

数字图书馆技术体系结构的研究历程 与特征分析

黄国彬¹ 郑霞¹ 孙一钢² 刘燕权³

(1. 北京师范大学政府管理学院, 北京 100875; 2. 国家图书馆, 北京 100081;
3. 美国南康涅狄格州立大学, 纽黑文 06515)

摘要: 数字图书馆技术体系结构承载着数字图书馆的功能实现, 决定了数字图书馆的发展范式。梳理数字图书馆技术体系结构的研究历程, 有利于认清其发展脉络, 为推动数字图书馆技术体系结构的更新与完善提供理论支撑。本文主要采用文献调研法、内容分析法梳理国内外学者有关数字图书馆技术体系结构的研究, 根据现有研究的内容侧重点, 划分数字图书馆技术体系结构的发展阶段, 并就每一阶段的研究特点进行深入剖析。将1992年至今的数字图书馆技术体系结构研究历程划分为四个阶段。第一阶段: 基于试验型项目的数字图书馆技术体系结构研究(1992—1998年)。第二阶段: 基于应用型项目的数字图书馆技术体系结构研究(1999—2003年)。第三阶段: 基于集成服务的数字图书馆技术体系结构研究(2004—2008年)。第四阶段: 基于云计算、智慧服务的数字图书馆技术体系结构研究(2009年至今)。

关键词: 数字图书馆; 技术体系结构; 研究历程

中图分类号: G250.76 **DOI:** 10.3772/j.issn.1673-2286.2021.03.006

引文格式: 黄国彬, 郑霞, 孙一钢, 等. 数字图书馆技术体系结构的研究历程与特征分析[J]. 数字图书馆论坛, 2021(3): 40-50.

数字图书馆相关研究最早可以追溯到20世纪90年代, 自此之后, 世界各国争相投入到数字图书馆研究浪潮中, 并取得许多令人欣喜的成果。现今数字图书馆研究已经走过近三十年, 然而图书馆人对其热情却有增无减, 相关研究成果的数量仍保持稳定增长趋势。相应的, 作为数字图书馆长久以来的研究重点, 以信息技术为支撑、以实现数字图书馆功能为导向的核心组件逻辑框架, 即数字图书馆技术体系结构, 也得到了研究人员的广泛关注与重视。

1 研究方法

本文采用文献调研法, 于2021年2月在CNKI、Web of Science、Scopus等数据库中进行中外文文献搜索。使用“数字图书馆”并含“体系结构”及其对应的单词“‘digital library’ AND ‘architecture’”进行标题检

索, 剔除涉及数字图书馆资源体系结构、服务体系结构和标准规范体系结构等非直接相关文献, 搜集20世纪90年代以来国内外涉及数字图书馆技术体系结构的研究文献。同时, 考虑到近年来在新兴技术推动下数字图书馆技术体系结构的发展与演进, 笔者又以“数字图书馆”并含“云计算”“大数据”“物联网”“增强现实”“虚拟现实”“人工智能”及其对应的英文表达进行标题检索, 获取2009年至今发表的相关文献, 最终共得到与数字图书馆技术体系结构相关的文献136篇, 其中中文文献70篇、英文文献66篇。

从文献检索结果可知, 国内外最早出现该主题研究论文的时间均为1997年, 且发文总量随着时间推移呈先增长后下降的趋势。其中, 2004—2008年国内外发文总量年均达到10篇左右, 反映出这一时期国内外学者积极探索并推进数字图书馆技术体系结构的相关研究工作。2012年以来, 国内外学者发表的关于数字图书馆体

系结构的论文数量逐年减少。这说明在数字图书馆产生初期学者对其技术体系结构研究的热情较高,而随着数字图书馆逐渐发展成熟,学者开始将研究视角聚焦到数字图书馆技术体系结构的具体模块构成。不过,作为数字图书馆技术架构的基础,数字图书馆的技术体系结构仍是一个值得探讨的话题,研究其发展沿革与技术演进对于实现新技术引进背景下数字图书馆技术体系结构的创新发展十分必要。

在此基础上,深入研读上述论文的研究角度与内容侧重点,以确定其阶段性研究特征。同时,考虑到数字图书馆的建设实践影响着学者对其技术体系结构的研究主题,本文进一步采用网站调研法和案例分析法,

通过选取国内外典型的数字图书馆原型项目与实践项目,从环境因素、设计理念、功能定位、设计原则、层次结构、基本组件、支撑技术等角度出发,总结其基本特征,为划分数字图书馆技术体系结构的阶段性研究特点提供依据,并结合数字图书馆研究与建设的发展阶段特征、国内外学者有关数字图书馆技术体系结构的研究主题演化,将1992年至今的数字图书馆技术体系结构的研究历程划分为四个阶段(见图1),以时间为线索系统梳理并深入分析每一发展阶段的研究成果与研究特点,以实现1992年至今的数字图书馆技术体系结构相关研究的全面扫描。

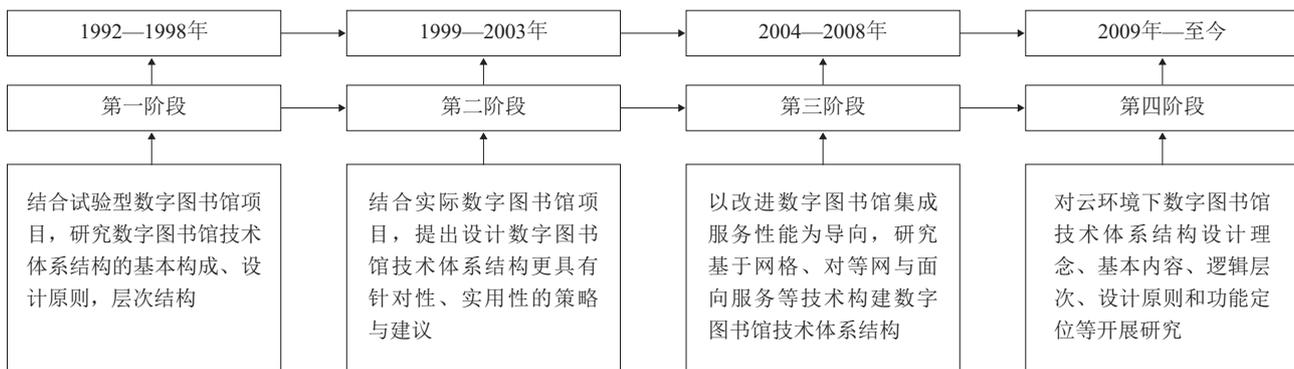


图1 数字图书馆技术体系结构研究的四个发展阶段

2 第一阶段: 基于试验型项目的数字图书馆技术体系结构研究(1992—1998年)

2.1 发展概况

从数字图书馆的研究历程来看,最早开启数字图书馆研究的是英国,1993年开启电子图书馆(Electronic Library)项目的研究^[1]。随后美国意识到这一领域的重要性,紧接着于1994年由美国国家科学基金委(National Science Foundation, NSF)资助了数字图书馆(Digital Library)项目的研究与建设,即美国数字图书馆先导计划一期工程(1994—1998年)。根据美国数字图书馆专家Arms的研究,1992年启动、1996年结项验收的“计算机科学技术报告项目”(Computer Science Technical Reports Project, CSTR)^[2],将数字图书馆技术体系结构纳入其研究范围,并通过Kahn等^[3]于1995年发表的《分布式数字对象服务框架》,将早期的数字图书馆技术体系结构概括为用于管理网络环境下各种数

字资料的数字对象、用于识别数字对象和其他网络资源的句柄、用于存储数字对象的存储库3个组成部分。从这一角度看,有关数字图书馆技术体系结构的最早研究时间可以追溯到1992年。

美国数字图书馆先导计划一期工程结束后,美国在此研究成果的基础上,于1999年启动了数字图书馆先导计划二期工程,并围绕以人为中心、以资源为中心和以技术为中心,深化数字图书馆技术体系结构的研究与应用,从立足于试验型数字图书馆项目、初步应用于建设型数字图书馆项目,拓展到深入结合试验型数字图书馆项目、广泛应用于建设型数字图书馆项目。基于上述分析,可以将1992—1998年,作为数字图书馆技术体系结构研究的第一阶段。

2.2 代表性研究观点

Arms认为,灵活的信息组织是任何数字图书馆的一个核心设计问题。20世纪90年代初期,Arms就开始

和美国国会图书馆的国家数字图书馆项目(National Digital Library Project, NDLP)成员合作,开展数字图书馆技术体系结构方面的研究,以构建一个实验性的系统来组织和存储图书馆的各类馆藏,并于1995年进一步提出构建数字图书馆技术体系结构需要考虑的八项原则^[4]:在社会、法律框架中设计数字图书馆技术体系结构框架;使用规范化的定义对技术体系结构中的术语进行详细说明;从技术上把数据、结构和服务分离开;名称和标识符是数字图书馆的基本构件;数字图书馆馆藏资源并不局限于“位”字节;投入使用的馆藏资源与存储的馆藏资源不同;存储库须全面管理馆藏资源;用户希望获得智能化资源而非数字对象。这八项原则已发展成为第一代数字图书馆研究型项目或应用型项目在构建技术体系结构时普遍考虑的标准。

经过早期的实验开发,到1995年,部分数字图书馆项目已开始使用Kahn等专家提出的理论与体系,开发并构建完整的数字图书馆系统。与此同时,Lagoze等^[5]深入探讨数字图书馆技术体系结构的设计理念,并对数字图书馆存储库的定位与功能做了全面描述。同年,Lagoze等^[6]合作完成的技术报告《互操作安全对象存储的设计》则成为首个数字图书馆存储库设计与部署的理论指南。而在数字图书馆技术体系结构另一重要组件即句柄的研究方面,有关学术论文与技术报告在1992年就已出现^[7]。

1996年,Arms^[8]突破之前有关学者关于数字图书馆技术体系结构的研究立足于试验型数字图书馆项目的研究局限,开始以美国国会图书馆启动的NDLP项目为研究对象,深入分析其技术体系结构的要素^[9]。作为数字图书馆建设早期在世界范围内较有影响力的大型项目,该项目的建设目标是将美国国会图书馆收藏的具有历史价值的文物与资源数字化,从而满足大众在线访问需求。因此,对于早期以资源数字化并提供在线访问为建设重点的数字图书馆项目而言,其技术体系结构的设计具有较高的实际参考价值。而美国国会图书馆也意识到这一点,将有关该项目的技术体系结构的研究资料加以系统整理,并对外发布,以此带动了全球数字图书馆的建设与发展^[10]。

1997年,Arms等^[11]在其有关数字图书馆技术体系结构奠基性研究论文中指出,数字图书馆技术体系结构主要由用户界面、存储系统、句柄系统和检索系统等构成。

1998年,Payette^[12]就数字图书馆技术体系结构的

研究意义、基本构成和以服务为导向的数字图书馆技术体系结构等开展研究,指出数字图书馆技术体系结构的实际意义在于解决富文本模型,即数字对象的存储与管理、永久与唯一的命名机制、定义清晰的数字图书馆服务、资源发现的便利性、数字图书馆馆藏的管理、分布式内容与服务的管理、知识产权的管理等问题。

2.3 研究特征

概括来看,在数字图书馆技术体系结构研究发展的第一阶段(1992—1998年),相关研究主要呈现出以下特征。

(1) 研究素材来源较为集中。这一时期国内外有关数字图书馆技术体系结构的研究,多基于英、美等国家规划或启动的数字图书馆建设实践项目进行案例分析和归纳,试图发现数字图书馆技术体系结构的设计理念和未来发展。且从相关研究成果的影响力来看,美国数字图书馆研究与建设主体产出的科研成果占据重要地位。

(2) 研究重点突出。虽然在此阶段国外相关学者对数字图书馆技术体系结构的研究内容较为广泛,但也有较为明显的研究重点,如强调数字图书馆技术体系结构的构成要素、功能定位、建设原则等内容。这与数字图书馆在此时期的发展特点存在较大关联。早期数字图书馆建设的核心目标是通过促进数字资源的收集、存储、处理和组织,以便于用户通过网络对其进行查询、检索和使用,与数字图书馆有关的服务架构尚处于发展和完善阶段。而作为数字图书馆信息系统的构成基础,数字图书馆技术体系结构的实践探索主要聚焦在宏观设想、内容构成的研究层面。

(3) 研究层次不断向着纵深发展。随着数字图书馆建设项目的不断成熟,学者对数字图书馆技术体系结构的研究也逐渐从理论探讨深入到实践分析的层面,即突破传统基于试验型数字图书馆项目研究的局限,开始以应用型数字图书馆建设项目为研究主体,相关研究理论推陈出新,也进一步促进了数字图书馆技术体系结构的实施建设与设计开发。

(4) 国内外研究水平存在较大差距。与国外学界或业界对数字图书馆技术体系结构的研究相比,可以说,在数字图书馆技术体系结构第一阶段的研究中,我国学者并没有过多的参与。这一方面缘于我国数字图书馆项目在此时期的建设仍有待深入,另一方面是因为我

国数字图书馆建设者与研究者仍未能充分了解数字图书馆技术体系结构的实质与价值, 故对其研究基本处于待开垦阶段。

3 第二阶段: 基于应用型项目的数字图书馆技术体系结构研究 (1999—2003年)

3.1 发展概况

数字图书馆技术体系结构研究发展的第二阶段为1999—2003年。在这一研究阶段, 数字图书馆技术体系结构领域较为突出的研究成果是美国数字图书馆先导计划二期工程 (DLI-II, 1999—2004年) 的研究。DLI-II将拟资助的数字图书馆项目划分为3个子领域: 以人为中心的数字图书馆项目, 以资源为中心的图书馆项目, 以信息技术为中心的图书馆项目。根据不同的研究定位, 资助不同的研究型数字图书馆项目, 并由此衍生出不同功能定位与层次结构的数字图书馆技术体系结构。

在这一发展时期, 世界范围内围绕数字图书馆技术体系结构的理论研究、试验型数字图书馆项目的研究与应用型数字图书馆项目的建设不断得到深入推进。与此同时, 我国数字图书馆界在前期数字图书馆的理论研究、试验型或研究型数字图书馆的探索以及借鉴国外数字图书馆研究与建设经验的基础上, 开始全面启动建设型或工程型数字图书馆项目的建设, 学术界也开始进一步对数字图书馆技术体系结构给予高度重视, 研究的关注度明显提升。此时段国内外围绕数字图书馆技术体系结构研究的突出特点体现为: 结合应用型数字图书馆项目, 提出更具有针对性、实用性的数字图书馆技术体系结构设计策略与建议。

3.2 代表性研究观点

1999年, 来自国际技术研究学会 (International Technology Research Institute, ITRI) 世界技术部的学者研究指出^[13], 数字图书馆技术体系结构应该有利于: 为数字图书馆系统提供采集、编目、存储、搜索、保护和检索信息的集成服务; 可使海量的数据有序化并改变由于信息管理不善造成许多重要的信息出现不连贯、不可用或丢失的现状; 数字图书馆技术体系结构应注重组织、采集、保存和利用各种信息; 数字图书馆技术

体系结构可方便地部署在某个特定的硬件、网络和软件上。同时指出, 数字图书馆技术体系结构由数字图书馆操作体系结构、技术体系和系统架构构成。

1999年, 根据NSF的合作协议ENG-9707092, 由NSF和美国国防部高级研究计划局 (DARPA) 资助, 并由美国马里兰州洛约拉学院国际技术研究院开展研究的数字图书馆技术体系结构项目课题组发表研究成果。该课题组对日本早期启动的数字图书馆的技术体系结构、系统和软件组件等开展深入调研^[13]。同年, 国内学者郑巧英等^[14]结合上海交通大学数字图书馆的建设, 介绍该原型项目的技术体系结构, 开创了我国学者围绕数字图书馆技术体系结构开展研究的历史先河。

2000年3月, 加利福尼亚数字图书馆项目组发布《数字图书馆技术体系结构与标准》研究报告^[15]。同年, 国内学者进一步围绕数字图书馆技术体系结构开展研究。如欧洁等^[16]分析了数字图书馆的数字对象体系结构, 张健等^[17]研究了开放体系分布式数字图书馆原型设计, 王军等^[18]初次较为全面地介绍了数字图书馆技术体系结构的主要类型与支撑技术。

2001年, Suleman等指出^[19], 需要建立一个开放的数字图书馆技术体系结构, 以支持组件的内部封装。同年, 孙一钢^[20]首次提出数字图书馆“技术体系结构”的确切表达。与此同时, 郑彦宁^[21]也进一步研究了具有集中索引的数字图书馆系统结构。

2002年, 美国学者Janéc等^[22]在研究亚历山大数字图书馆项目 (Alexandria Digital Library Project) 技术体系结构的报告中指出, 亚历山大数字图书馆技术体系结构包括组件、数据模型、服务和用户界面。同年, 我国学者真漆^[23]、马自卫^[24]、王军^[25]、盛小平^[26]对数字图书馆技术体系结构的研究进一步深入, 发表了一系列较有影响力的论文; 而陈梅^[27]的博士学位论文《数字图书馆体系结构参考模型》, 将我国数字图书馆学界对数字图书馆技术体系结构的研究进一步推向成熟。

进入2003年, 来自印度统计研究院 (Indian Statistical Institute) 的Pandey^[28]研究指出, 数字图书馆是一个复杂系统, 一个数字图书馆可以具有多层的技术体系结构, 不同的数字图书馆具有不同的技术体系结构。同年, 刘炜等^[29]在数字图书馆的体系结构与元数据领域开展了深入研究, 孙魁明等^[30]发表了《数字图书馆体系结构的研究与实现》, 牛振东等^[31]梳理了数字图书馆体系结构的发展, 叶鹰等^[32]进一步构建了数字图书馆的体系

结构与理论模型。值得一提的是,我国学者韩毅^[33]根据网格技术的特点及其对数字图书馆系统可能带来的影响,在国内首次发表了在网格技术平台上如何设计与部署数字图书馆技术体系结构的研究论文。

3.3 研究特征

总的来说,在数字图书馆技术体系结构研究发展的第二阶段(1999—2003年),相关研究主要呈现出以下特征。

(1) 在研究方法中更多应用案例分析进行归纳总结。基于上述分析可知,在数字图书馆技术体系结构研究的第二发展阶段,随着各类新概念、新技术和新工具引入到数字图书馆项目建设实践中,国内外有关数字图书馆技术体系结构的研究也积极引入数字图书馆项目工程,如开始关注到日本早期启动的数字图书馆、加利福尼亚数字图书馆项目、亚历山大数字图书馆等发展前景广阔的数字图书馆的技术体系结构,从而通过案例分析法从中归纳提炼出数字图书馆技术体系结构的基本构成、设计策略以及功能定位等内容。同时,且与第一阶段研究特征类似,以美国为代表的数字图书馆研究与建设主体依然是此领域的世界引领者。

(2) 国内研究迎来高速发展阶段。总体而言,相较于第一阶段,在数字图书馆技术体系结构第二阶段的研究中,我国学者已高度意识到数字图书馆技术体系结构的重要性,并给予充分重视。这主要是因为,随着我国大规模数字图书馆建设项目的立项与实施,解决数字图书馆技术体系结构的问题,已成为数字图书馆建设的首要环节。可以发现,自2001年开始,我国学者对数字图书馆技术体系结构的研究进入一个高速发展时期。这股研究热潮,一直持续至今。

(3) 研究内容更加广泛。与第一阶段数字图书馆技术体系结构的研究内容相比,第二阶段对于数字图书馆技术体系结构的研究内容更加广泛,包含数字图书馆技术体系结构的基本概念、核心特点、功能定位、构成要素、突出优势、建设标准、实施规范、支持技术以及服务形式等多个维度。进一步结合数字图书馆的发展特点来看,在此期间数字图书馆的应用领域、建设规模、服务内容及运作形式等方面都取得了极大拓展,促使不同功能定位和结构层次的数字图书馆技术体系结构不断出现和衍生,为学术界从不同维度探索数字图书馆技术体系结构提供了基础资料和参考依据。

(4) 研究成果更具针对性和实用性。值得注意的是,由于在数字图书馆技术体系结构研究的第二阶段期间,各类数字图书馆建设项目也在如火如荼地开展,因此学者对数字图书馆技术体系结构的研究更加注重实践与应用,研究既有针对正在投入使用的数字图书馆技术体系结构的现状总结,也有对未来数字图书馆技术体系结构的设计优化,这对后期数字图书馆技术体系结构的多技术融合、多学科交叉的纵深发展奠定了重要基础。可以说,数字图书馆技术体系结构研究的第二阶段产出的研究成果更具实用价值。

4 第三阶段: 基于集成服务的数字图书馆技术体系结构研究(2004—2008年)

4.1 发展概况

就在美国数字图书馆先导计划二期工程结束的同一年,即2004年,世界数字图书馆界有关数字图书馆技术体系结构的研究进入了一个新的发展时期。之所以将2004年划定为数字图书馆技术体系结构研究第三发展阶段的起始年,缘于从2004年开始,国外开启了以网格为平台的数字图书馆技术体系结构项目的研究,同时开展了以网格为支撑技术的数字图书馆试验型项目的建设。

4.2 代表性研究观点

首先,在理论研究方面,2004年,Knežević等^[34]指出,下一代数字图书馆的技术体系结构将是一种全新的技术体系结构,将由集成的、集中控制的数字图书馆系统转向可动态配置的数字图书馆联盟。面向服务的体系结构(SOA)、网格技术以及对等网(P2P)技术为数字图书馆的这种转变提供了技术基础。此类技术使得下一代数字图书馆更加开放、具备更强的适应性,数字图书馆的集成性将更强,无须集中控制就能以动态的方式发挥数字图书馆功能。同年,Petinot等^[35]提出了基于SOA的数字图书馆技术体系结构构想,设计了新一代的更开放式数字图书馆系统——CiteSeer,以更加方便快捷地满足21世纪多变、复杂的用户需求。

其次,在学术会议方面,2004年6月在意大利召开了“数字图书馆的体系结构: P2P、网格和面向服务”的学术会议。德国学者Knežević等^[36]指出,借助网格、

P2P与面向服务等技术构建下一代数字图书馆技术体系结构已成为可能。

最后,在试验型项目方面,于2004年1月启动DELOS (DELOS Network of Excellence on Digital Libraries)项目由欧盟第六框架计划赞助,该项目通过联合欧洲从事下一代数字图书馆技术开发的主要研究团队力量,合作开发数字图书馆通用技术并进行整合,尝试获得基于网格平台的数字图书馆技术体系结构的设计策略。同年9月,由欧盟资助630万欧元的DILIGENT项目也开始启动,其研究目标定位于创建一个由网格技术支持的先进数字图书馆技术体系结构。

自此,国内外学者开始以改进数字图书馆集成服务性能、降低运行成本、高效部署系统架构为导向,探讨新一代数字图书馆技术体系结构的设计与改进。需要指出的是,在2004年,与国外学者热切研究基于网格的数字图书馆技术体系结构相比,我国多数学者仍普遍关注属于数字图书馆技术体系结构第二阶段的研究议题。如李爽^[37]分析了数字图书馆体系结构的设计原则与系统构成,王亮^[38]在其博士论文中研究了数字化图书馆基本体系及实现,龚萍^[39]则在其博士学位论文中探讨了第三代数字图书馆体系结构分析及其建设中的知识产权问题,王东霞^[40]、赵芳等^[41]就数字图书馆技术体系结构的设计原则、建设内容、功能定位开展研究。不过,此时期我国学者围绕网格与数字图书馆技术体系结构的研究已开始萌芽^[33]。

在2005年, Ioannidis等^[42]依托于2004年投入建设的DELOS (项目时间为2004—2007年),将该项目于2005年召开的专题研讨会的会议论文,以《未来数字图书馆的管理系统:系统架构与信息访问》为题,编辑出版了一本论文集,详细介绍了DELOS项目的研究与建设内容。自2005年起,我国学者开始广泛结合网格、SOA、P2P等技术,分析在这些技术影响下数字图书馆技术体系结构的设计与部署问题。包括董慧^[43]出版专著《数字图书馆网格应用模型研究》、梁小吟^[44]以《基于组件的数字图书馆系统体系结构的研究》为硕士论文选题、张兴军^[45]则围绕网格支持的数字图书馆研究开展博士后研究工作。

到了2006年,以Trnkoczy等为代表的数字图书馆研究学者,进一步探讨网格环境下数字图书馆技术体系结构的设计。Trnkoczy等^[46]研究指出,当前基于网格技术建设的数字图书馆迅速发展,而基于网格的数字图书馆技术体系结构,应允许基于OAI-PMH协议的并行

收割,支持收割节点的动态配置,以及收割任务的动态调度。同年,毕强等^[47]研究指出,体系结构或者系统设计是数字图书馆运行的核心,剖析了该时期国内学者围绕网格研究数字图书馆技术体系结构的主要特点。

2007年, Candela等^[48]结合网格在数字图书馆的应用,指出网格环境下的数字图书馆技术体系结构应该能够融合不同数字图书馆之间的所有功能,还可以对满足特定场景与特定服务的子系统进行有机整合。从总体上看,网格环境下数字图书馆技术体系结构应分解为数字图书馆、数字图书馆系统、数字图书馆管理系统3个基本模块。

2008年, Artini等^[49]在DELOS建设的基础上,提出基于网格的DRIVER技术体系结构。

4.3 研究特征

基于上述分析可知,在数字图书馆技术体系结构研究发展的第三阶段(2004—2008年),相关研究主要呈现出以下特征。

(1) 研究形式更加多样。基于前述分析可知,在数字图书馆技术体系结构研究的第三发展阶段中,国内外围绕基于网格等新兴技术在图书馆技术体系结构中的应用开展了诸如理论探讨、学术会议、试验项目等多种研究形式,且产出了众多令人欣喜的研究成果,这对于前沿信息技术在数字图书馆技术体系结构中的应用研究意义重大。

(2) 新技术的应用对研究选题产生较大影响。综合这一研究阶段国内外有关数字图书馆技术体系结构的研究可以发现,这一时期新技术对数字图书馆建设的影响不断加强,包括SOA、P2P、网格、物联网、移动技术、泛在技术和Web2.0技术等,有效解决了数字图书馆资源建设中的信息组织、统一检索和知识服务等问题,将不同领域、载体中的数字化资源以网络化的形式整合和联结起来,进而推动数字图书馆技术体系结构实现异构性、分布式和高度自治的网格特点。在这类新技术的影响下,数字图书馆技术体系结构的集成化、动态化的发展特征描述成为此阶段的主要研究选题。可以说,这一阶段的研究更集中于从理论上探讨如何在新技术环境下设计数字图书馆技术体系结构,或者是启动试验型项目研究新兴信息技术与数字图书馆系统结合的可行性,如欧洲的DELOS项目、DILIGENT项目。

(3) 研究主体发生变化。与上述数字图书馆技术体系结构的前两个发展阶段不同,基于对这一领域研究的启动时间与关注程度,在数字图书馆技术体系结构的第三研究阶段,来自欧洲主要国家(如意大利、德国等)数字图书馆领域的研究与建设主体,无论是在基础理论研究,还是在试验型项目的研究方面,都比美国同行更积极投入。这反映出世界各国对于新技术应用环境下数字图书馆技术体系结构的探索热情空前高涨,以及信息技术在数字图书馆技术体系结构应用过程中的巨大发展潜力。

(4) 国内外研究主题类似,但国内更侧重基础理论研究。概括来说,在数字图书馆技术体系结构研究的第三发展阶段中,国内外相关学术研究主题基本一致,均围绕网格、SOA、P2P和Web2.0等技术阐述数字图书馆技术体系结构的创新发展、标准协议、设计部署、技术应用等问题。与国外相比,这一时期我国多数数字图书馆建设项目、应用项目进入二期建设,其建设理念如何明晰、服务功能如何定位、用户需求如何跟踪,都直接影响到数字图书馆技术体系结构的设计,此时期虽然我国基于新兴信息技术来建设的数字图书馆应用项目数量极少,但我国学者开展数字图书馆技术体系结构的基础理论研究仍然相当活跃。

5 第四阶段:基于云计算、智慧服务的数字图书馆技术体系结构研究(2009年至今)

5.1 发展概况

在欧洲主要国家于2004年率先启动以网格、SOA和P2P技术为支撑的下一代数字图书馆技术体系结构的研究之后,进入2009年,美国以OCLC为代表的数字图书馆研究者或建设者,开始富有战略眼光地探讨云平台环境下、面向大数据处理的数字图书馆技术体系结构的设计与建设问题。可以说,有关数字图书馆技术体系结构的引领性研究,再次由美国“接棒”。虽然在此阶段出现了包括云计算、大数据、物联网、增强现实和虚拟现实等在内的多项革新性技术,但相关研究多集中于探讨云计算技术给数字图书馆技术体系结构带来的影响。

云计算由Google于2006年提出,经过几年的发展,目前已经得到越来越多的厂商与用户的认可和追捧,云计算被提升到前所未有的高度^[50]。有媒体认为:

2010年是云计算元年^[51]。事实上,早在2009年,图书馆界就开始探讨如何在云计算平台上部署、设计适用于云服务技术环境的数字图书馆技术体系结构。

可以说,2009年,Google、Amazon等商业云在美国图书馆的应用,以及OCLC的“Web级协作型图书馆管理服务”和美国国会图书馆DuraCloud项目的推出,标志着图书馆领域最早的云服务在美国诞生。作为数字图书馆领域建设与研究的领跑者,美国图书馆界在云计算领域的研究与搜索,也标志着数字图书馆界迎来了云的时代,并朝着大数据处理与服务的方向迅速发展。而2009年9月中国图书馆学会学术委员会举办的“云计算与图书馆”专题研讨会在上海召开^[52],则标志着我国图书馆界正式涉足云计算领域。

5.2 代表性研究观点

2010年,来自OCLC的Goldner^[53]在其有关云计算与数字图书馆的研究报告中,概括出云计算环境下数字图书馆技术体系结构需要重新思考的7项理由,包括信息查找者只能在一个公共网络环境下查找信息,而且分布式系统使得其难以将数字图书馆纳入其工作流程中。2011年,OCLC提出云计算环境下数字图书馆技术体系结构由5个模块构成,即物理环境管理、人员管理、网络安全管理、灾难数据备份、恢复与销毁管理、隐私保护与第三方政策。

值得注意的是,在此发展阶段我国学者也开始积极探索云计算环境下数字图书馆系统的理论构建和宏观架构。早在2010年,叶小榕等^[54]就设计并开发了一项结合云计算和PKI技术的数字图书馆系统框架,数字图书馆用以提供信息资源,云计算技术用于提供整体支撑框架,PKI用于提供多层次安全服务模型。随后,马晓亭等^[55]探讨了数字图书馆云存储应用系统的理论构建与应用原理,认为在建设数字图书馆云存储方案的过程中,应当遵循系统设计的安全性原则、保证数据存储和用户访问效率原则、确保存储区域网络健壮性原则以及云存储可操作性原则等。赵美凯^[56]以云计算和虚拟化技术为理论依据,在Eucalyptus开源云计算平台基础上,构建了数字图书馆私有云基础架构。胡昌平等^[57]基于移动图书馆云平台的基本架构和终端系统特点,提出云环境下移动数字图书馆应当由门户层、应用层、管理层和适配层4个基本模块构成。张艳等^[58]从资源检索的角度出发,基于开源云计算Hadoop技术,

设计了专门针对数字图书馆的云检索系统架构。

随着大数据、物联网、增强现实、虚拟现实、人工智能等技术与方法的发展,也有学者从上述角度出发,研究数字图书馆技术体系结构的整体性设计或子模块创新。比如,王宁等^[59]从大数据的特征及其与数字图书馆的结合出发,构建了符合用户需求特征的数字图书馆知识发现系统。刘传玺等^[60]阐述了智慧数字图书馆的概念、特征及设计原则,提出智慧数字图书馆的架构设计。石婷婷等^[61]提出引入数字孪生技术,如何设计智慧图书馆的应用场景,并据此设计了智慧图书馆的体系架构。

从发展阶段和功能定位上看,图书馆管理系统、图书馆资源管理系统均是传统图书馆向数字图书馆发展过程的产物,只有部署这些系统的图书馆,能够基于互联网提供服务,并成为新时期数字图书馆技术体系结构研究的来源素材。

5.3 研究特征

概括来看,自2009年以来,数字图书馆技术体系结构研究发展的第四发展阶段中,相关研究主要呈现出以下特征。

(1) 较关注理论探讨,而基于应用案例的实证研究稍显不足。云计算与云服务的实现,需要更先进发达的网络基础设施、相关的信息安全政策与法律的配套支持,显然,在这方面,我国仍有一段路要走。因此,即便国内数字图书馆学界已从理论上对云环境下数字图书馆技术体系结构的设计理念、基本内容、层次结构、设计原则和功能定位等开展了较为全面的研究,甚至也有部分研究型或试验型项目结合这一命题开展研究,并取得一定的研究成果,但因为建设项目或应用项目的缺乏,使得迄今为止我国围绕云计算技术探讨数字图书馆技术体系结构的设计问题等研究选题在很大程度上显得实证性不强。类似的问题同样摆在国际数字图书馆界学者和业者面前——虽然国外数字图书馆在应用云计算技术方面比国内走得更远,但真正基于云计算与云服务理念而建立起来的大型应用性数字图书馆项目,或者是有影响力的试验型数字图书馆,目前仍较为缺乏。

(2) 研究结构复杂并形成新的研究体系。与前三个发展阶段相比,数字图书馆技术体系结构研究的第四个发展阶段所涉及的研究成果呈现出相互渗透、交

叉融合的研究态势,研究结构具有一定的复杂性,从数字图书馆技术体系结构的基础设施、软件服务,到资源与服务集成整合,再到资源共享和馆际互借,最后到虚拟参考咨询与共建共享,逐层深化,形成云计算环境下资源、管理、技术、服务等全方位融合的数字图书馆技术体系结构的新研究体系。

(3) 研究成果推陈出新,不断发展。近年来,随着云计算的进一步发展和应用,云技术服务为数字图书馆提供了新的建设场景和运行方式。从前述分析可以看出,云计算环境下数字图书馆技术体系结构的发展与变革已成为学界、业界普遍关注的话题,且众多科研人员突破传统研究模式的局限,更多考虑到数字图书馆的发展现状及其与云计算技术的契合点,在此基础上开始探讨数字图书馆技术体系结构的未来发展方向、设计原则、基本模块等观点,紧扣云计算环境带动各领域协同发展的时代背景,较之以往从宏观层面阐述数字图书馆技术体系结构的基本概念、内容构成等方面的研究是新的突破与发展。

6 结语

数字图书馆技术体系结构承载着数字图书馆的功能实现,决定了数字图书馆的发展范式。作为建设数字图书馆的基础,数字图书馆技术体系结构是数字图书馆在网络和计算机技术上的具体实现,是未来信息社会处理、存储和应用数字化信息的基本构架。由于数字图书馆技术体系结构涉及数字图书馆的诸多方面,因此其成为数字图书馆建设与研究的一个难点。一方面,数字图书馆所收藏、处理的数字资源内容多样、规模不一、服务对象各异、知识产权复杂等技术和非技术因素使得数字图书馆技术体系结构的设计各具特色;另一方面,作为互联网信息资源组织者与服务提供者,数字图书馆要为所有的用户提供服务,又必须在满足特定功能需求的同时,遵循数字图书馆技术体系结构设计的基本规则。随着数字图书馆相关研究的不断深入推进,数字图书馆的技术体系结构方面的研究也将进一步取得丰硕的成果^[62-63]。系统梳理这些成果不仅对全面回顾数字图书馆技术体系机构的研究发展历程具有重要作用,对于日后开展相关研究也具有重要的启示意义。需要指出的是,对于图书馆界而言,数字图书馆是计算机技术与图书馆相结合的一个产物,其技术体系结构的形成与发展,始终是理论模型、原型系统和实践

系统共同推动的结果,三者关联交织而又相互促进。后续的相关研究,可以根据这三种类型的特点,进一步对数字图书馆技术体系进行细化和梳理,以期得到更为专深的研究发现。

参考文献

- [1] 汪冰. 电子图书馆理论与实践研究 [M]. 北京: 北京图书馆出版社, 2000.
- [2] CNRI. Computer Science Technical Reports Project (CSTR) [EB/OL]. [2021-02-04]. <http://www.cnri.reston.va.us/home/ctr.html>.
- [3] KAHN R, WILENSKY R. A Framework for Distributed Digital Object Services [EB/OL]. [2021-02-04]. <http://www.CNRI.Reston.VA.US/home/ctr/arch/k-w.html>.
- [4] ARMS W Y. Key Concepts in the Architecture of the Digital Library [EB/OL]. [2021-02-04]. <http://www.dlib.org/dlib/July95/07arms.html>.
- [5] LAGOZE C, ELY D. Implementation Issues in an Open Architecture Framework for Digital Object Services [R]. Cornell Computer Science Technical Report TR95-1540, 1995.
- [6] LAGOZE C, MCGRATH R. A Design for Inter-Operable Secure Object Stores (ISOS) [R]. Cornell Computer Science Technical Report TR95-1558, 1995.
- [7] CNRI. Handle Home Page [EB/OL]. [2021-02-06]. <http://www.handle.net/>.
- [8] ARMS C R. Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress [J/OL]. D-Lib Magazine, 1996 (4) [2021-02-06]. <http://www.dlib.org/dlib/april96/loc/04c-arms.html>.
- [9] ARMS C R. Historical Collections for the National Digital Library: Lessons and Challenges at the Library of Congress [EB/OL]. D-Lib Magazine, 1996 (5) [2021-02-06]. <http://www.dlib.org/dlib/may96/loc/05c-arms.html>.
- [10] Library of Congress. American Memory Background Papers and Technical Information [EB/OL]. [2021-02-06]. <http://lcweb2.loc.gov/ammem/ftpfiles.html>.
- [11] ARMS W Y, BLANCHI C, OVERLY E A. An Architecture for Information in Digital Libraries [J/OL]. D-Lib Magazine, 1997 (2) [2021-02-06]. <http://www.dlib.org/dlib/february97/cnri/02arms1.html>.
- [12] PAYETTE S. Digital Library Architecture: A Service-Based Approach [EB/OL]. [2021-02-12]. <http://www.cs.cornell.edu/payette/presentations/DL-architecture.ppt>.
- [13] WTEC Hyper-Librarian. PRINCIPLES OF DIGITAL LIBRARIES [EB/OL]. [2021-02-12]. http://www.wtec.org/loyola/digilibs/04_02.htm.
- [14] 郑巧英, 杨宗英. 数字图书馆原型体系结构初探 [J]. 情报学报, 1999 (6): 523-529.
- [15] California Digital Library Workgroup. California Digital Library Technical Architecture and Standards [R/OL]. [2021-02-12]. <http://besser.tsoa.nyu.edu/howard/Classes/208-s00/CDL/CDL-Arch-031000.pdf>.
- [16] 欧洁, 罗治国, 林守勋, 等. 数字图书馆的数字对象体系结构 [J]. 中国科学院研究生院学报, 2000 (1): 93-99.
- [17] 张健, 覃嵘, 顾君忠. 开放体系分布式数字图书馆原型设计 [J]. 计算机应用, 2000 (6): 38-41.
- [18] 王军, 杨冬青, 唐世渭, 等. 数字图书馆的体系结构 [J]. 情报学报, 2000 (6): 564-572.
- [19] SULEMAN H, FOX E A. A framework for building open digital libraries [J/OL]. D-Lib Magazine, 2001 (12) [2021-02-12]. <http://www.dlib.org/dlib/december01/suleman/12suleman.html>.
- [20] 孙一钢. 数字图书馆的技术体系结构 [J]. 现代图书情报技术, 2001 (5): 8-10.
- [21] 郑彦宁. 具有集中索引的数字图书馆系统结构 [J]. 情报学报, 2001 (6): 642-647.
- [22] JANÉE G, FREW J. The ADEPT Digital Library Architecture [R/OL]. [2021-02-06]. <https://dl.acm.org/citation.cfm?id=544306>.
- [23] 真溱. 数字图书馆: 体系结构、问题及分析 [J]. 图书馆杂志, 2002 (3): 3-7.
- [24] 马自卫. 数字图书馆及其体系结构实现的研究 [J]. 现代图书情报技术, 2002 (2): 3-5.
- [25] 王军. 数字图书馆体系结构比较研究 [J]. 大学图书馆学报, 2002 (6): 5-12, 90.
- [26] 盛小平. 数字图书馆体系结构的比较研究 [J]. 图书馆杂志, 2002 (12): 36-40.
- [27] 陈梅. 数字图书馆体系结构参考模型 [D]. 上海: 华东师范大学, 2002.
- [28] PANDEY R. Digital library architecture [J]. World Digital Libraries-An International Journal, 2003 (2): 165.
- [29] 刘炜, 张亮. 数字图书馆的体系结构与元数据方案 [J]. 情报学报, 2003 (2): 148-154.

- [30] 孙魁明, 李春霞. 数字图书馆体系结构的研究与实现 [J]. 北京师范大学学报(自然科学版), 2003 (2): 280-284.
- [31] 牛振东, 赵四友. 数字图书馆体系结构的发展 [J]. 现代图书情报技术, 2003 (3): 20-23.
- [32] 叶鹰, 金玮. 数字图书馆的体系结构与理论模型 [J]. 图书情报工作, 2003 (9): 45-47.
- [33] 韩毅. 基于知识网的区域数字图书馆建设框架 [J]. 大学图书馆学报, 2003 (6): 11-14.
- [34] KNEŽEVIĆ P, MEHTA B, NIEDERÉE C, et al. Supporting Information Access in Next Generation Digital Library Architectures [EB/OL]. [2021-02-08]. http://link.springer.com/chapter/10.1007%2F11549819_13#page-1.
- [35] PETINOT Y, GILES C L, BHATNAGAR V, et al. A Service-Oriented Architecture for Digital Libraries [EB/OL]. [2021-02-08]. <http://clgiles.ist.psu.edu/papers/ICSOS-2004-service-oriented-arch.pdf>.
- [36] KNEŽEVIĆ P, MEHTA B, NIEDEREE C, et al. Supporting Information Access in Next Generation Digital Library Architectures [EB/OL]. [2021-02-02]. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.747&rep=reprl&type=pdf>.
- [37] 李爽. 数字图书馆体系结构的设计原则与系统构成 [J]. 图书情报知识, 2004 (1): 52-54.
- [38] 王亮. 数字化图书馆基本体系及实现 [D]. 武汉: 华中科技大学, 2004.
- [39] 龚萍. 第三代数字图书馆体系结构分析及其建设中的知识产权问题研究 [D]. 武汉: 武汉大学, 2004.
- [40] 王东霞. 数字图书馆: 体系结构与建设 [J]. 大学图书情报学刊, 2004 (3): 39-41.
- [41] 赵芳, 任瑞娟. 数字图书馆建设中体系结构问题 [J]. 情报杂志, 2004 (10): 62-63.
- [42] IOANNIDIS Y, SCHEK H J, WEIKUM G. Future Digital Library Management Systems: System Architecture and Information Access [EB/OL]. [2021-02-08]. http://www.delos.info/files/pdf/Proceedings/Dagstuhl_2903_010405/delos-dagstuhl-handout-all.pdf.
- [43] 董慧. 数字图书馆网格应用模型研究 [M]. 武汉: 湖北人民出版社, 2005.
- [44] 梁小吟. 基于组件的数字图书馆系统体系结构的研究 [D]. 北京: 北京交通大学, 2005.
- [45] 张兴军. 网格支持的数字图书馆研究 [D]. 北京: 北京航空航天大学, 2005.
- [46] TRNKOCZY J, TURK Z, STANKOVSKI V. A grid-based architecture for personalized federation of digital libraries [J]. Library Collections, Acquisitions, & Technical Services, 2006, 30: 139-153.
- [47] 毕强, 韩毅, 朱亚玲. 国外基于网格技术的数字图书馆结构体系比较研究 [J]. 情报学报, 2006, 25 (6): 330-336.
- [48] A reference architecture for digital library system: principles and applications [EB/OL]. [2021-02-08]. http://link.springer.com/chapter/10.1007%2F978-3-540-77088-6_3#page-2.
- [49] ARTINI M, CANDELA L, CASTELLI D, et al. Sustainable Digital Library Systems over the DRIVER Repository Infrastructure [EB/OL]. [2021-02-08]. http://link.springer.com/chapter/10.1007/978-3-540-87599-4_23#page-1.
- [50] 米飙. 2010云计算元年 [EB/OL]. [2021-02-08]. <http://www.forbeschina.com/review/201008/0003352.shtml>.
- [51] 中关村在线. 2010云计算元年各大厂商云布局谁更胜 [EB/OL]. [2021-02-08]. <http://www.csdn.net/article/2010-12-28/286403>.
- [52] 上海市图书馆学会. “2009图书馆前沿技术论坛: 云计算与图书馆”会议通知 [EB/OL]. [2021-02-08]. <http://www.libnet.sh.cn/tsgxh/list/list.aspx?id=6011>.
- [53] GOLDNER M R. Winds of change: Libraries and cloud computing [J]. BIBLIOTHEK Forschung und Praxis, 2010, 34 (3): 270-275.
- [54] 叶小榕, 邵晴. 一项结合云计算和PKI的数字图书馆系统框架设计 [J]. 科技导报, 2010, 28 (9): 63-67.
- [55] 马晓亭, 陈臣. 数字图书馆云存储应用系统研究与实现 [J]. 图书馆理论与实践, 2012 (5): 8-13.
- [56] 赵美凯. 基于Eucalyptus的数字图书馆私有云服务系统的研究与实现 [D]. 北京: 北京邮电大学, 2012.
- [57] 胡昌平, 万华. 云环境下移动数字图书馆跨系统服务平台构建与实现 [J]. 国家图书馆学报, 2013, 22 (2): 40-48, 71.
- [58] 张艳, 潘吴斌. 基于Hadoop的数字图书馆云检索系统的设计与实现 [J]. 图书馆理论与实践, 2014 (4): 92-95.
- [59] 王宁. 浅析大数据背景下的数字图书馆知识发现系统 [J]. 图书馆工作与研究, 2016 (4): 58-61.
- [60] 刘传玺, 王珏. 基于大数据的智慧数字图书馆设计 [J]. 信息与电脑(理论版), 2020, 32 (14): 133-135.
- [61] 石婷婷, 徐建华, 张雨浓. 数字孪生技术驱动下的智慧图书馆应用场景与体系架构设计 [J]. 情报理论与实践, 2021, 44 (3): 149-156.
- [62] LIU Y Q. Surfing the World of Knowledge-Analytical Case Studies of Digital Liberties in the United States [M]. Beijing:

Ocean Press, 2014: 33-44.

California: CreateSpace, 2016: 12-14.

[63] LIU Y Q. Digital Libraries in Theory and Practices (2nd Ed) [M].

作者简介

黄国彬, 男, 1979年生, 博士, 副教授, 研究方向: 信息法学、信息分析, E-mail: huanggb@bnu.edu.cn.

郑霞, 女, 1994年生, 硕士研究生, 研究方向: 信息资源管理。

孙一钢, 男, 1961年生, 研究馆员, 副馆长, 研究方向: 数字图书馆。

刘燕权, 男, 1952年生, 终身教授, 研究方向: 数字图书馆、智慧图书馆。

Analysis on Research Process and Characteristics of Technology Architecture of Digital Library

HUANG GuoBin¹ ZHENG Xia¹ SUN YiGang² LIU YanQuan³

(1. School of Government, Beijing Normal University, Beijing 100875, China; 2. National Library of China, Beijing 100081, China;
3. South Connecticut State University, New Haven 06515, USA)

Abstract: The technical architecture of digital library carries the function realization of digital library and determines the development paradigm of digital library. Combing the research stage of technology architecture of digital library is helpful to recognize its development context and provide theoretical support for promoting the renewal and improvement of technology architecture of digital library. This article mainly applies the methods of literature research and content analysis to sort out the domestic and foreign scholars' research on technical architecture of digital library, divides the development stages of the technical architecture of digital library according to the existing research contents, and deep analyzes the research characteristics of the technical architecture of digital library in each research stage. The research stage of technology architecture of digital library from 1992 to now can be divided into four stages. The first stage: the research on the technical architecture of digital library based on experimental projects (1992-1998). The second stage: the research on the technical architecture of digital library based on application-based projects (1999-2003). The third stage: the research on the technology architecture of digital library based on integrated services (2004-2008). And the fourth stage: the research on the technology architecture of digital library based on cloud computing and intelligent services (2009-present).

Keywords: Digital Library; Technical Architecture; Research Process

(收稿日期: 2021-02-25)