

CALIS高等教育数字图书馆技术体系

□ 王文清 张月祥 陈凌 / CALIS管理中心 北京 100871

摘要: 中国高等教育文献保障系统 (CALIS) 在“十一五”三期项目建设期间面临着巨大挑战: 如何为近2000个高校成员馆提供有效的标准化、低成本、自适应、可扩展并且馆馆之间可互联互通的数字图书馆平台服务。在云计算发展的浪潮中, CALIS采用了云计算技术作为自身的解决方案的基础, 搭建了由CALIS中心级云平台、共享域级云平台和本地平台组成的多云服务体系。文章从CALIS云服务体系、统一认证体系、资源交换与处理体系、资源与服务整合体系、资源发现与获取体系以及咨询与培训体系等六个方面详细介绍了CALIS高等教育数字图书馆的整个技术体系。

关键词: 数字图书馆, 技术体系, 云计算, 云服务, CALIS

DOI: 10.3772/j.issn.1673—2286.2013.01.006

1 概述

CALIS历经了“九五”、“十五”和三期的建设, 特别是在三期项目建设期间, 面临着巨大挑战^[1], 主要包括: (1) 为全国近2000个高校成员馆提供标准化、低成本、自适应、可扩展的数字图书馆统一服务和集成平台; (2) 这些馆应能通过彼此互联, 构成全国高校数字图书馆三级共建和共享服务以及多馆服务协作的联合体系, 共同为高校师生提供全方位的文献服务、咨询服务、电子商务和个性化服务。

面对上述挑战, CALIS提出了以“云计算”技术为基础的高校数字图书馆技术体系以及由全国中心、省中心、共享域中心和核心馆组成的服务体系。共享域指的是一种图书馆联盟, 其成员馆可以彼此实现资源共享。共享域既可以是区域性图书馆联盟, 如省级、城市级、地区级或园区级联盟, 也可以是面向行业、领域或学科的联盟, 如外语联盟、旅游院校联盟等。在技术体系中, CALIS建立了由CALIS中心级云平台、共享域级云平台和图书馆本地平台组成的彼此互联的多级云服

务体系; 在此基础上, CALIS进一步建立了统一认证体系、资源交换与处理体系、资源与服务整合体系、资源发现与获取体系以及咨询与培训体系, 从而为成员馆提供较为完整的数字图书馆解决方案。

2 多云服务体系

CALIS数字图书馆在技术上而言是一种包含公有云、私有云、社区云和混合云的多云的服务体系, 包括CALIS中心级云服务平台、CALIS共享域级的云服务平台和本地平台。CALIS多云服务体系以用户的统一认证为基础, 以资源交换、资源与服务的整合为主要技术手段, 借助咨询与培训服务, 为读者提供便利的资源发现和获取方式。

2.1 云服务平台

CALIS云服务平台的整体架构^[2]包括三个层次:

(1) CALIS中心级云服务平台, 为所有成员馆和读

者提供服务。(2) CALIS共享域级云服务平台, 基于 Software as a Service (SaaS) 技术, 用于构建省级和地区级的共享域, 共享域成员馆可以免费使用云服务平台中的各个应用系统, 为本馆的馆员和读者提供服务。(3) 本地平台, 狭义而言, 是指可以在本地部署的 CALIS 本地应用, 以用于构建本地数字图书馆; 广义而言, 是指将 CALIS 中心级和共享域级的云服务平台提供的统一 API 嵌入到图书馆本地已有系统中, 构建出“由 CALIS 助力”(即 Powered by CALIS) 的本地数字图书馆平台。

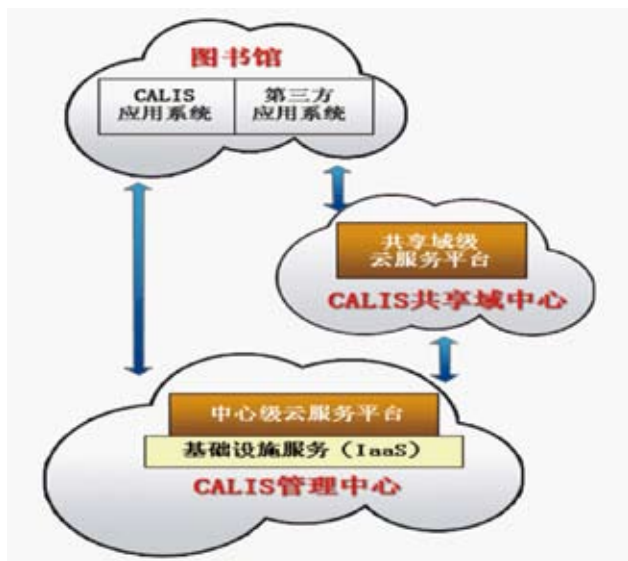


图1 CALIS云服务平台总体架构

根据美国国家标准与技术研究院 (National Institute of Standards and Technology, NIST) 给云计算部署方式的分类^[3], 云计算分为四种部署方式: 私有云 (Private cloud)、社区云 (Community cloud)、公有云 (Public cloud) 和混合云 (Hybrid cloud)。在 CALIS 多云服务体系中, 对于高校用户而言, 部署在 CALIS 管理中心的 CALIS 国家级云中心属于全面开放的公有云, 部署在各个省中心或共享域中心的共享域级云中心属于社区云, 图书馆自身如果采用云计算的话, 构建的是私有云。图书馆利用自身的私有云的同时, 整合 CALIS 国家级云中心 (公有云) 以及 CALIS 共享域级云中心 (社区云) 后, 构成一个混合云的部署方式。各馆的私有云、CALIS 共享域级云中心的社区云和 CALIS 国家级云中心的公有云整个形成完整的 CALIS 的多云服务体系。

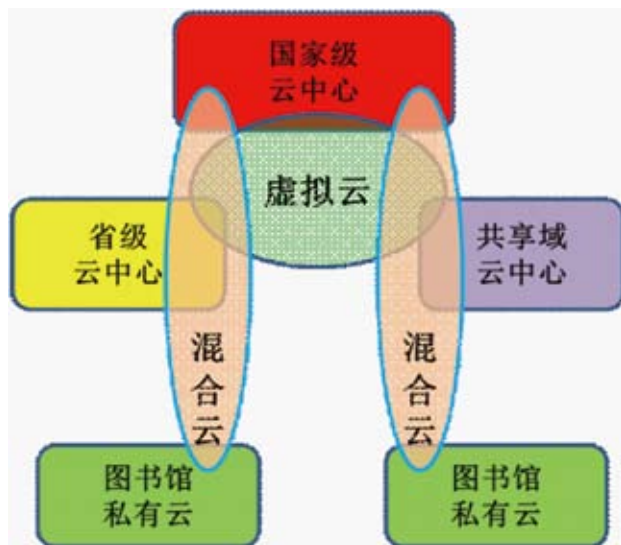


图2 CALIS多云服务体系

(1) CALIS中心级云服务平台

CALIS 中心云服务平台从逻辑上分为基础层、数据层、支撑层、馆员服务层、读者服务层以及门户层。基础层包括基本的云基础设施, 如硬件、软件、网络、存储和系统软件等等; 数据层包括数据处理、数据管理、数据挖掘和数据保存; 支撑层包括应用系统和服务信息的统一注册、数据的统一交换、用户的统一认证以及统一计费; 馆员服务层有馆员联合编目、联合订购、古文献、参考咨询工具以及培训与考试系统。读者服务层由 e 读、e 问、e 得和 e 学四大块组成, 其中 e 读是 CALIS 各类学术搜索服务的统称, 具体包括 e 读中心服务、e 读本地化服务、外文期刊网服务 (CCC) 和联合目录 OPAC 服务, 支持图书、期刊、期刊论文、学位论文、特色库、名师、课程和古文献等资源的检索, 对于电子全文馆藏服务的调度通过资源调度实现。e 问包括联合问答、业务咨询和定制推送服务; e 得包括 e 得中心门户、馆际互借与文献传递 (ILL) 服务中心、馆际互借与文献传递调度中心、全文资源下载以及相应的全文数据库等; e 学包括面向读者的信息素养教育平台和教学参考信息服务系统。门户层主要由 CALIS 中心门户、基于 OpenSocial Gadgets 规范的个性化门户、CALIS 直通车以及供门户嵌入的各类 OpenAPI 组成。

(2) CALIS共享域级云服务平台

CALIS 共享域级云服务平台为共享域中的成员提供了以统一认证为基础的六大服务, 包括馆际互借、参考咨询、数据交换、云盘、教学参考以及特色库, 并且提



图3 CALIS国家中心级云服务平台

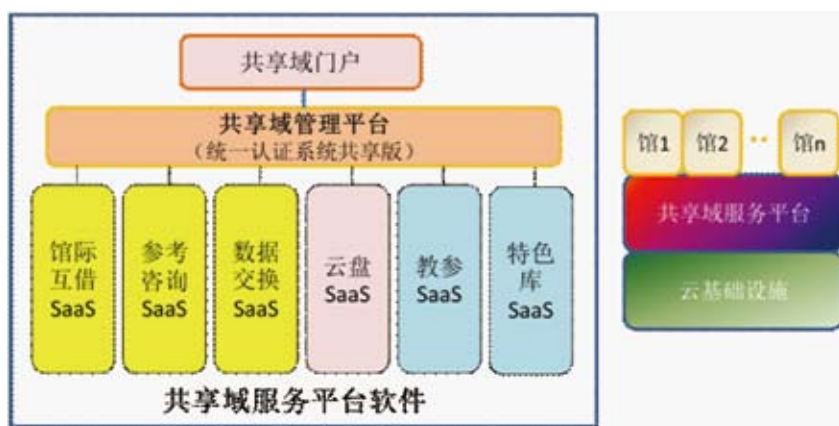


图4 CALIS共享域级云服务平台

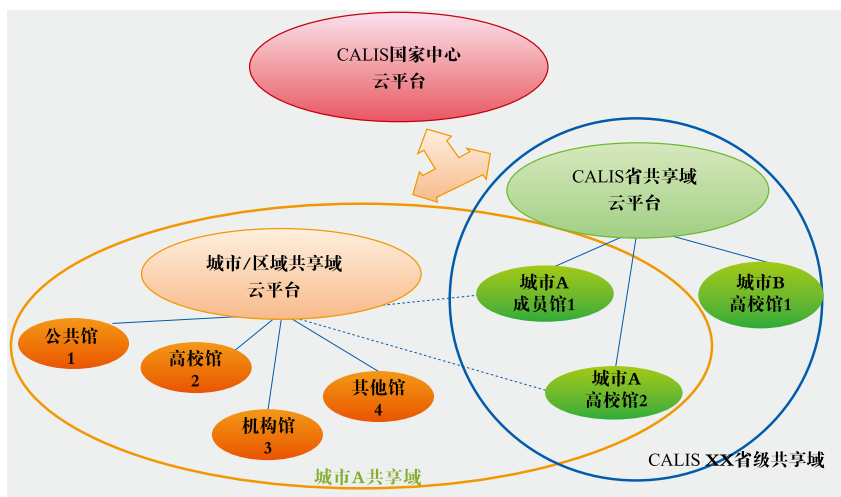


图5 采用CALIS云平台的城市/区域共享域

供能整合这些服务的共享域门户。共享域中心可根据自身特点对共享域门户进行定制改造,以便更好地整合其他服务。共享域的成员馆可以免费使用CALIS共享域级云服务中心提供的各类服务,无需安装部署。

基于CALIS中心级云服务平台和共享域级云服务平台,区域中心很容易构建能整合高校和非高校机构的城市/区域级的资源共享服务平台,该共享平台很容易实现各类信息服务机构(包括高校图书馆、公共图书馆和其他企事业单位或机构)的服务与资源整合,促进城市或区域级共享域云服务平台的建设以及资源、服务和知识的共享,实现各馆之间的资源互补。

总体而言,CALIS共享域服务模式具有如下的优势:1)一个账号、域内访问、全国漫游:即在共享域中使用一个账号(比如图书馆原来的图书证卡或一卡通账号)即可实现全域访问,并可全国漫游;2)SaaS服务模式,部署简单,使用方便,成效显著:只需在实体共享域中心部署和维护云服务平台,各馆/机构开通使用,即开即用,专注于业务服务,降低技术维护成本,推广迅速;3)轻松整合域内各馆的资源与服务,部分馆可以在其他实体共享域开通,只需接入,无需重复改造;4)整合域外其他优质资源,可整合上海图书馆、NSTL、国家图书馆(丰富的馆藏资源)、国内外各种全文期刊库(全文调度或馆际互借);5)增强成员馆服务能力,提高各馆参与积极性,成员馆采用CALIS云平台API很容易嵌入本馆门户,使得CALIS提供系统服务变成本馆服务一部分;6)可整合图书馆馆藏(电子、纸本等等),通过图

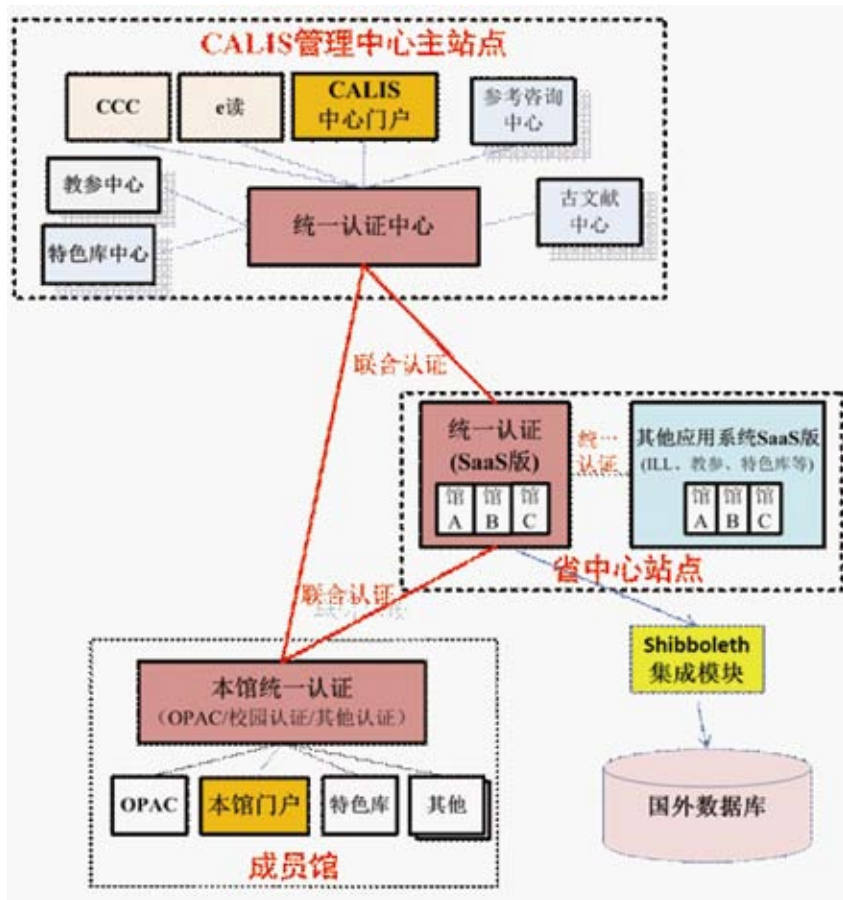


图6 CALIS统一认证服务体系

书馆与CALIS中心以及共享域中心服务的集成，图书馆用户可通过馆际互借、按篇付费、全文下载等方式分享其他成员馆和共享域中的资源，弥补本馆不足。

(3) 本地平台

对于应用情况复杂、规模大并且有一定技术力量的成员馆，可以采用本地部署部分CALIS应用系统的方式，整合本馆已有系统和第三方系统，整合CALIS国家级云中心（公有云）以及CALIS共享域/省级云中心（社区云）中的资源和服务后，构成一个面向本馆用户的混合云。

一般情况下，成员馆无需在本地安装CALIS系统，而是充分利用CALIS两级云服务平台提供的免费租用服务以及OpenAPI，与图书馆已有应用服务进行整合。这种整合可集中体现在本馆门户中。

2.2 统一认证

作为CALIS云服务体系中的基础服务，统一认证覆盖CALIS中心云服务平台、CALIS共享域云服务平台以及本地平台，CALIS中心统一认证系统、各个共享域统一认证系统和图书馆本地统一认证系统（或本地联合认证集成模板）构成一个完整的“一次登录，域内访问，全国漫游”的CALIS统一认证服务体系。

CALIS统一认证支持联合认证、代理认证和直接认证这三种方式。其中，直接认证方式和代理认证方式用于应用系统本身与统一认证系统之间的相互集成，直接认证适用于能够进行统一认证接口改造的系统，代理认证则是用在无法进行改造的系统。而联合认证用于两

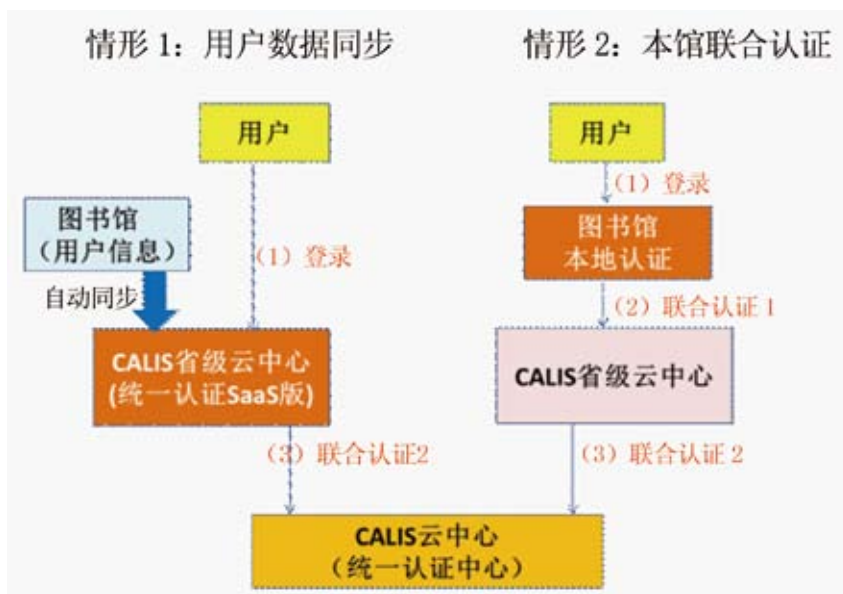


图7 集成统一认证的两种情形

个认证系统之间的相互认证。

从图书馆的统一认证集成角度看,主要有两种整合方式:用户数据同步和图书馆联合认证。1) 图书馆已有系统实施整合很困难的情形,可采用用户数据同步的方式,将用户数据定期自动同步或手动导入到统一认证SaaS版的租用系统中去;用户在登录时候直接登录CALIS统一认证租用系统。2) 对于图书馆可以对已有认证系统进行联合认证改造和集成的情形,采用系统联合认证方式,无需同步用户数据。当用户登录的时候,直接引导用户登录图书馆原有认证系统,然后由图书馆认证系统和CALIS云平台中的统一认证系统进行联合认证,此时用户便可访问共享域中心提供的应用和服务,无需再次登录;如果用户继续访问CALIS云中心,则共享域统一认证系统又会与CALIS中心统一认证进行第二次联合认证,最后引导用户回到所需访问的目标系统。整个联合认证过程对用户而言几乎是无感知和透明的。

2.3 资源交换与处理

对于大量不同来源和格式的数据,在CALIS原有

分散的图书、论文、期刊等资源采集工具的基础上,建立了统一资源交换平台(UES),建立高校分布式节点之间海量文献资源的共享交换体系,支持各类数据文件的快速交换,实现资源数据在CALIS成员馆、CALIS共享域与CALIS中心之间的双向传输,为CALIS其他应用系统提供资源数据汇集与分发的支持。

对于汇集后的数据,采用统一数据管理系统(DMS)进行进一步的处理(质量检查、查重比对合并、转换、收发、交换),形成新的资源集合,便于提供服务。目前处理支持的资源类型有中外文图书、中文期刊、中外文学学位论文、特藏资源(包括舆图、拓片、古籍、地方志、家谱)、会议论文、期刊论文、音频资料、网络资源、学位论文、电子书。

UES和DMS均采用Master/Worker架构,支持多机部署模式。其中UES支持热升级、断点续传、双向交换和多系统集成。统一数据管理系统数据处理任务可以并行执行,大大提高数据处理速度,并且支持快捷的适配器安装,通过组件化的安装/更新升级模式,可以快速地完成适配器的升级或安装工作。

本地平台中的UES、共享域云服务平台中统一交换分中心、CALIS中心云服务平台中的统一交换中心系统

以及相应的DMS,共同构成了完整的CALIS资源交换和处理体系。该体系解决了原来各类数据采集各自孤立、未经初步处理而富含大量错误、冗余数据导致中心压力过大、效率低下的问题,使得数据采集易于使用、便于交互、管理及数据的快捷汇总和处理,提供了有效的数据汇集和交换功能以及完善的数据质量检测控制功能。

2.4 资源与服务整合

(1) 资源与服务的双向整合

在CALIS多云服务体系中,资源与服务的整合是双向的,即一方面CALIS依托各馆进行资源共建,整合各馆和第三方的资源与服务,在各馆“共建、共享”的共识下又在各馆之间进行共享;另一方面各馆也整合CALIS多云服务体系中提供

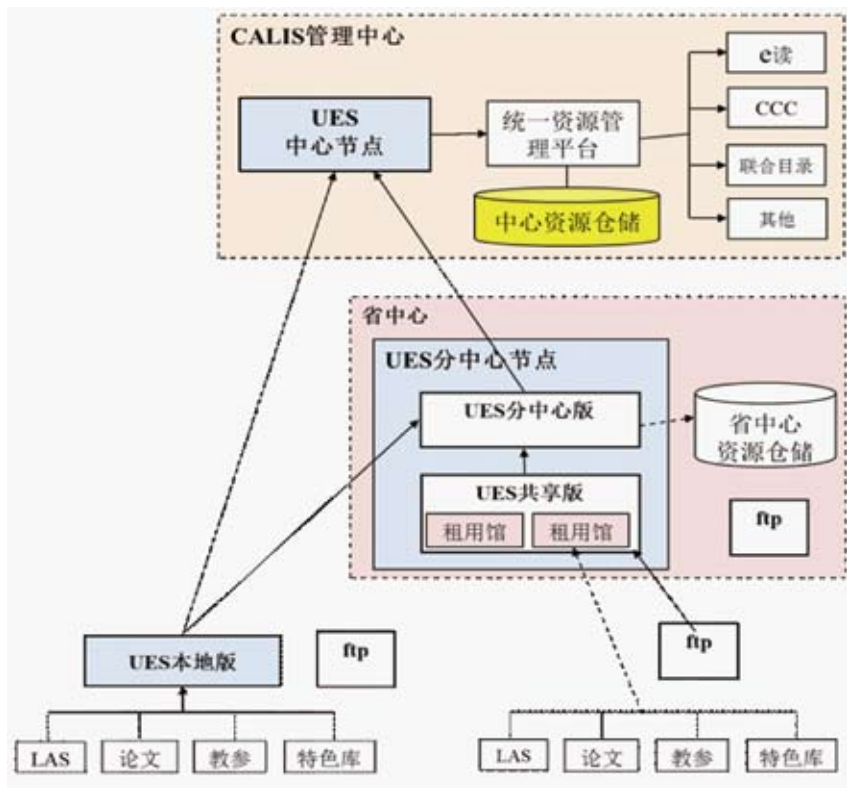


图8 CALIS资源交换与处理体系

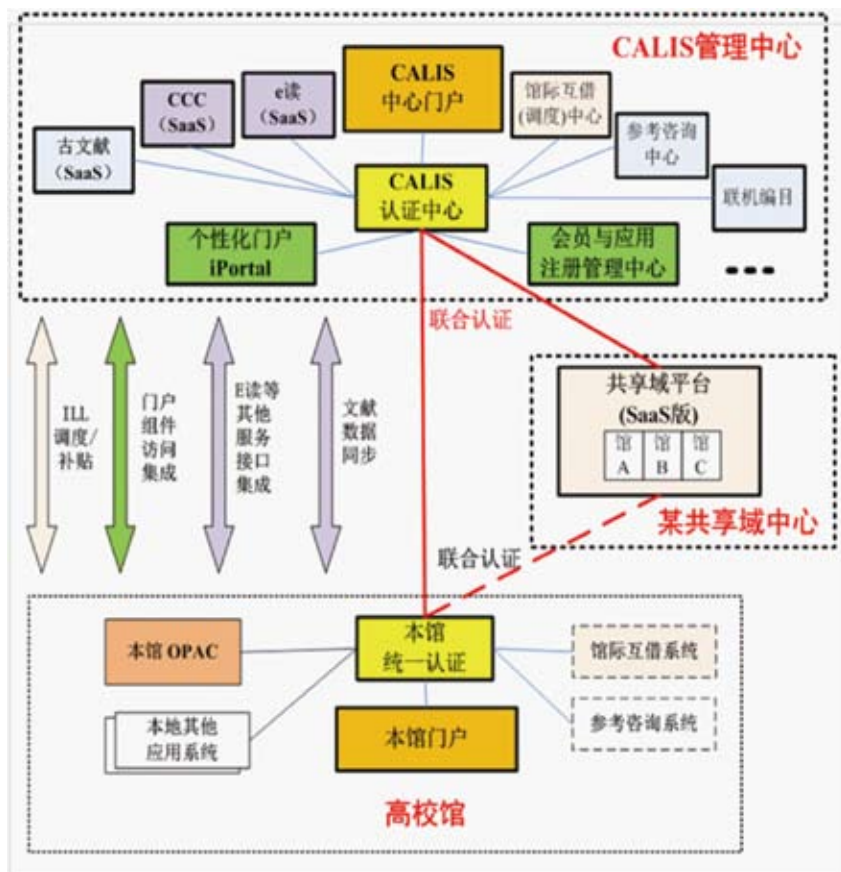


图9 资源与服务的双向整合

的各种资源和服务,这些资源和服务包括各馆共建共享的、联合采购外,还有第三方授权许可范围内的。最终而言,资源与服务整合的目标在于为向读者提供的资源发现和获取服务提供有效的服务手段和技术保障。

各馆的纸本和电子的馆藏、电子资源元数据和购买国内外数据库的馆藏信息、本馆OPAC接口地址、电子图书与期刊访问地址和其他接口等各类信息通过CALIS资源交换与处理体系中的系统汇集和处理后,根据资源类型分发到CALIS云服务体系中的各类资源检索系统(e读、CCC、教学参考和特色库等等)中去。各馆的应用系统和服务接口则是注册到会员与应用注册系统中,作为应用和服务元数据信息,供需进行应用服务集成的系统使用。

在本馆整合CALIS云服务体系中的资源与服务过程中,图书馆的本地系统改造和集成包括:用户认证集成、文献数据同步、本馆门户网站集成改进(即集成CALIS中心云服务、共享域云服务)、本馆资源检索系统改进(即嵌入e读/CCC等服务,实现CALIS云服务深度集成)以及本馆系统API开发与集成(包括门户组件

开发、移动服务Apps开发等)。

(2) 资源与服务整合粒度

资源与服务的整合既包括浅层次、粗粒度整合(如网页链接、资源/服务介绍等),也包括深层次、细粒度整合(利用复杂的接口和互操作方式来实现)。

资源与服务整合深度(即对目标系统访问的深度)取决于接口的复杂度和用户访问权限。对于并无权限控制和其他限制的资源和服务,应尽可能提供细粒度层次的内容,对于有权限控制的资源与服务,在CALIS的统一认证体系中能够解决权限控制下的细粒度资源与服务整合问题,以便实现用户一次登录,域内访问各种资源与服务。

CALIS多云服务体系中,支持多种粒度的双向集成,包括粗粒度的(如系统网站链接方式的),也包括中粒度的(比如数据库导航接口、期刊信息接口等等),还有细粒度的(比如全文图书和文章篇目调度接口、篇目级别检索OpenAPI等)。

整合粒度可以是动态变化的。当一个目标系统的获取方式发生变更时,无法再按原有方式进行整合,需按新的方式重新注册和整合。在CALIS多云服务体系中,支持资源、应用与服务的注册、更新与通知,以适应目标系统的访问方式的变化。

整合粒度可以是动态变化的。当一个目标系统的获取方式发生变更时,无法再按原有方式进行整合,需按新的方式重新注册和整合。在CALIS多云服务体系中,支持资源、应用与服务的注册、更新与通知,以适应目标系统的访问方式的变化。

(3) 资源与服务整合方式

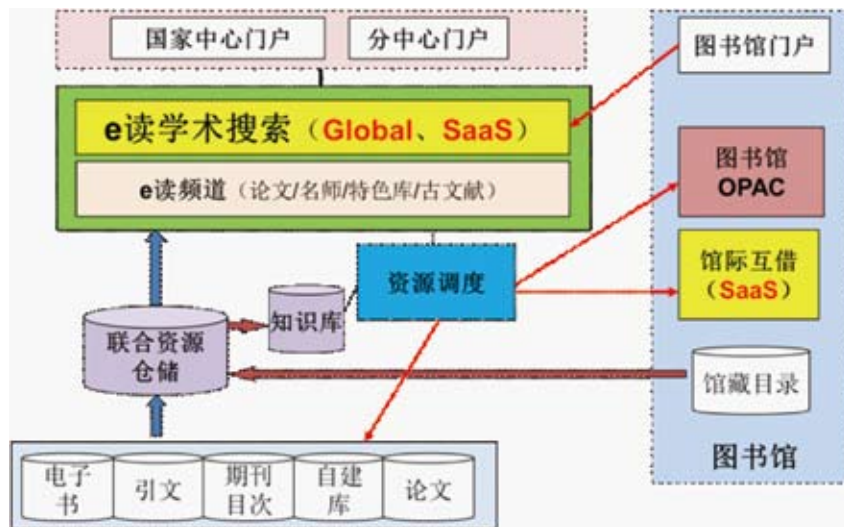
CALIS多云服务体系中,资源与服务整合的方式可以分为“直接整合”和“间接整合”两种:“直接整合”指通过服务界面模块可以直接到达目标资源系统或服务系统;“间接整合”指服务界面模块需要通过资源调度系统才到达目标资源或服务系统。

直接整合可以分为:导航链接方式(比如嵌入CALIS系统e读、e问等链接)、CALIS OpenAPI接口动态整合方式(采用CALIS OpenAPI后台调整合,或采用CALIS提供的AJAX接口Web端前台整合)、查询表单提交方式(比如内嵌e读的查询表单)以及混合方式(即综合灵活采用以上多种直接整合方式)。

间接整合则是通过资源调度提供的OpenURL接

口进行集成,通过资源调度系统,实现与目标资源与服务的间接整合。CALIS多云服务体系中,提供了两种资源调度的接口方式,包括典型的OpenURL链接方式和AJAX内嵌式调度接口方式。

直接整合与间接整合不仅仅应用在各馆整合CALIS多云服务体系中的资源与服务,也应用在CALIS多云服务体系内部以及CALIS多云服务体系整合各类第三方资源与服务的时候。



2.5 资源发现与获取

图10 CALIS资源发现与获取体系

CALIS统一认证体系、资源交换与处理体系以及资源与服务整合体系都为CALIS多云服务体系中的资源发现与获取提供了非常重要的基础。基于整合后的各类资源（电子书、期刊目次、学位论文、特色资源等等）的联合仓储检索技术，e读学术搜索2.0系统提供了一站式、多频道的资源发现功能，支持20多种资源类型的发现，包括图书、期刊、学位论文、网络资源、电子图书、会议论文、舆图、音频、地方志、家谱、期刊文章、古籍、拓片、图像、视频、名师、课件、手稿、国内课程、教参书、国外课程等。

e读学术搜索2.0系统不仅仅能对单个图书馆提供对各类资源（纸本馆藏、电子馆藏和其他特色资源）的整合和发布的一站式资源发现服务；也能为图书馆联盟提供对其各个成员馆之间的各类资源进行整合和共享的功能，并实现面向图书馆联盟的统一资源发现服务；还能部署在图书馆联盟，由其构建出自己的统一资源发现云服务平台，并能自行为其成员馆提供相应的云服务。

e读学术搜索2.0系统提供灵活的资源共享方式，包括单个图书馆能自行设置哪些资源库共享给图书馆联盟；图书馆联盟能自行设置成员馆的哪些资源库可以在本联盟中进行发布；并且提供具有丰富的智能化的搜索引擎功能、强大的定制能力（具有个性化的页面定制功能）和强大的集成能力（通过与图书馆其他系统集成，从而构成下一代图书馆的知识服务门户；通过与CALIS共享域云服务平台中的其他云服务系统集成，共同构成完整的面向图书馆联盟的数字图书馆云服务平台；与CALIS e读中心系统集成，能动态访问和获取

CALIS高校图书馆资源；能动态获取高校图书馆馆藏信息）。

在资源发现后，e读学术搜索2.0系统通过综合直接整合和间接整合（通过资源调度）这两种方式，根据资源的特点和用户权限信息，提供资源的全文链接、本馆/其他馆馆藏、章节试读、馆际互借功能，并且同时无缝整合了参考咨询、统一认证，为其他系统提供集成的OpenAPI，整体构成一个完整的CALIS资源发现与获取体系。

2.6 咨询与培训

在CALIS多云服务体系中，还提供了咨询与培训服务系统，作为整个服务体系的支持类服务，由e问、虚拟参考咨询共享版、考试培训以及信息素养教育平台组成。

其中e问提供了多馆联合咨询的在线问答、面向咨询馆员的课题服务工具箱、面向咨询馆员服务的咨询工具箱、面向读者的信息素养教育课件、查收查引与查新协作网、学科信息定制服务和重大课题咨询服务。

虚拟参考咨询共享版的定位在于迅速为单一图书馆搭建网络咨询服务平台，支持24/7咨询服务模式，并且构建多馆联合参考咨询服务体系，为单馆咨询服务提供后备支持。提供的主要功能包括馆员与读者交互（实时交互、异步表单咨询交互）、咨询问题知识库积累、读者参与解答（征答区，实现多方互动）以及省内多馆联合咨询（共享咨询员和共享知识库）。

考试培训系统面向的培训对象有共享馆、成员馆的人员;培训师资有管理中心、技术中心、省中心特邀专家;培训内容涉及业务管理、系统管理、技术开发;支持技术分享(程序代码、实施经验等方面的分享),提供在线培训、在线资格考试功能。

信息素养教育平台是提供信息素养教育网络的学习平台,用来打造信息素养教育“第一门户”,包括课程建设(课程体系设置、教学方法、教学大纲、师资队伍、教学评估),提供教学资料(课程教材、参考资料,支持与相关数据库的集成),展示教学过程(比如教学视频、作业平台),支持考试评测(比如题库、考试系统),并且为馆员和读者提供一个开放的学习环境。

3 结语

作为包含公有云、私有云、社区云和混合云的CALIS高等教育数字图书馆,是一个由多级云服务、统一认证、资源交换与处理、资源与服务整合、资源发现

与获取以及咨询与培训这六大体系组成的分布式数字图书馆系统。CALIS基于云计算的数字图书馆解决方案一方面切实解决了CALIS项目中成员馆数量多、差异大(资源、服务和人员技术)所导致的系统过多、安装部署复杂、维护困难等技术问题;另一方面也大大促进了高校图书馆之间的资源与服务共享,催生了高校图书馆的新型管理与服务模式,促进了馆员之间的交流与协作,促进图书馆联盟的建设和共享,大大降低了图书馆的设备和人员的投入。通过整合和利用CALIS各级云服务和相应的云计算解决方案,实现了高校数字图书馆的技术与服务创新。

随着信息爆炸式增长和复杂度的非线性上升,以及用户对于数字图书馆服务需求的细化和深入,高校图书馆必然会面临新的问题和挑战,CALIS高等教育数字图书馆的技术体系也将不断改进和完善,在云计算、知识处理和服务、大数据、关联数据、移动图书馆等方面继续深入研究和拓展服务,以更好应对未来的变化和发展。

参考文献

- [1] 王文清,陈凌.CALIS数字图书馆云服务平台模型[J].大学图书馆学报,2009,27(4):13-18.
- [2] WANG W, CHEN L. Building the New-generation China Academic Digital Library Information System (CADLIS): A Review and Prospectus[J]. D-Lib Magazine,2010 (5/6).
- [3] MELL P, GRANCE T.The NIST Definition of Cloud Computing [OL]. [2012-11-10]. <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>.

作者简介

王文清,男,教授级高工,CALIS管理中心总工程师。E-mail: wangwq@calis.edu.cn
张月祥,男,CALIS技术中心副主任。E-mail: zhangyx@calis.edu.cn
陈凌,男,研究馆员,CALIS管理中心副主任,北京大学图书馆副馆长。E-mail: chenl@calis.edu.cn

A Technical Overview of China Academic Digital Library Information System of CALIS

Wang Wenqing, Zhang Yuxiang, Chen Ling / CALIS Administrative Center, Beijing, 100871

Abstract: As one of the national projects of China's 11th Five-year Program, CALIS faced great challenges: building standardized, low-cost, adaptive, scalable digital library platform enabled with interoperability for nearly 2000 member libraries in China. With the development of cloud computing technology, CALIS proposed total solutions which is based on cloud computing technologies, and built multi-cloud service platforms including CALIS National Central Platform, Regional Platform and Local Platform. This paper gives a technical overview of China Academic Digital Library Information System including the CALIS cloud service platform, unified authentication system, resource exchange and processing system, resources and services integration system, resource discovery and access system, as well as consulting and training system.

Keywords: Digital library, Technology systems, Cloud computing, Cloud service, CALIS

(收稿日期: 2012-12-26)