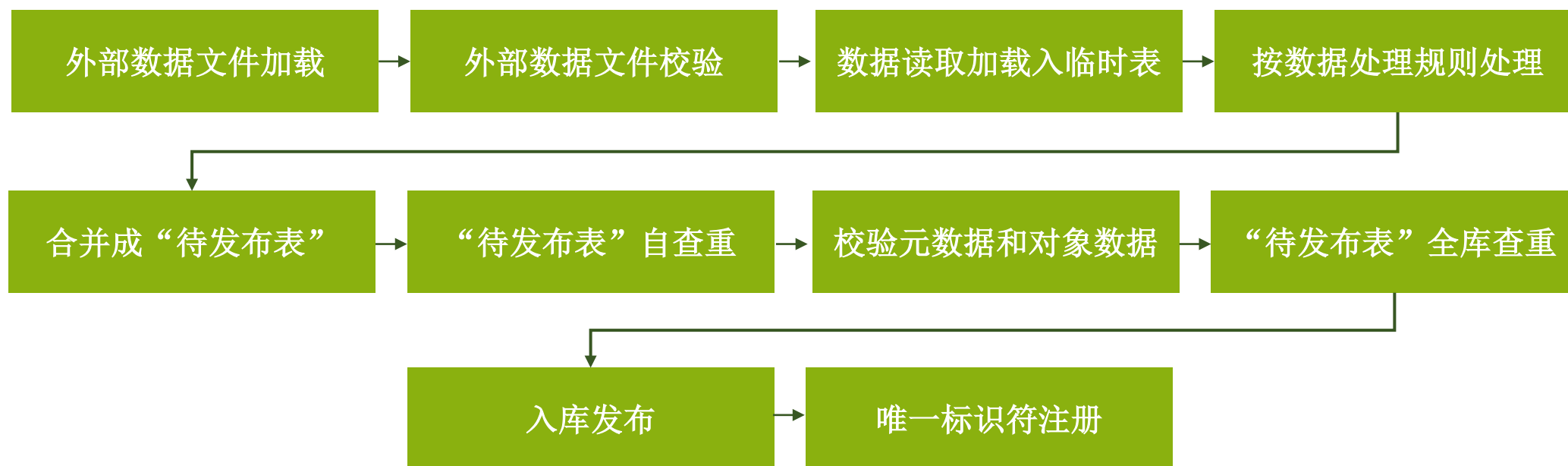




三、功能设计

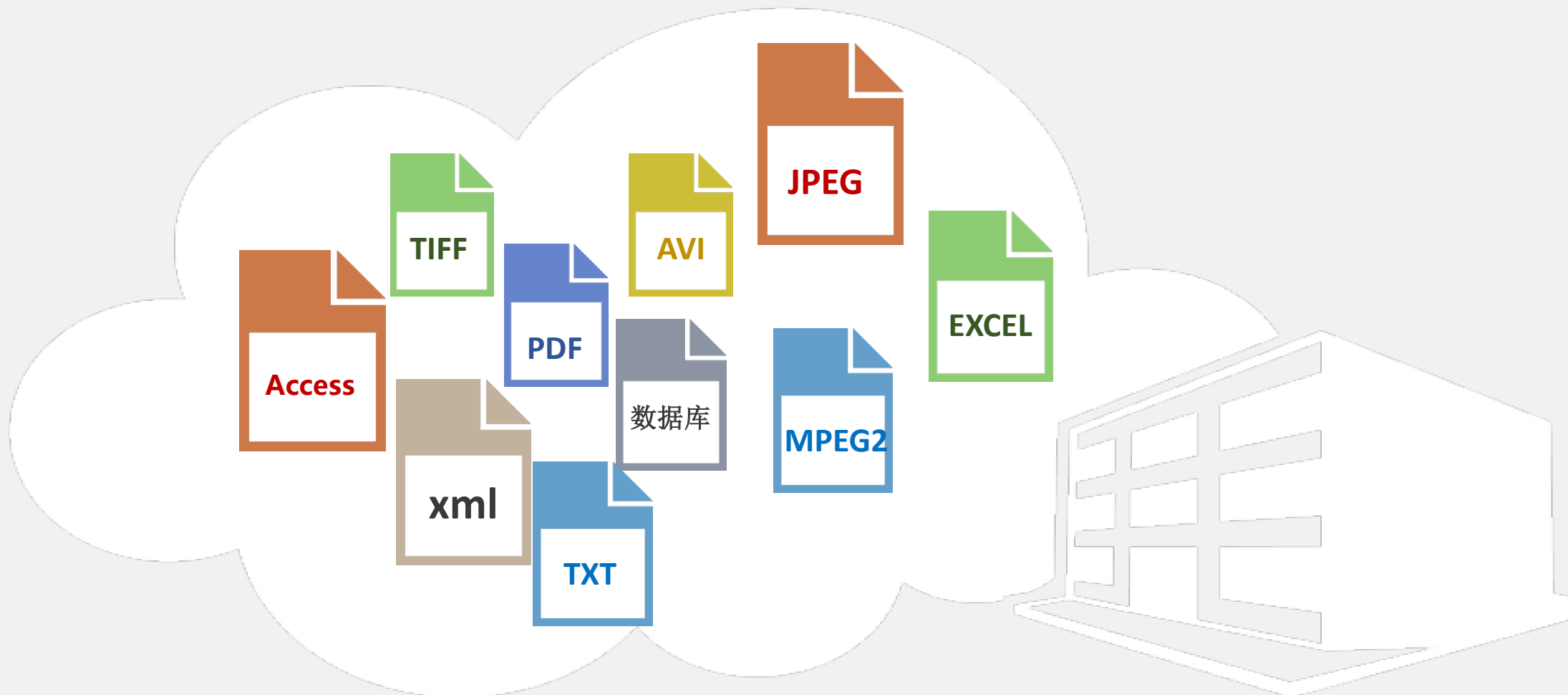
(一) 数据加载和处理流程设计

数据加载和处理流程



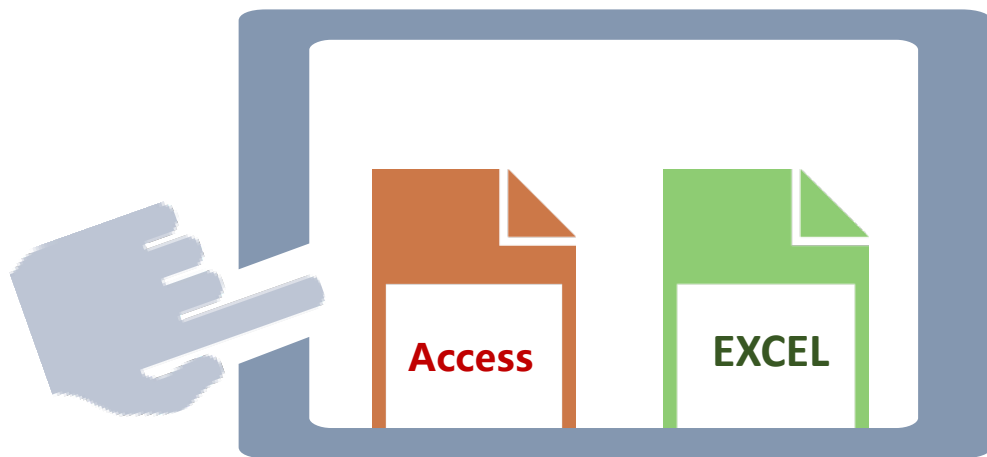
1.外部数据文件加载

外部数据文件加载的含义：要处理的元数据和对象数据是多种多样的，其中元数据文件包括**Access文件**，**EXCEL文件**、**数据库文件**、**文本文件**、**xml文件**等多种格式的文件。

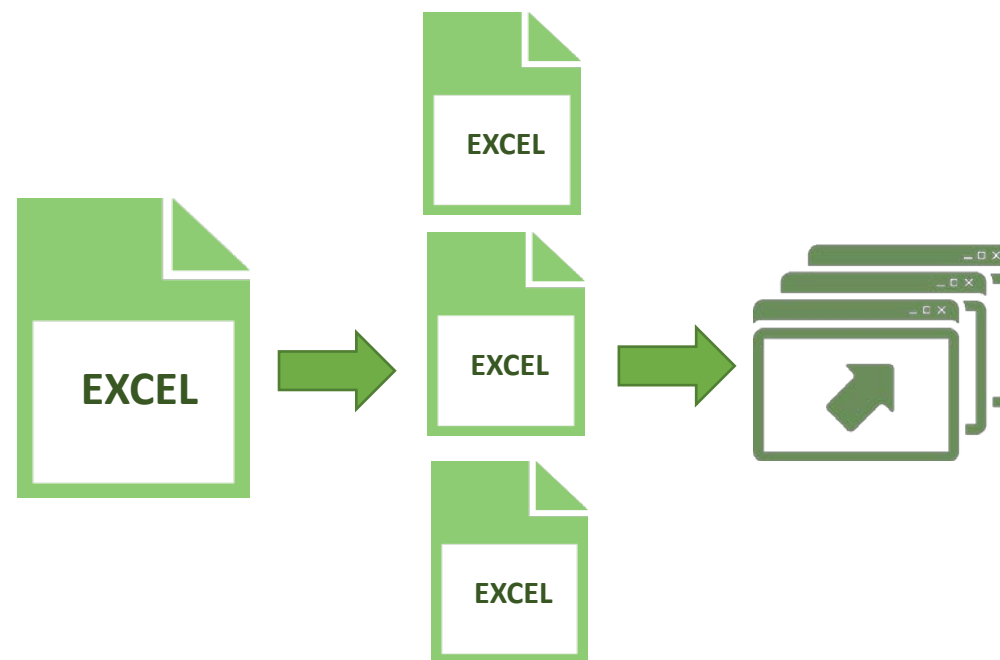


系统要针对不同类型的文件进行加载以读出该文件中的数据。系统在加载文件前**首先要预定义好被加载文件的数据格式和数据类型等。**

◆ 加载文件类型（Access、EXCEL等）

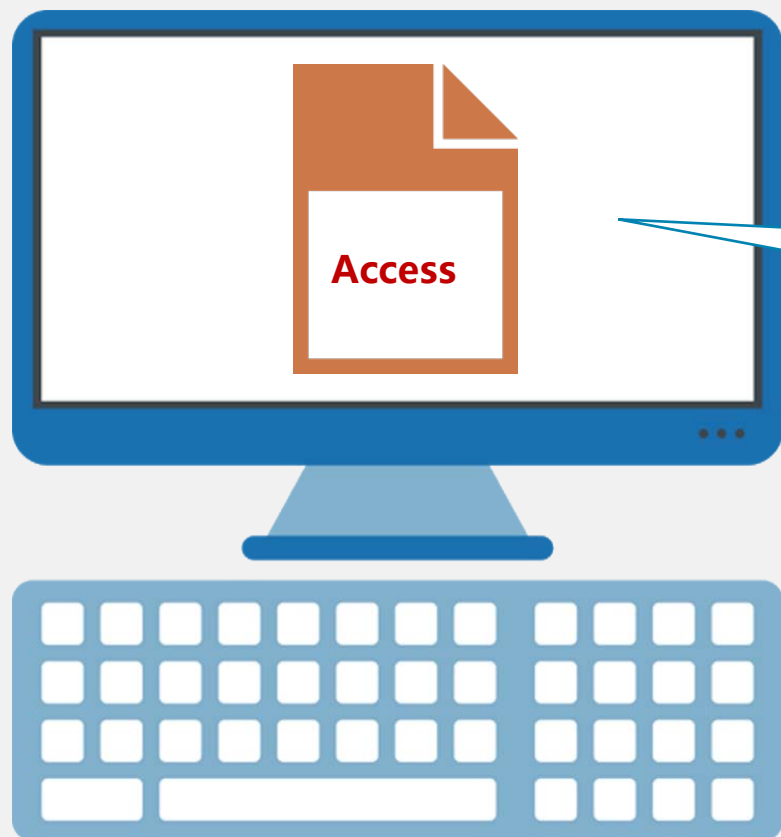


◆ 读取数据文件



2.外部数据文件校验

对于外部数据文件来说，会针对不同的数据来源定义好不同的数据校验格式。



◆ 不同图书馆不同专题库ACCESS
文件字段结构不一样

- ◆ 选择校验规则，如字段名称的匹配，字段个数匹配，字段完整性匹配等



- ◆ 校验规则自定义



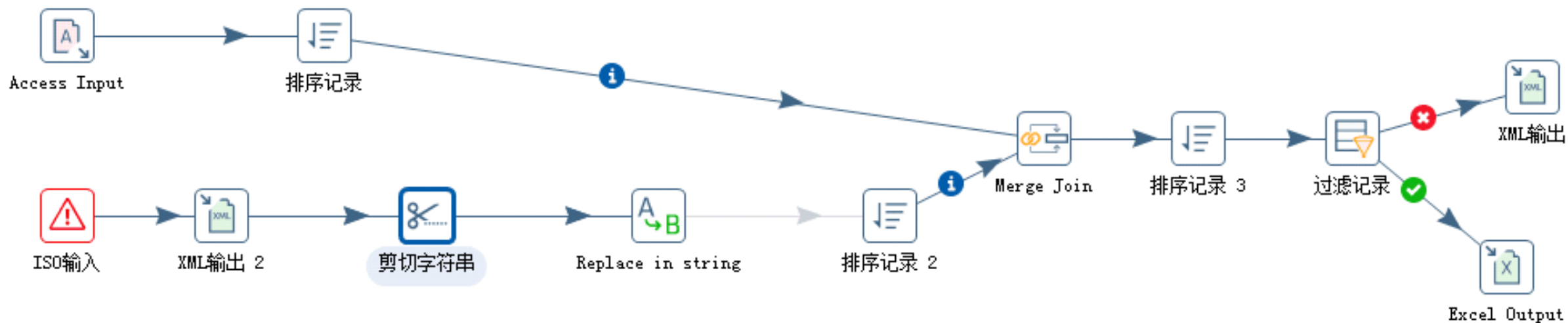
3.数据读取加载入临时表

数据读取加载入临时表的过程是指数据完成校验之后，对该数据文件进行解析，分析该数据文件的数据结构，提取该数据文件各记录中各字段内容，导入到临时表。



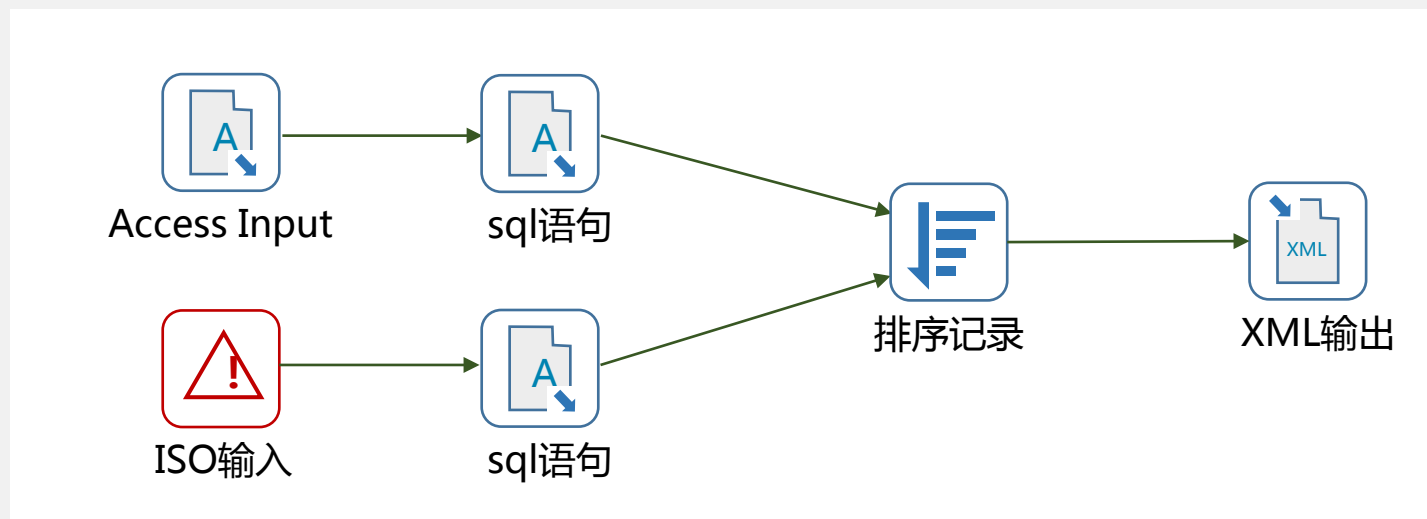
4.按数据处理规则处理

按照数据处理规则，对临时表中的数据进行处理。由于要加载的元数据格式有多种类型，例如中文图书元数据有Access文件和ISO文件组合形成元数据，因此，需要对原数据和内容进行一定处理，例如对Access文件入库后进行了字符串截取、字符串替换等一系列操作。



5.合并成“待发布表”

合并成“待发布表”就是将不同的数据合并成一张数据表，由于系统最终在前台发布供读者检索使用的是一张最终的数据表，所以**不管来源数据有多少种类型，多少个文件，最终都要合并到一起，合并成“待发布表”。**

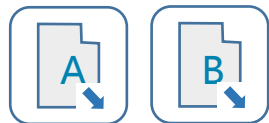


6. “待发布表” 自查重

“待发布表”的自查重功能：由于数据是分批次更新入库，所以在入库之前首先要保证本批次数据内部不能有重复数据，因为在后期数据录入到总库时需要批量比对，所以要在批量入库之前进行本次数据的自查重功能。



排序记录

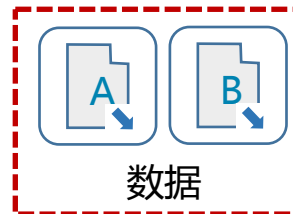


待发布数据

先逐条比对
再整批次写入



排序记录



数据

“待发布表” 自查重规则自定义

7.校验元数据和对象数据

当“待发布表”合并完成后，实际上对于元数据的处理操作就已经初步完成。由于发布数据由元数据和对象数据构成，因此需要将元数据和对象数据进行校验。如：元数据中标识对象数据的存储信息是否合法，是否和实际对象文件形成一致的对应关系等，是否对象数据有所缺失等一系列校验工作。



8. “待发布表” 全库查重

在完成“待发布表”自查重之后，“待发布表”会比对总库全部数据进行查重处理，即在写入总库之前先进行查重工作，当所有的记录查重工作完成之后才进行数据的写入。

◆ 查重工具

在系统后台操作界面上提供该查重工具的启动开关，由查重工具在后台进行异步查重，当查重完成后将结果返回。



◆ 单线程查重

查重工作涉及到数据的并发处理，例如，如果有两个数据处理人员同时提交查重工作，两部分如有重复数据，可能就会导致重复数据进行入库，因此再进行查重时对要查重的源表和目标表进行锁定排队处理。



9. 入库发布

该操作是指在数据进行查重工作完成后一次性进行写入，并创建索引。操作分为两步：一是将数据发布到云数据库中；二是管理员进行数据审核。

- ◆ 数据的发布通过专用的数据索引工具，从mysql库发布到云搜索引擎，该工具可以人工触发进行索引也可以定时进行索引。

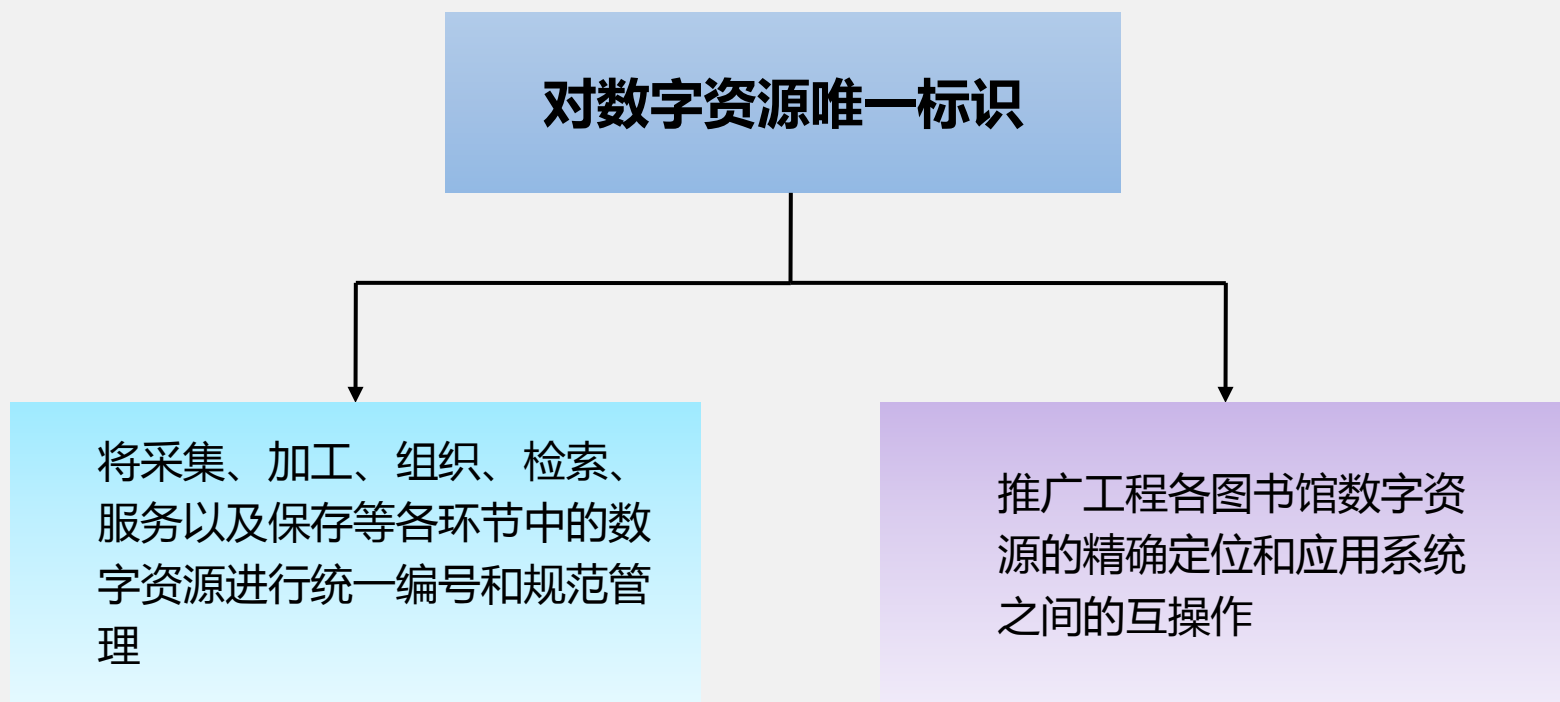


- ◆ 管理员根据云发布系统最终发布到引擎库中的数据进行预览和审核。



10. 唯一标识符注册

数字资源的唯一标识符是按统一的规则和程序授予数字资源的唯一和永久的名字，形象的被称为数字资源的条形码。

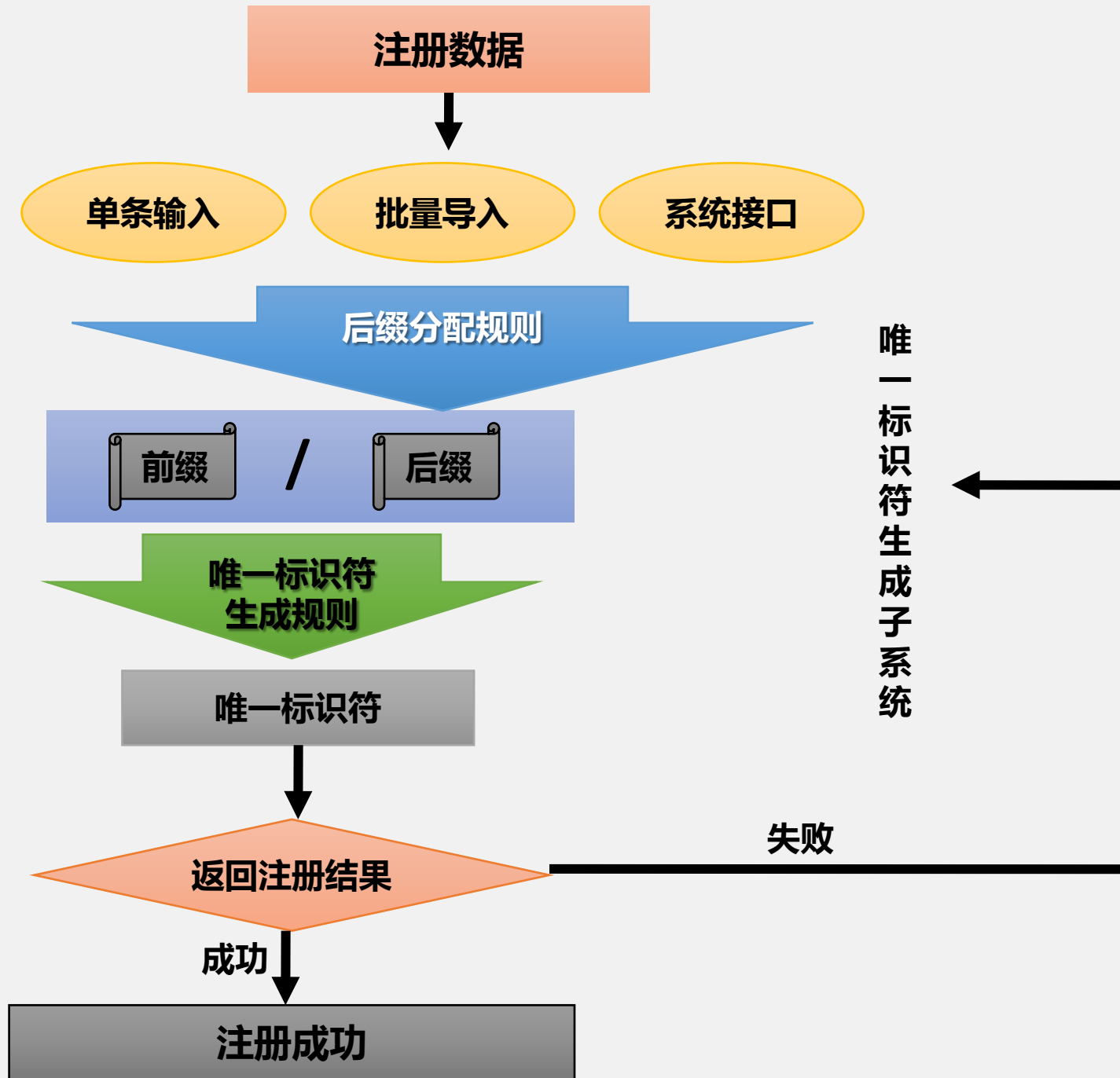


推广工程数字资源唯一标识符命名规则采用“前缀/后缀”的方式，前缀和后缀各段采用英文半角“.”分隔。

前缀

后缀

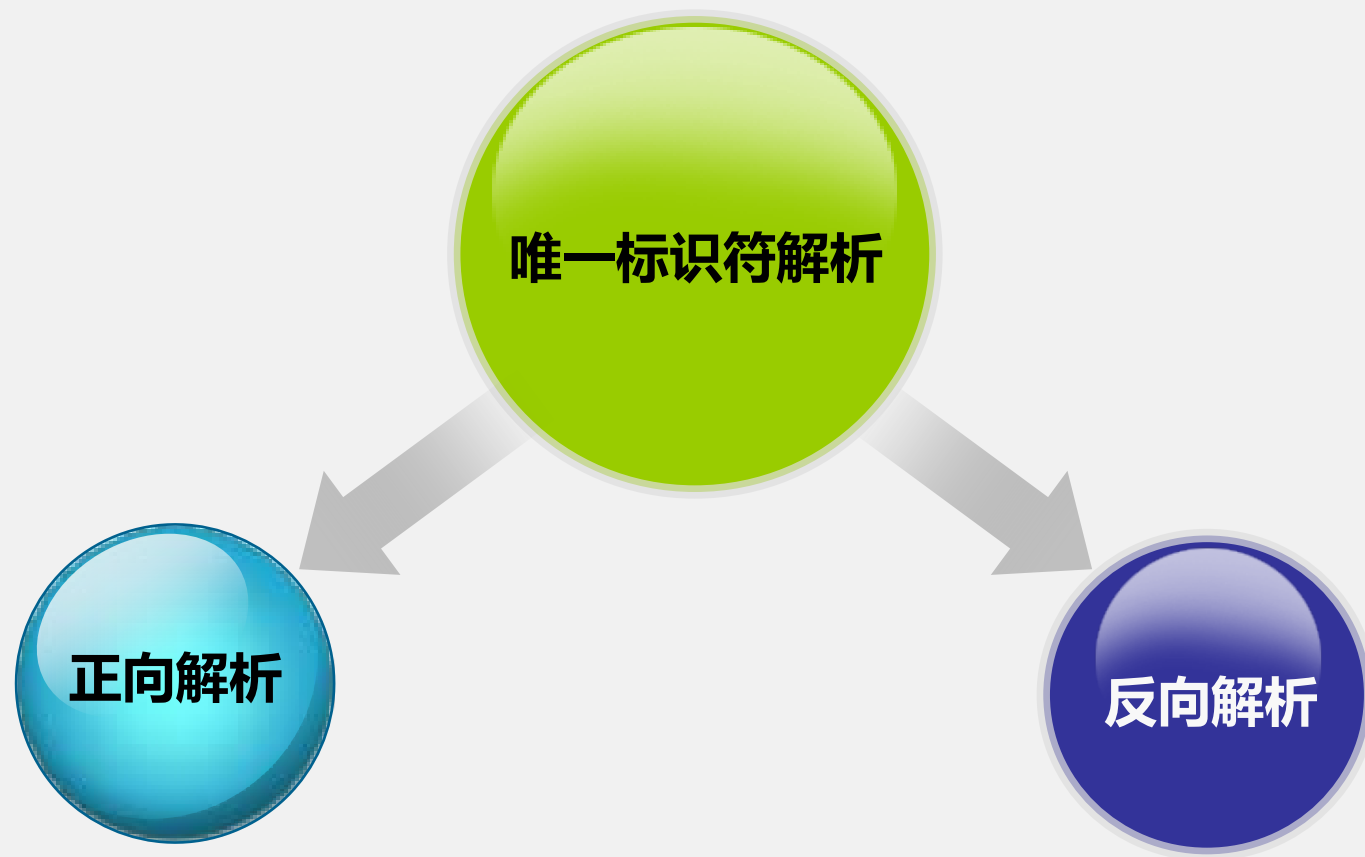
108.ndlc.2.1100009031010001/T1F23.0196011586



唯一标识符解析包括正向解析和反向解析。

正向解析：用户通过输入唯一标识符，获得与标识对象相关的信息。

反向解析：用户输入资源描述信息，获得与资源信息相符的唯一标识符。



国家图书馆数字资源唯一标识符系统于2012年11月完成建设，在近五年的运行中，完成全国48个省、市级图书馆唯一标识符系统的部署，完成250万条唯一标识符的注册，基本构建了公共图书馆数字资源唯一标识符体系。

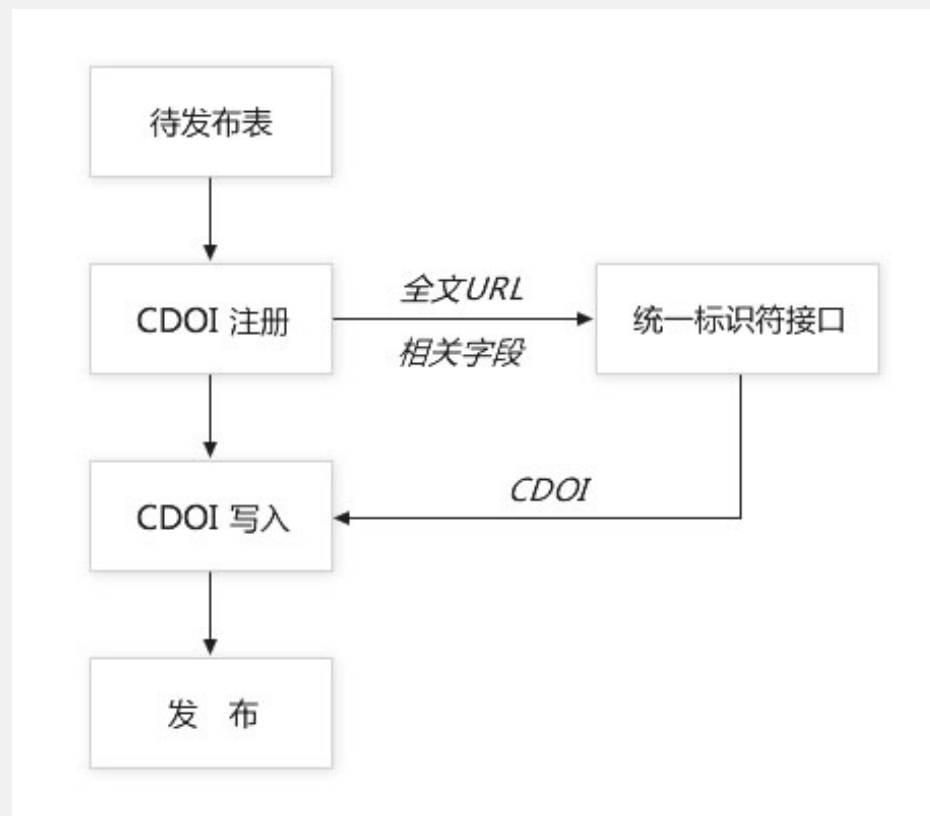
系统应用情况：

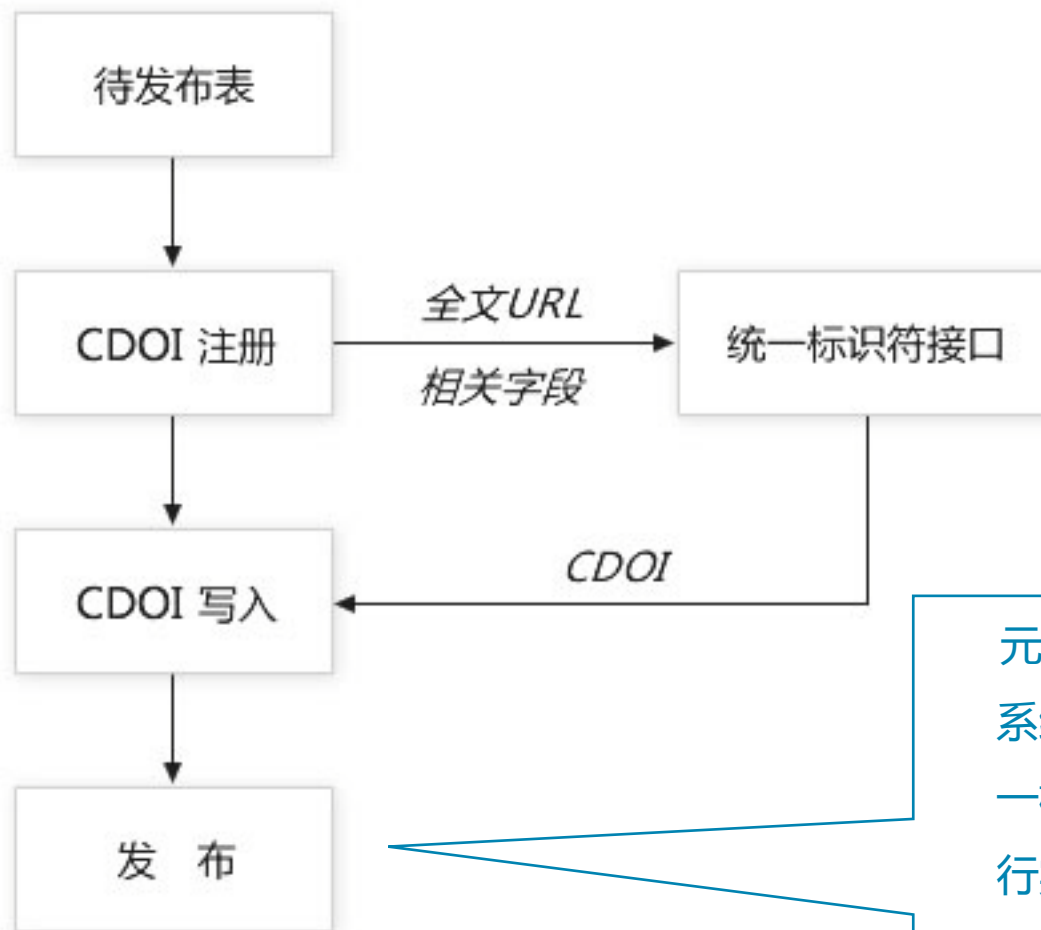
- 参与推广工程资源联建项目：实现了地方图书馆30多万条唯一标识符的注册。
- 与国图Aleph系统对接：实现了50多万册纸本图书与电子图书的数据关联展示。
- 与国图发布与服务系统对接：实现了数字资源的唯一标识和统一管理；
- 外购资源整合：完成35万条方正电子书唯一标识符的注册，并在国图发布与服务系统中实现了与自建资源的整合及可视化展示。

推广工程云发布系统将**通过唯一标识符（CDOI）来实现数字资源的统一调度**。该系统将与唯一标识符系统进行接口对接，实现发布资源唯一标识符的自动注册。**通过该唯一标识符（CDOI）定位资源。**

具体流程：

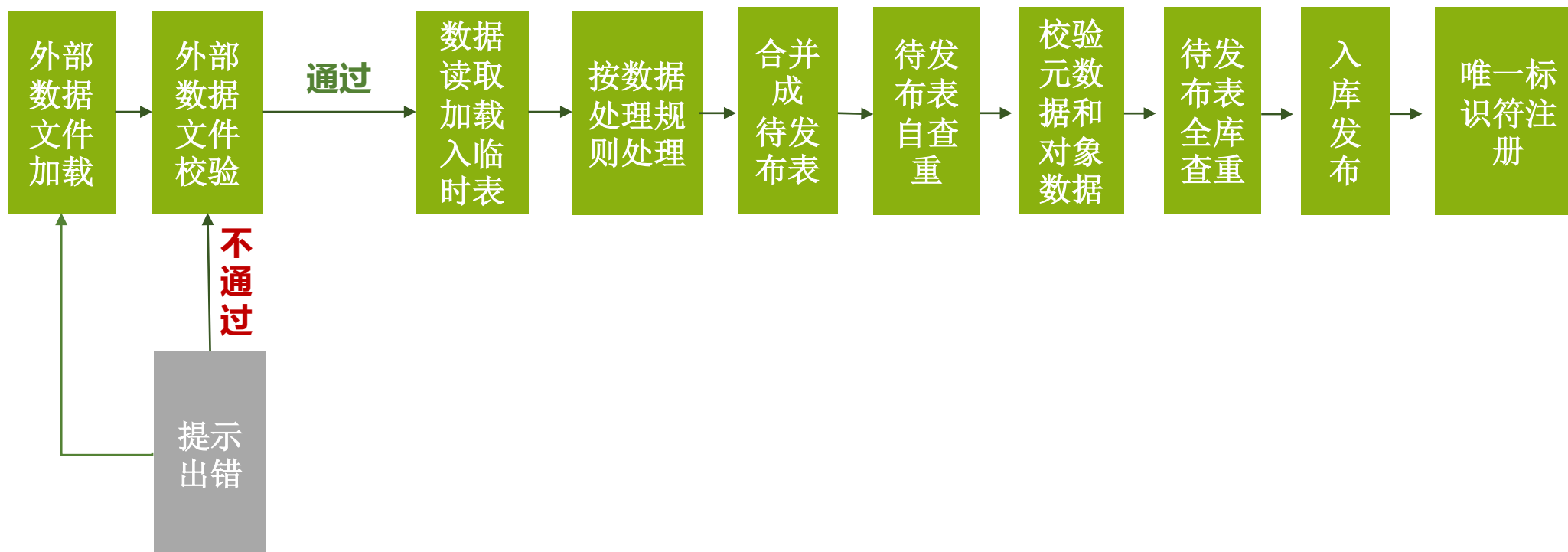
- ◆ 注册CDOI（云发布系统-->唯一标识符系统）
- ◆ 回写CDOI（唯一标识符系统-->云发布系统）
- ◆ 回写URL（云发布系统-->唯一标识符系统）



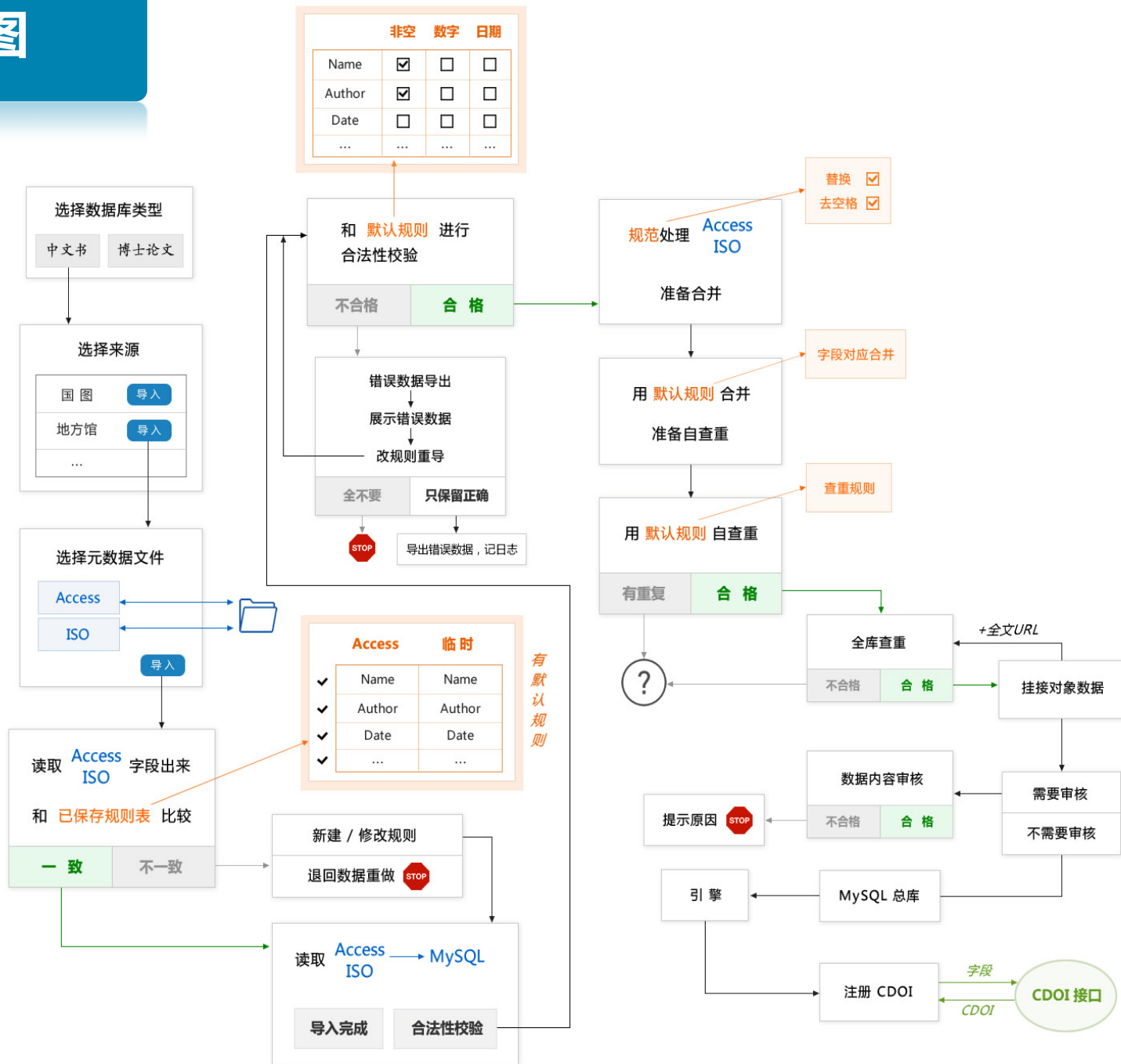


元数据在发布到系统后，读者通过系统访问数字资源时，可以通过唯一标识符**解析接口**对唯一标识符进行实时解析，获取资源元数据及对象数据，完成资源访问。

数据处理流程图

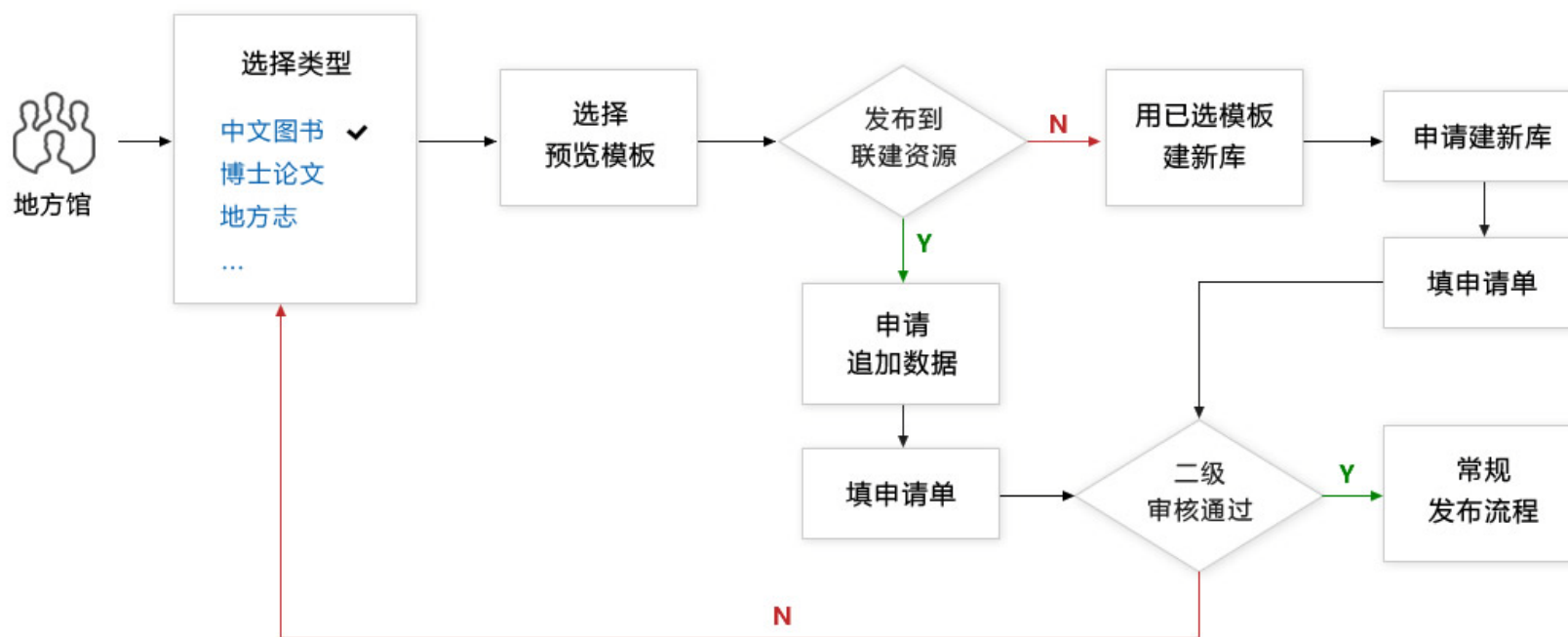


数据流程详细设计图

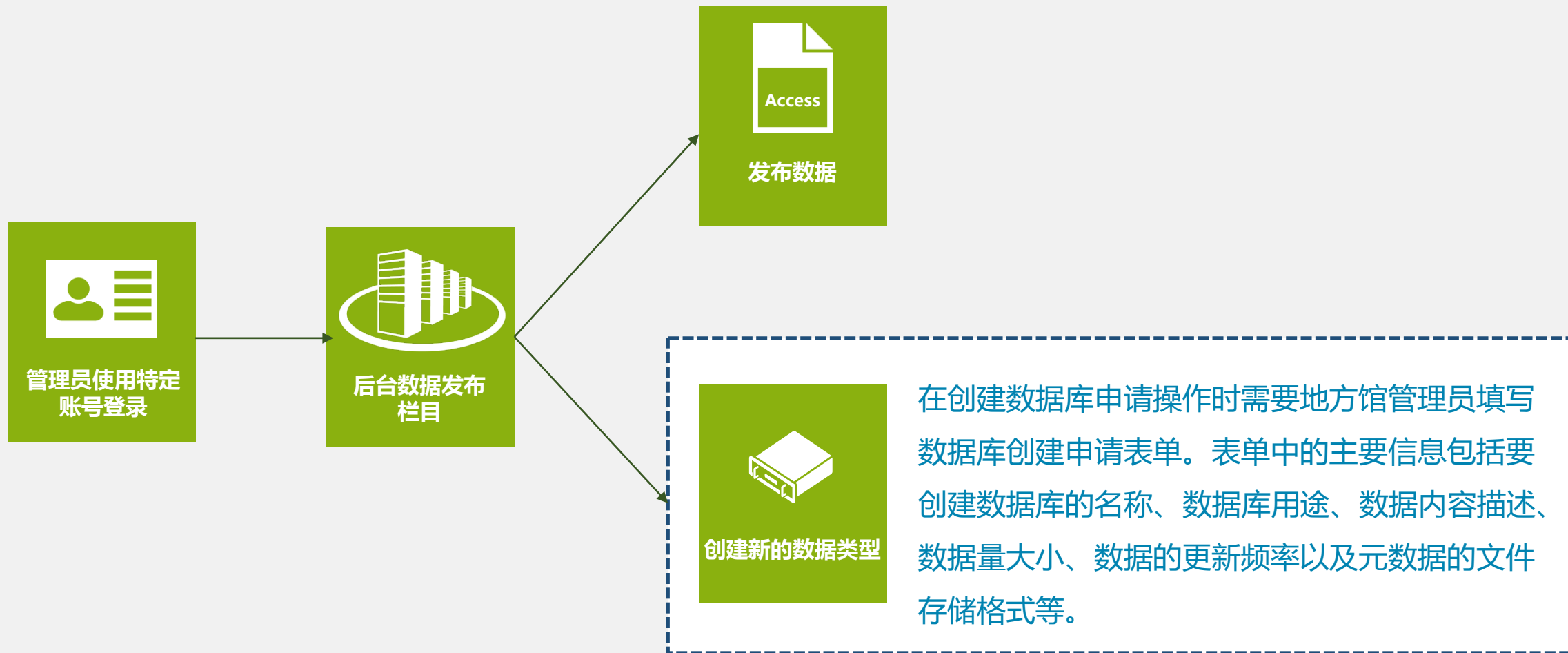


地方馆数据发布流程

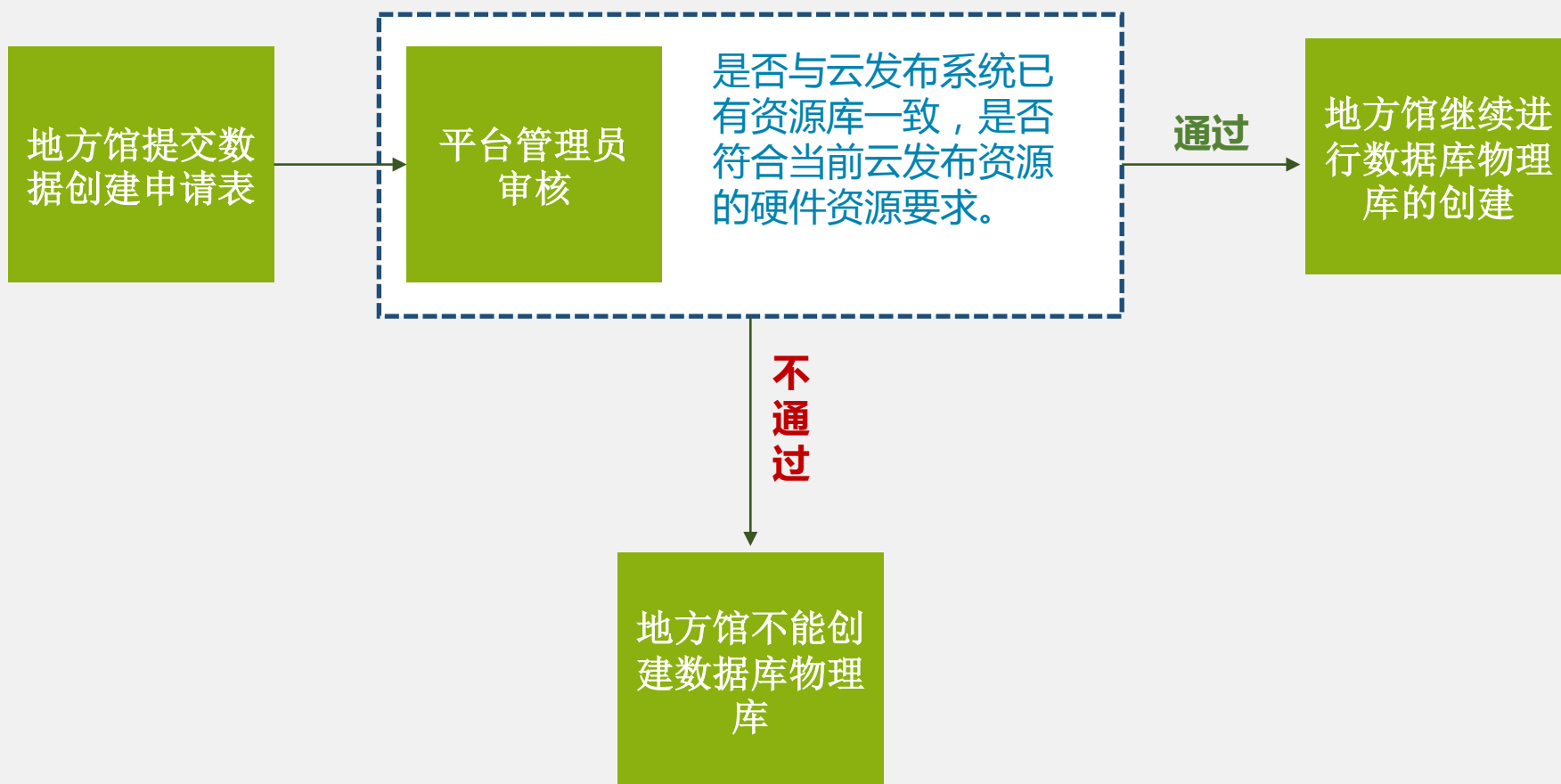
- ◆ 各省、市图书馆都可以借助推广工程云发布系统在本地对本馆数字资源进行发布与服务。但考虑到云发布系统的资源安全性和可控性，**各图书馆在发布数据前必须要进行发布任务的申请**。如果地方馆要发布的数据是现有云发布平台已有的资源类型，可以按照常规数据发布流程进行发布；如果地方馆要发布的数据类型不在云平台中（地方馆自有特色数据），需要在云发布系统新建资源库。**新建资源库必须经过平台管理员审核**，流程如下：



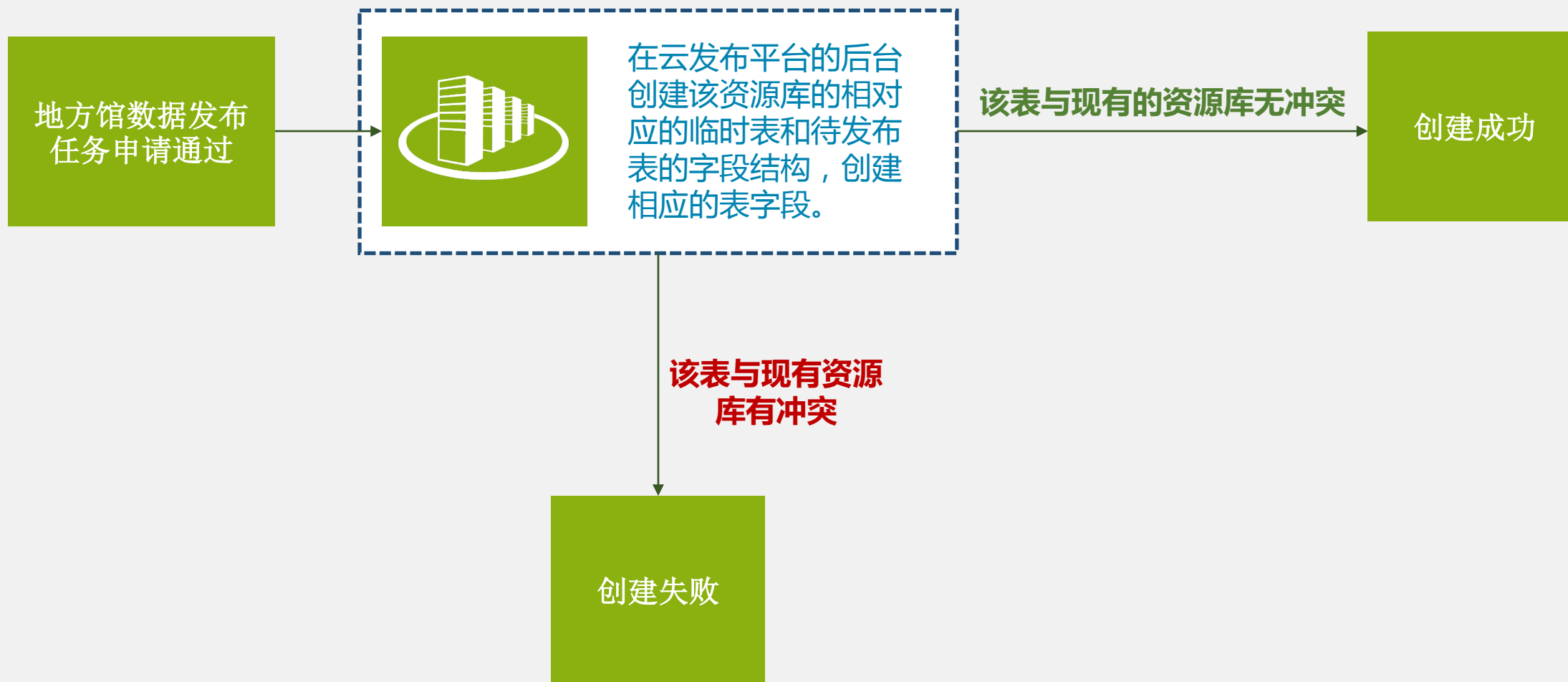
- ◆ 地方馆系统管理员使用国图开设的特定用户账号进行登录，登录成功后进入到云发布系统后台的数据发布栏目，地方馆系统管理员首先要选择需要发布的数据类型或者是创建新的数据类型。如果地方馆管理员要创建新的数据类型需要先提交创建数据库申请操作。



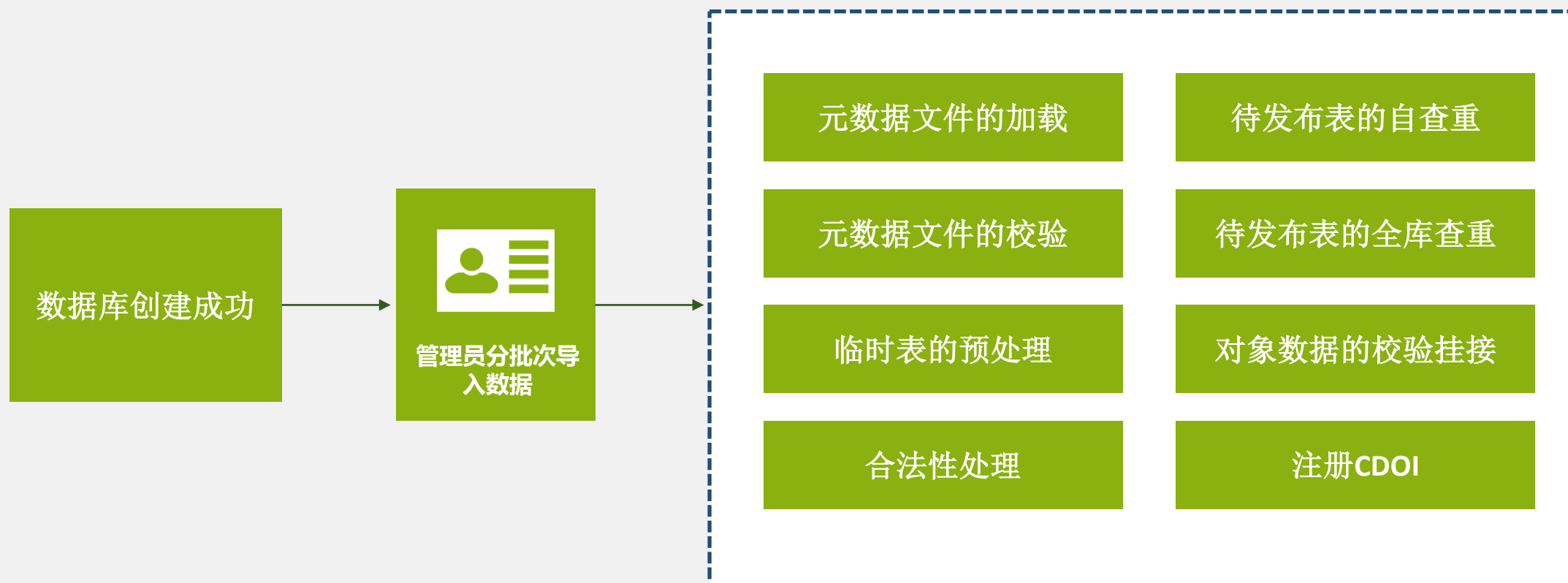
- ◆ 地方馆系统管理员填写完数据发布申请表后提交至平台管理员进行审核。平台管理员根据申请表中的内容来判断地方馆要创建的资源是否与云发布系统已有资源库一致，是否符合当前云发布资源的硬件资源（如存储空间、网络条件等）要求。



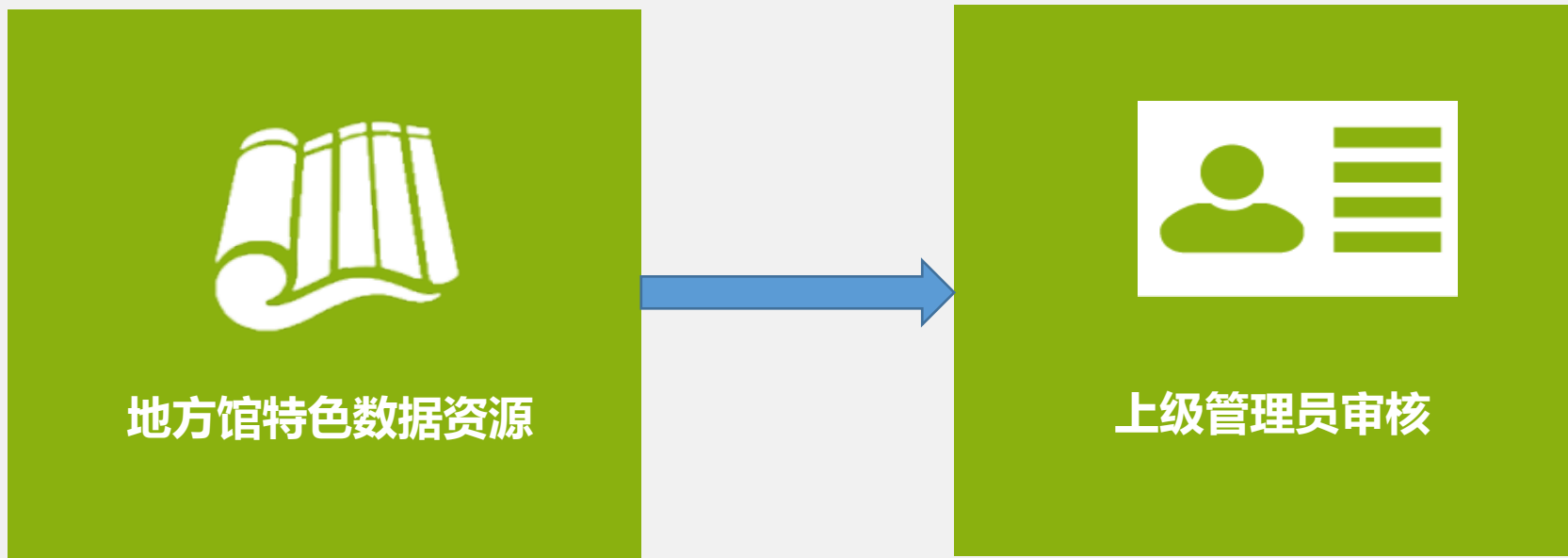
- ◆ 当申请被通过后地方馆系统管理员根据其要上传的资源类型和资源库的结构，在云发布平台上创建该资源库相对应的临时表和待发布表的字段结构，创建相应的表字段。在创建完表后系统会通过程序进行自动校验，判断该表是否与现有的资源库有冲突，如果没有冲突的话则提示对方管理员该表创建成功。



- ◆ 当创建完数据库之后，地方馆系统管理员就可以分批次的往该库中进行数据导入。其过程包括元数据文件的加载，元数据文件的校验，临时表的预处理、合法性处理，待发布表的自查重，待发布表的全库查重，和对象数据的校验挂接，以及注册CDOI等操作。



- ◆ 数据审核工作：可以由本馆管理员自行审核，也可以由上级机构的管理员进行。举例来说，市馆发布的数据，数据的合法性审查可以由市馆自行完成，也可以由省馆管理员操作。



(二) 用户权限功能设计

用户类型

角色类型

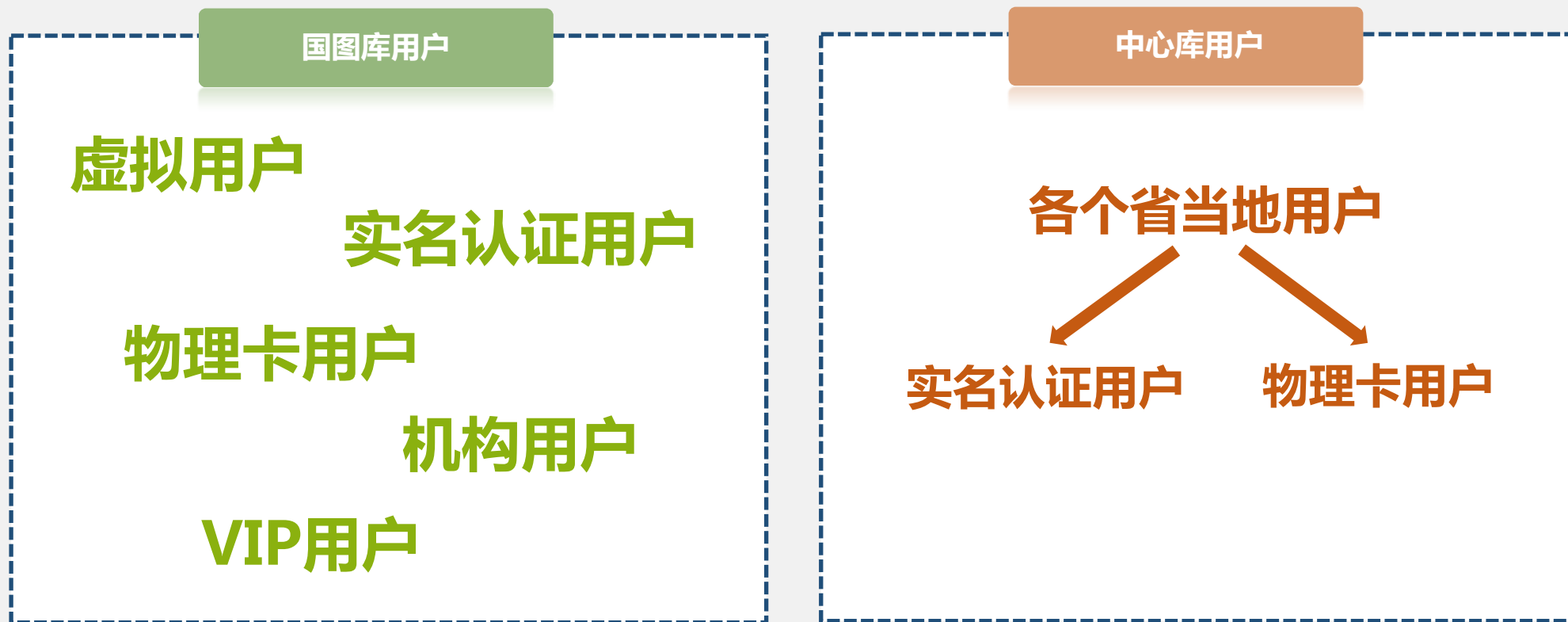
功能操作 (后台)

权限认证流程

用户系统日志相关功能等

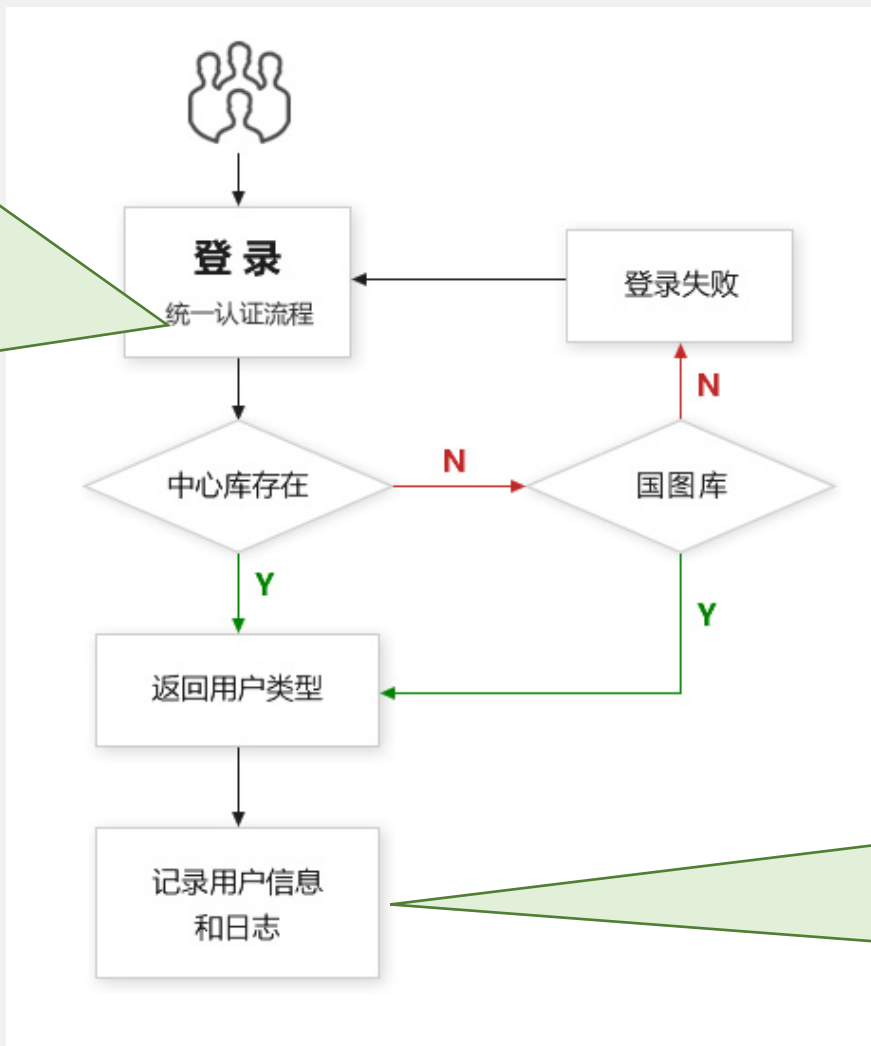
用户类型

云发布系统与推广工程统一用户管理系统进行对接，即发布平台不存在自己的用户，所有用户添加、用户的管理都通过统一用户管理系统接口来实现。目前统一用户管理系统的用户数量已经达到了一千万。其中包括中心库用户和国图用户两种来源的用户。



云发布系统进行用户登录判断时，首先先通过统一认证接口访问中心库，根据返回的用户信息来判断是否在中心库中，如果在中心库中找不到该用户，再通过接口访问国图库，根据国图库中返回的用户信息来具体判断该用户的类型，具体流程具体如下：

◆ 所有和用户操作相关的界面如用户信息修改，找回密码等功能都沿用现有统一用户管理系统界面。在用户登录界面中根据用户填入的登录信息，通过接口发送到统一用户管理系统进行认证，根据返回的登录信息判断是否认证成功，进行后续的登录操作。

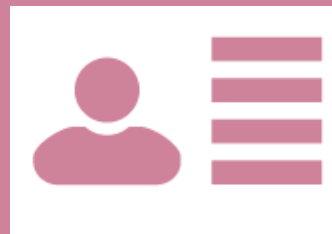


◆ 系统每次访问统一认证接口后获取的**用户基本信息**会同步到云发布平台中，以便根据该用户信息进行今后用户的各种操作日志的记录。

系统的用户类型从应用的角度来说可以划分为内部用户和外部用户。



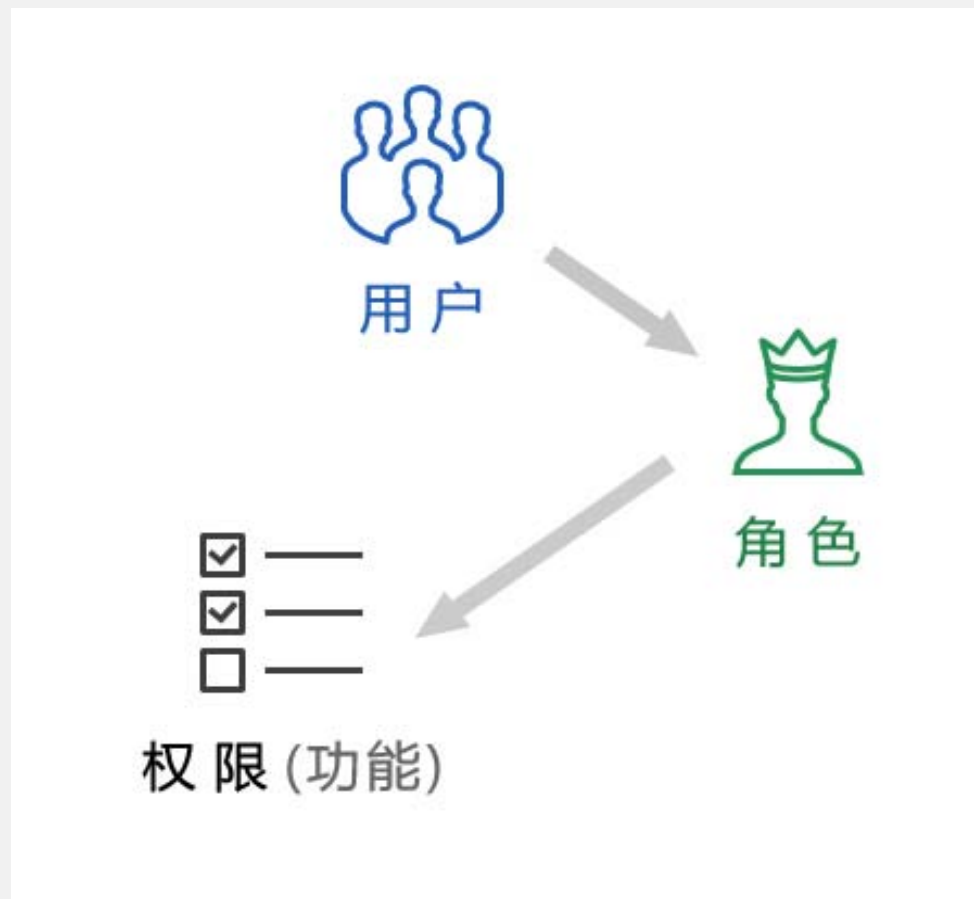
内部用户:系统后台用户,包括各省、市图书馆系统管理用户。



外部用户:读者。

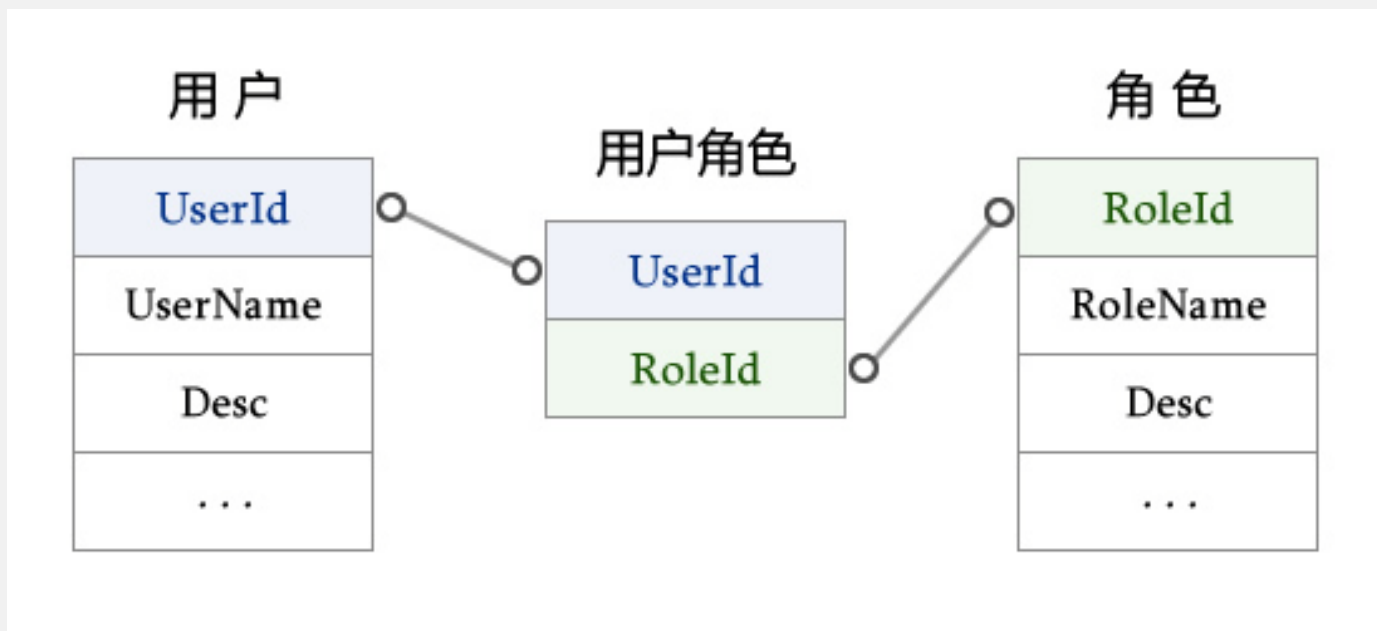
- ◆ 内部用户通常是和后台管理操作相关的用户。这部分用户的创建主要是由超级管理员在系统后台进行创建和管理,内部用户可以按照不同的用户角色来划分为各种不同类型的用户,比如有些用户可以负责对数据内容进行管理,有些用户则负责维护系统的权限或者系统的一些常规配置参数管理,有些用户则可以划分成为成员馆或分馆分机构的管理员,拥有不同层级的权限。

- ◆ 云发布平台对于用户的设计主要通过角色来进行区分，既不同的用户对应不同的角色，同一个用户也可以对多个角色。一个角色也可以设置给多个用户。用户和角色之间是多对多的关系。用户的不同权限的区分主要体现在用户的所属角色上。



角色类型

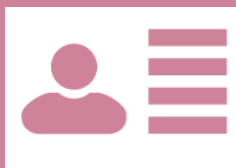
- ◆ 用户角色在后台对应一张角色表，每一个角色的属性包括：角色名称、角色的描述、角色的唯一标识。角色和用户的关联则是通过“用户角色表”来完成，用户角色表用来存储用户的唯一标识和角色的唯一标识之间的关联。





内部用户

用户角色可以由管理员在后台进行动态添加和管理。但是超级管理员角色在系统当中是预定义的，不能够修改和删除。其他所有角色都可以由超级管理员角色进行创建。例如数据管理员、权限管理员、审核员等角色。

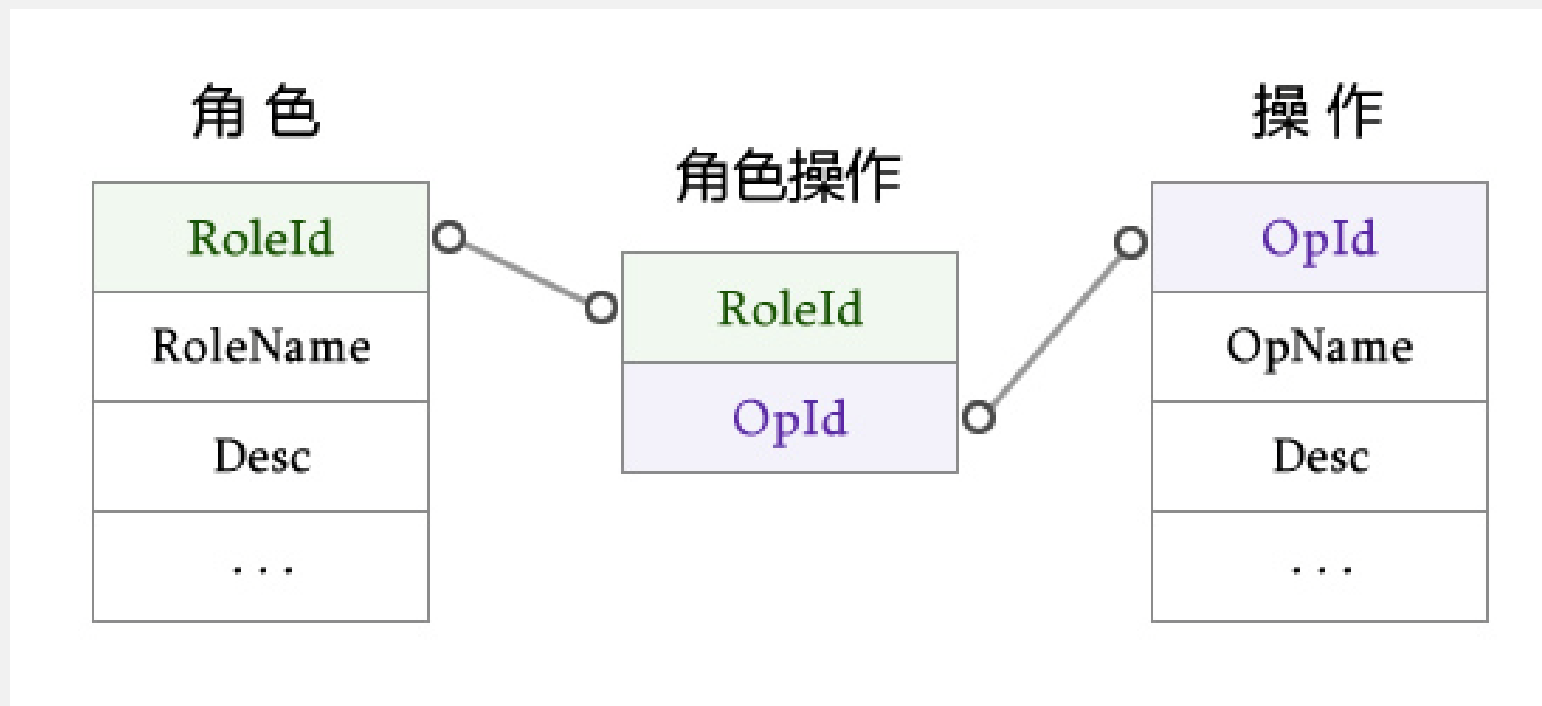


外部用户

外部用户也可以根据不同的读者的权限来设定不同的角色。外部用户可以根据读者的类型来进行权限划分，可以对不同读者所在的不同区域，对某些资源设置不同的访问权限。总之，可以是多维度的进行设置和管理。

后台功能

后台功能包括数据管理、用户管理、用户操作权限、统计分析、接口管理等功能。所有的后台功能都可以设置到或绑定到某一角色当中，一个角色可以有多个功能操作，同时一个功能操作也可以设置给多个角色，他们之间的关系是多对多的关系。



权限认证流程

云发布系统的前台权限认证是多维度的认证。基本上可以分为按照**按终端类型认证**、**按资源库级别认证**、**按照来源出版社认证**、**按照具体某一篇级资源认证**，同时还要根据系统的**用户并发情况认证**、**用户所在的IP范围认证**、**用户所访问的资源的版权时间范围认证**以及**VIP用户授权认证**。



按终端类型
认证设计



资源库级别
认证



按照出版社进
行认证



按照篇级资源或
书籍进行认证



系统的用户并发
情况进行认证



用户所在的IP范
围进行认证



资源的版权时间
范围进行认证



VIP用户授权
认证



按终端类型认证设计

在判断某一类资源是否有权限访问时，首先要判断访问该资源用户的来源终端情况。如果终端是移动端，则根据移动端的访问规则进行验证，如果来源是PC端则按照PC端的认证流程进行。



资源库级别
认证



按照出版社进
行认证



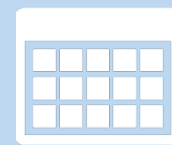
按照篇级资源或
书籍进行认证



系统的用户并发
情况进行认证



用户所在的IP范
围进行认证



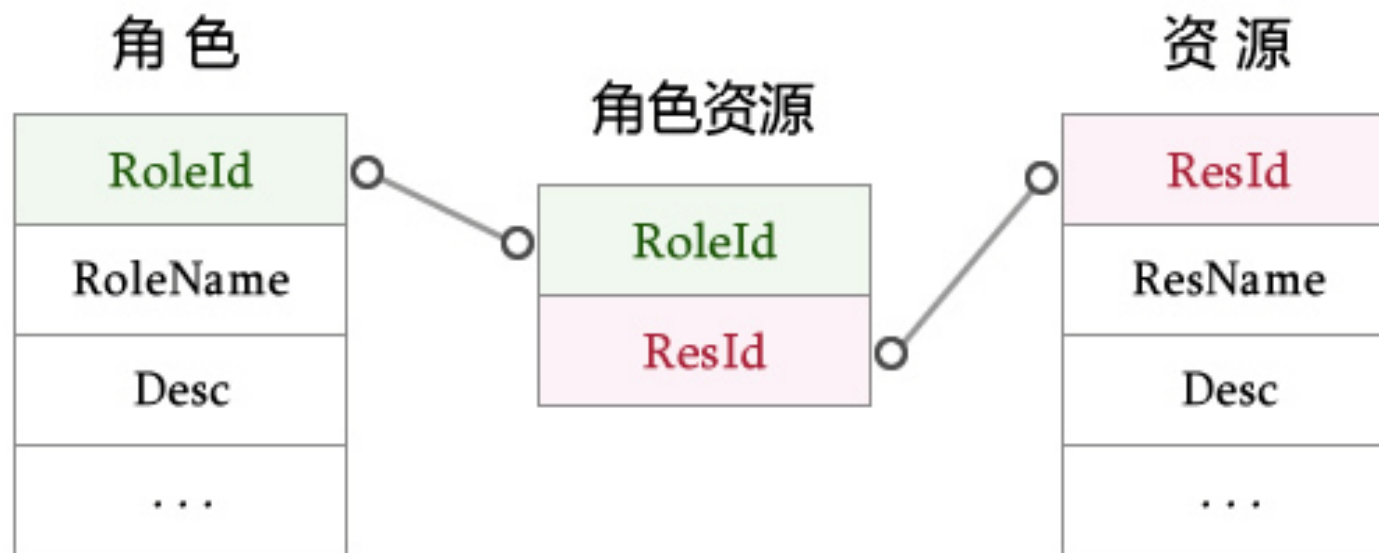
资源的版权时
间范围进行认证

- ◆ 不论是移动端还是PC端其基本的验证规则和维度都包括按资源库级别、按不同的出版社级别、按照系统的并发权限控制、按照IP范围权限控制、按照版权时间范围许可进行控制。



按资源库级别认证

云发布平台有多种不同类型的数据，例如中文图书、博士论文、古籍、地方志、老照片、民国文献等类型的数据。系统要求首先能够根据资源类型进行用户的权限认证授权。



每一类文献资源可能针对不同的角色有不同的访问权限和许可。

数据库类型	未登录用户	实名认证用户	物理卡用户	机构用户	发卡用户	VIP
中文图书	●	▲	▲	▲	▲	★
博士论文	●	▲	▲	▲	▲	★
民国图书	●	▲	▲	▲	▲	★
民国期刊	●	▲	▲	▲	▲	★
民国法律	●	▲	▲	▲	▲	★
数字方志	●	▲	▲	▲	▲	★
矢量数字方志	●	▲	▲	▲	▲	★
碑帖菁华	●	▲	▲	▲	▲	★
甲骨拓片	●	▲	▲	▲	▲	★
老照片	●	▲	▲	▲	▲	★

权限类型：

●资源检索

▲结合其他维度判断

★资源检索、在线阅读



按出版社授权

针对图书类资源，每一种数字资源都有不同的出版社，不同出版社的访问权限也是不同的，这些权限都可以由管理员在后台进行设置，例如某些用户只能访问某些出版社的资源，没有权限的出版社的资源对于该用户来说不能访问。

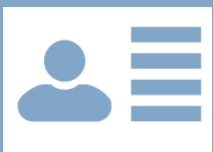
数据库类型	未登录用户	实名认证用户	物理卡用户	机构用户	发卡用户	VIP
出版社1	●	▲	▲	▲	▲	★
出版社2	●	▲	▲	▲	▲	★
.....	●	▲	▲	▲	▲	★
出版社N	●	▲	▲	▲	▲	★

权限类型：●资源检索 ▲结合其他维度判断 ★资源检索、在线阅读



按篇级授权

用户的篇级权限主要是指某一类角色浏览某一篇文章（某一本书）或元数据的全文时所获得的权限。系统根据不同的用户所属的不同权限对某一篇文章或文献的权限是不同的。



用户角色



用户浏览的文献或书籍

系统会创建记录的唯一标识和用户角色的关联表。

该用户对该篇文献或该本书有全文浏览的权限

用户继续浏览

该用户对该篇文献或该本书无全文浏览的权限

用户未经授权不可浏览

- ◆ 系统对篇级的权限控制还可以进行更细粒度的划分，例如，当用户已经授权可以访问某篇或者某一本书的全文浏览权限时，但是还要限制对该书和该篇进行全文浏览的页数。此时就需要对该用户或该类角色进行更细粒度的限制。这种情况一般可以分为两种权限类型：



部分权限
授权

如：可以浏览某一篇pdf文件前多少页或占整本书的百分比；按整本书的某些章节授权。



全部权限
授权

某一篇pdf全文的所有页（正常情况下只能看24页）。

- ◆ 在对篇级的权限控制时，可以设置某个出版社所有书在一天能看的总页数，也可以设置某一本书在一天能看的总页数，如果是多媒体资源，可以设置音视频播放时长。



文本

如：国家图书馆出版社出版的所有图书每天提供读者阅读10万页；百部经典图书《论语》每天提供读者阅读1000页等。



多媒体

如：单个音视频节目时长控制；某个出版单位所有音视频节目时长控制等



系统并发访问授权

云发布系统需要针对系统的特定资源访问进行并发控制，可设定某类资源允许用户同时访问的次数上限。



并发数控制

用户访问同一本书或查看同一篇文献时，同时在线的用户不能超过一定数量。即用户具备某一篇文献或某一本书的浏览权限时，还需要结合当前系统该资源的并发数进行判断。

并发数未达到上限

用户正常浏览

并发数达到上限

用户浏览受限



按IP范围进行授权

按照IP范围对数字资源进行授权也是云发布系统的资源授权访问的主要需求之一，即可以按照某一具体IP或IP地址段对某一类用户进行授权访问，也可以结合IP地址范围或IP段和某一资源及某一用户进行授权访问。



地方馆IP授权

某一地方馆的资源接收到用户的访问请求时，可以针对于该用户进行IP范围判断。

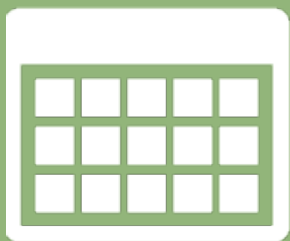
该用户的IP范围是属于该地方馆的IP段内

用户正常浏览

该用户的IP范围不在属于该地方馆的IP段内

用户不可浏览

◆ IP范围进行授权，也可以支持统一资源设置多个不同的IP段。具体又分为**内网**，**互联网**，**专网**。



版权时间范围许可授权

数字资源都是有版权许可的。版权许可具有一定的时间范围。例如某一出版物的版权时间许可范围是一年，那么在该时间范围内用户可以有权对其进行浏览和访问，超出该时间段则该出版物内容将进行下架或者访问受限。



系统版权控制

管理员可以具体设置某一类型资源的版权许可时间段。当前的用户进行访问时，根据该访问资源的时间段与当前时间进行判断。

当前时间段在许可的范围内

资源正常访问

超出时间范围

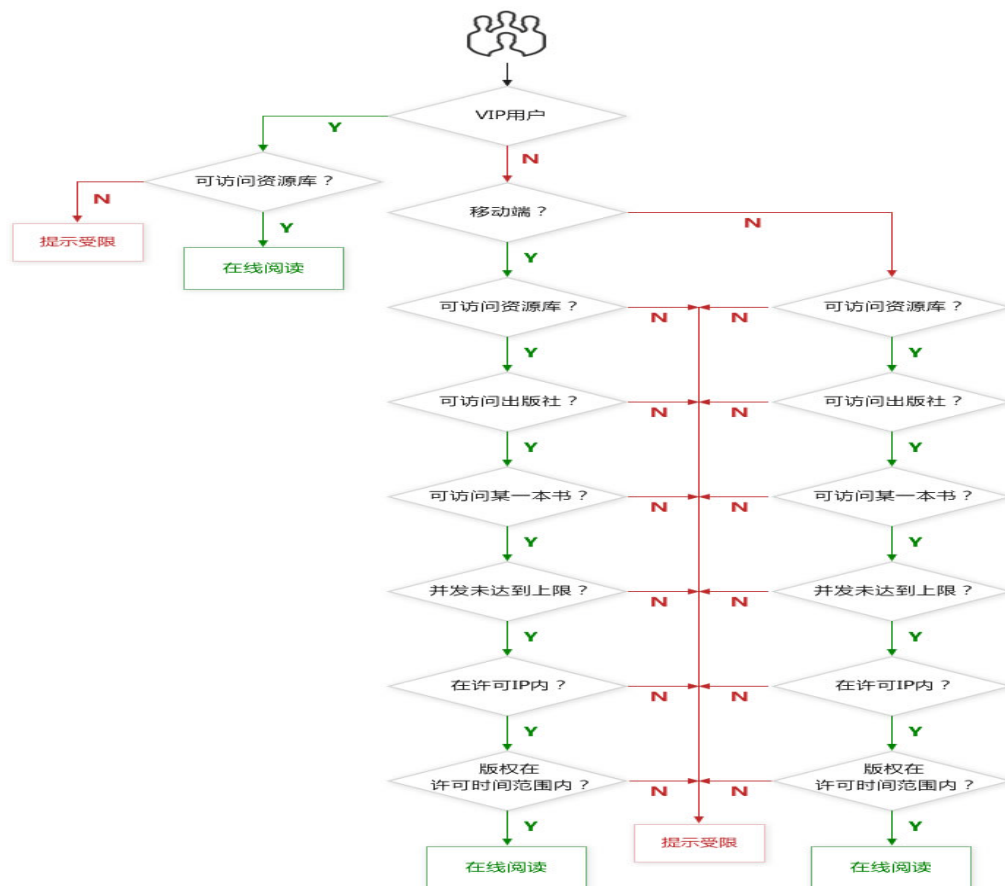
资源不可访问



VIP用户授权

数字资源系统的用户类型当中有一类权限级别较高的特殊用户，系统将这类用户定义为VIP用户，VIP用户针对资源库进行开放，该用户拥有最高的数据访问权限，可以访问系统的大部分资源或者所有资源。

用户资源权限认证流程



综上所述，云发布系统资源权限控制是一个比较复杂的过程，该流程具有多种维度的判断。总体流程的原则是以用户作为出发点，结合具体某一资源的特定权限属性进行判断。



用户访问日志记录

不论是哪类用户的角色权限，访问某一类型资源，对于某一种数据的访问、检索、浏览，系统都会进行日志的记录。系统会记录下该用户的访问帐号、系统访问时间、访问的具体资源名称、访问的具体资源所属的数据库、该用户的IP地址，同时根据该用户IP地址可以判断出该用户所属的地区，如该用户所在的省份和城市等。此类信息全部记录到系统的后台日志表中。该日志表示作为以后系统对于用户行为统计分析的依据。

(三) 数据发布功能设计

数据发布检索功能

对于已发布的元数据进行检索，检索功能可以分为简单检索和高级检索。

简单检索

高级检索

- **高级检索**：支持数字资源类型选择，选择一种数字资源类型时，检索项为将当前所有的可检索字段，可以选择“主题”、“全文”、“目录”、“篇名”、“摘要”等不同的检索项，实现元数据检索、全文检索、目录检索、篇名检索、摘要检索等检索方式。支持繁简体通检和检索词自动匹配。

摘要检索

- ◆ 根据图书内容介绍或者摘要中的文字进行检索。
- ◆ 针对部分有中文摘要的期刊论文资源，系统提供摘要检索功能。
- ◆ 可检索中文摘要里出现检索词的资源，检索结果标红展示。

全文检索

- ◆ 系统支持全文检索功能。
- ◆ 第三方系统通过调用全文检索接口进行全文检索。

基于目录的检索功能

- ◆ 读者可以通过目录检索查询文献。
- ◆ 第三方系统通过调用目录检索接口进行目录查询。

其他

- ◆ 系统可对用户检索行为进行统计，并动态的将高频检索词作为检索推荐提示用词推荐给用户。
- ◆ 检索结果中支持检索词高亮显示，用户通过点击关键字词直接到达与搜索关键字最相关的文档中的段落或句子。

对象数据的展示功能

对于对象数据，例如原文pdf在线阅读测试功能，对音视频文件在线播放等功能进行展示。



通过调用第三方的WEBPDF插件对文本文献进行展示。



WEBPDF支持PDF格式资源在PC、PAD、手机等多终端上阅读，支持断点、多屏阅读。可通过目录快速定位到相应的资源页。

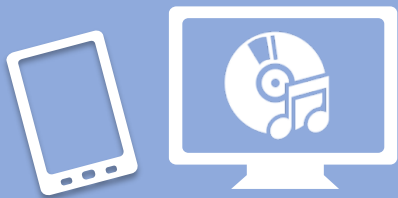


全文展示

矢量方志提供影像文件和矢量文件的对照显示，联动展示功能。

报纸资源能定位到篇目，并支持将定位的文章高亮显示。

古籍资源，提供合并阅读的功能，即通过卷次信息，定位各卷全文，而不用每读一卷，打开一个网页。



流媒体播放平台对接

对于视频和音频资源要与流媒体播放平台进行对接，调用流媒体平台接口，获取音视频数据流，实现资源的在线播放功能。视频播放支持高清、标清，读者可自行选择。支持PC、PAD、手机端等多终端多屏观看。



移动端图书对象数据全文展示

支持移动端图书对象数据为TXT和EPUB格式的全文展示。

数据分类导航

根据不同的数据分类进行分类展示。可以按照资源库分类展示；也可以按照不同分类的主题进行展示。

◀ 返回首页



图书



期刊



报纸



论文



古籍



音乐



影视



缩微

数据导出下载功能

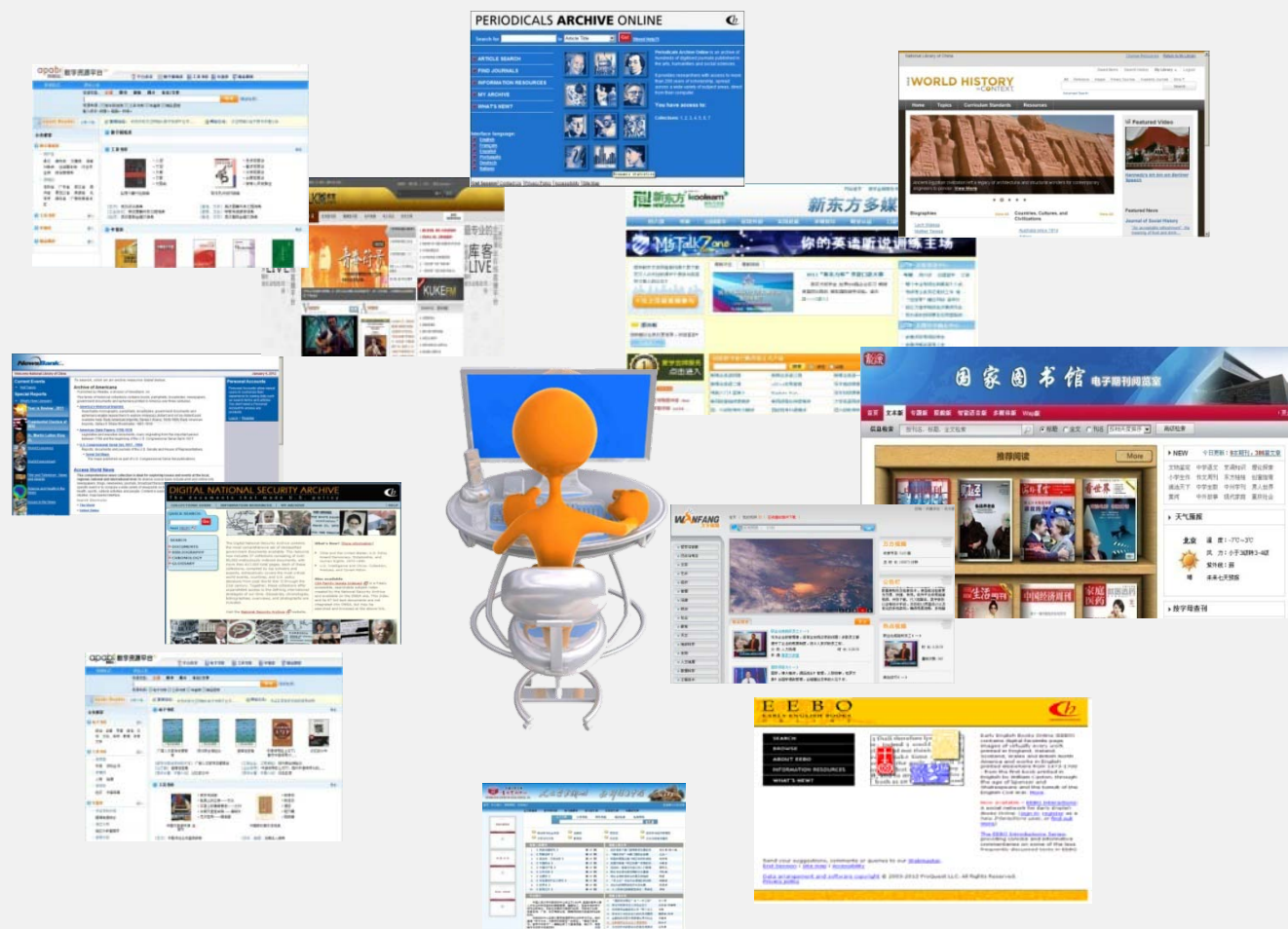
云发布系统提供元数据下载页面，在资源发布完成后，管理员可以在后台对已发布的数据进行批量导出。第三方系统在授权许可下，可选择需要的字段下载资源元数据，导出为需要的文件格式，如excel、txt、xml等。其中**唯一标识符字段为必备下载字段**。第三方系统可将元数据导入到本地数据库中，对外服务时，通过唯一标识符获取发布系统中的元数据和对象数据对读者提供服务。导出格式可以为特定的EXCEL或者xml。

注意：导出的数据数量在系统当中设定一定的上限，不能够将整库的数据全部导出，也不能够一次性导出大量的数据。只能导出本馆发布的数据。



(四) 统计分析功能设计

系统可以根据积累下来的用户行为日志数据对系统的所有使用和用户行为进行详细的统计分析，统计结果以图、表的形式在管理后台显示。



1.资源统计分析

基于资源的统计分析，主要是针对云发布系统的数字资源库的建设情况和使用情况进行统计分析。

资源统计指标包括资源大小、资源并发、资源流量、资源数量（种、册、页）。



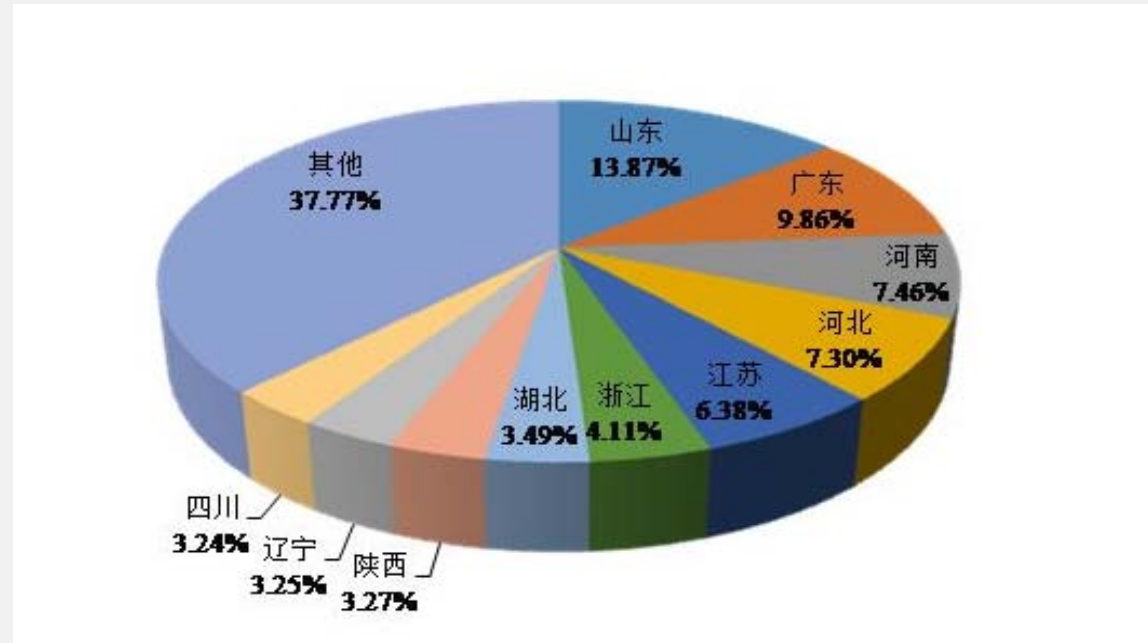
◆ **统计内容**包括资源发布量、资源使用量、资源并发量、资源流量。

2.访问统计分析

系统可以按浏览器、操作系统、终端类型、网络提供商、地域等数据来源对资源访问情况进行统计。

系统可以按不同的时间维度对**第三方应用系统访问**本应用系统内资源情况进行统计分析。

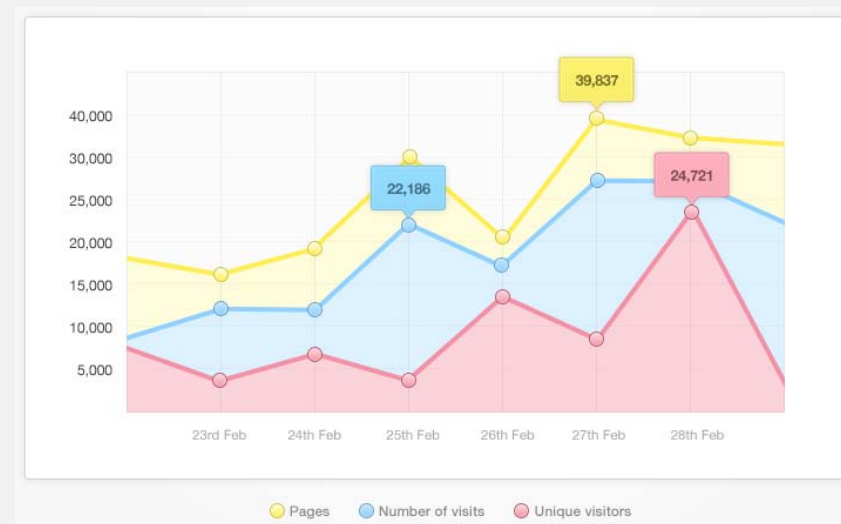
统计指标包括资源浏览量、资源访客数、IP数、资源跳出率、资源访问时长、资源访问地域。



按访问地区进行统计

3.内容统计分析

内容统计分析指对用户交互模块中产生的数据进行统计。统计指标包括资源评星数、资源评论数、资源笔记数、资源分享数、用户评价数、用户笔记数、用户分享数。





中国国家图书馆 · 中国国家数字图书馆
NATIONAL LIBRARY OF CHINA NATIONAL DIGITAL LIBRARY OF CHINA

谢谢！

