

# 数字图书馆机制的范式演变及其挑战

张晓林

文摘: 本文描述了数字图书馆从基于数字资源到基于集成信息服务再到基于用户信息活动的范式演变, 分析了集成数字信息服务体系和基于用户信息活动的数字图书馆结构, 并探讨了这种范式演变对图书情报理论与实践的挑战。

关键词: 数字图书馆, 范式演变, 图书情报服务

## The Paradigm Shift of Digital Library Mechanisms and Its Challenges

Abstract: The paper describes the paradigm shift in digital library mechanisms from Resource-based DL to Service-based DL and on to Work-based DL, analyses the basic structures of the Service-based and Work-based DLs, and explores the challenges this paradigm shift presents to library and information theories and practices in general.

Keywords: Digital libraries, Paradigm shift, Library and Information Services

### 1. 前言

迅速发展信息网络和数字信息资源体系正在造就一个全新的信息服务环境, 其中信息资源、信息组织工具、信息传递工具日益聚合在同一数字空间, 信息资源系统、信息服务系统和用户信息系统(例如电子邮件信箱、个人网页、课题网站、机构信息系统、业务信息管理系统等)日益连接在同一网络空间, 各种基于网络、基于知识、基于协作的信息处理机制也日益成熟, 它们之间的链接、交换、互操作、协作和集成也日益成为可能。

数字图书馆(以及所有信息服务系统)的根本目标是通过一系列服务机制有效支持用户利用信息来学习和创造知识。当信息资源、信息服务和用户(信息活动)都聚合在同一数字空间时, 就有可能从新的技术基础出发, 从用户信息利用全过程及其复杂信息活动的角度来重新审视信息服务系统的功能与结构, 构建全面和直接支持用户信息活动的信息服务机制。因此, 随着数字信息资源、信息服务系统和用户信息环境的不断发展, 数字图书馆机制也从基于数字信息资源的系统形态逐步过渡到基于集成信息服务的系统形态, 并开始向基于用户信息活动环境的系统形态过渡。本文将对这一范式演变过程以及它为数字图书馆建设和图书情报理论与实践带来的挑战进行初步分析。

### 2. 数字图书馆的范式演变

我们可以根据数字图书馆建设的基点、体系形式和所解决的关键任务等, 将数字图书馆分为不断递进和深化的三代范式<sup>[1-3]</sup>:

2.1 第一代——基于数字化资源的数字图书馆 (Resource-based digital library)

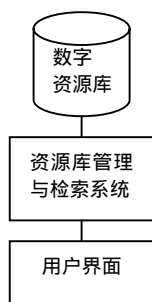


图 1

第一代数字图书馆主要在特定文献资源数字化的基础上建立数字信息资源系统,它们往往作为独立系统,往往嵌入到传统图书馆系统或上层机构信息系统中,将跨时空检索和传递特定数字化资源作为其主要任务,称为基于数字化资源的数字图书馆。这类数字图书馆的范例包括美国 LC 的 American Memory 系统 (<http://memory.loc.gov/>),密西根大学的 JSTOR (<http://www.jstor.org/>),拉斯阿拉莫斯国家实验室的 Physics E-Print (<http://www.arXiv.org/>)、美国 NASA 的 Astrophysics Data System (<http://ads.harvard.edu/>) 以及我国实验数字图书馆项目等。目前,许多图书馆、档案馆乃至博物馆等的数字化馆藏系统基本属于这一范围。我们可用图 1 来表示这一类数字图书馆的基本逻辑结构,这时资源库管理与检索系统往往是与数字资源库直接捆绑(Hard-binding),其功能形式、技术方法和操作管理机制往往取决于资源库的内容格式、元数据格式、知识组织体系和特定软硬件平台,可能利用专用甚至私密的代理模块来实现,最后通过 Web 平台供用户直接检索浏览。

这类数字图书馆的任务范畴一般包括:数字化对象的选择(例如珍贵文献、手稿、档案、地方特色文献、或经过授权的出版物等),文献数字化方法(Digitization),数字文献格式标准体系(例如数字文本标记格式、数字图像扫描格式等的标准化),描述和管理具体数字文献的元数据,数字资源库组织(包括标识符机制、内部标识与检索机制、存储管理系统等),检索与呈现方法(包括并行检索、基于内容检索和简单数字对象的呈现等),初步的知识产权管理(包括用户使用控制和数字水印保护等),数字化工作流程等。

## 2.2 第二代——基于集成信息服务的数字图书馆(Service-based digital library)

为有效利用数字信息环境中分布、多样化、往往是异构的数字信息资源,第二代数字图书馆致力于支持分布的数字信息系统间的互操作(interoperability),支持这些系统间无缝交换和共享信息资源与服务,并由此构造一个逻辑的集成信息服务机制,形成基于集成信息服务的数字图书馆。这一代数字图书馆不再以文献数字化和具体数字资源库建设为核心,而主要是面向分布和多样化数字信息资源(包括由出版商、学术机构、各类机构等拥有的“正式”和“非正式”资源),通过服务集成(包括虚拟资源体系建设、跨系统多系统检索、分布式使用管理、分布式权益管理、分布式数字参考咨询服务、长期保护协调等)构造统一的信息服务系统,将形成与传统图书馆不同的新系统形态和组织形态,是目前数字图书馆技术研究、应用试验和开发的主要趋势,其范例包括加州大学的 California Digital Library (<http://www.cdlib.edu/>),计算机科学领域的 Networked Computer Science Technical Reference Library (<http://www.ncstrl.org/>)、学位论文领域的 Networked Digital Library of These and Dissertations(<http://www.ndltd.org/>)、美国 OhioLink(<http://www.ohiolink.edu/>)、英国 National Electronic Site License Initiative(<http://www.nesli.ac.uk/>)、英国 Distributed National Electronic Resources (DNER)<sup>[4]</sup>以及我国的 CALIS 等。

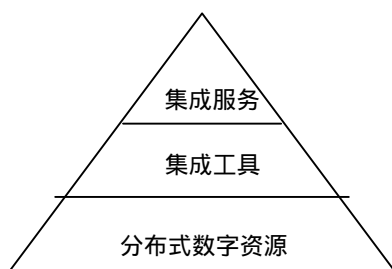


图 2

图 2 是 MOA2 项目提出的第二代数字图书馆体系架构<sup>[5]</sup>,将第一代数字图书馆作为具体数字资源系统,强调通过一系列搜寻、转换、整合工具来集成这些分布的系统,支持集成服务。

这一代数字图书馆的任务范畴将更集中于:分布式系统结构,系统互操作(例如基于分布式对象、基于代理与协调技术、基于搜寻或整合协议的互操作),数字对象与数字对象唯一标识符,元数据互操作,数字资源开放链接,分布式开放式使用管理,分布式开放式数字权益管理,网络化资源建设和资源组织,以及基于内容和集成的检索技术(基于内容检索、

基于知识体系和语义的检索、跨系统检索、跨语言检索等)。

### 2.3 基于用户信息活动的数字图书馆 (Work-based digital library)

随着聚合数字信息空间的逐步形成, 人们第一次有可能摆脱传统图书情报系统(甚至传统数字图书馆系统)单纯基于信息资源的服务形态和将信息系统与用户信息利用过程相对隔绝的局限, 以支持用户灵活地处理信息、提炼知识和交流协作为核心, 围绕用户信息活动和用户信息系统来组织、集成、嵌入数字信息资源和信息服务, 从而更直接、深入、有效地支持用户检索、处理、利用信息以解决问题的全过程。在这种理念推动下, 数字图书馆的前沿研究者们已开始探讨以用户信息活动为基础的数字图书馆机制。例如, 美国国家科学基金会的“国家科学、数学、工程和技术教育数字图书馆项目”(NSDL)<sup>[6-7]</sup>明确提出建设成围绕用户协作化学习过程的分布式资源网络和学习机制, 个人或集体可充分和动态调用各种数字化资源和工具(包括合作学习系统、远程实验室、虚拟实验室等)来个性化和协作地检索、集成、处理信息并以此支持合作学习。将于今年九月举行的欧洲数字图书馆会议(ECDL)已将研究开发嵌入到用户信息空间和用户合作过程的数字图书馆系统作为其三大主题之一<sup>[8]</sup>, 马里兰大学 MiND 项目提出新一代数字图书馆应直接支持用户在其信息利用过程中灵活处理信息对象<sup>[9]</sup>, 奥地利 Maurer 提出数字图书馆应成为用户交流媒介来支持用户对数字信息的注解、交流和协作处理<sup>[10]</sup>, 美国数字图书馆研究著名专家 Ed Fox 也提到数字图书馆的下一步发展将走向虚拟个性化数字图书馆和嵌入到用户工作环境中的数字图书馆<sup>[11]</sup>。实际上, 一些试验或者应用系统已出现, 例如 NCSU 图书馆的 MyLibrary 系统(<http://my.lib.ncsu.edu/>)支持建立个性化门户, Questia 数字图书馆(<http://www.questia.com/>) 在提供全文图书检索浏览的同时支持用户对图书内容的析取、注解和交流, SOSIG 主题信息网关(<http://www.sosig.ac.uk/>) 在提供网络精选资源导航的同时支持用户发布信息、围绕特定主题资源构造用户社区和进行协作。

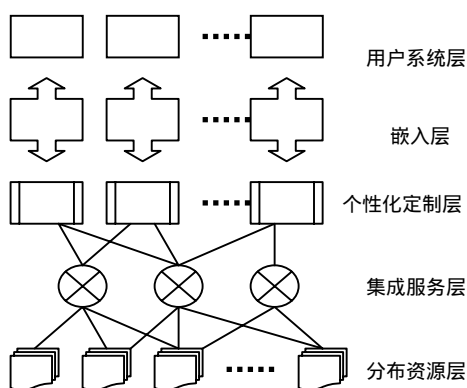


图 3

图 3 给出了第三代数字图书馆的可能模式。在分布式数字资源系统(分布资源层)和集成信息服务体系(集成服务层)基础上, 通过一定的个性化定制机制形成适应用户或用户群组需要的可能是动态过滤、析取和组合的资源、工具和服务集合(个性化定制层), 这些集合被有机地嵌入到用户信息系统或用户信息利用环境(用户系统层)中, 直接支持用户的信息利用活动。

这一代数字图书馆的任务范畴可能涉及: 数字对象与分布式对象代理技术, 智能代理技术,

个性化机制, 动态文献和动态文献集技术, 知识组织系统技术, 信息协作处理机制, 基于 XML 技术体系的信息处理技术和信息系统定义、构建、集成技术, 用户信息系统和信息处理流程中数字化信息资源与服务的嵌入与定制机制等。

### 3. 基于集成信息服务的数字图书馆模式分析

面对异构、多样的分布式资源系统和分布、移动的用户, 集成信息服务体系将:

- a. 支持分布和多样化资源系统的方便接入, 同时支持各资源系统的自主性和本地服务;
- b. 支持对这些资源系统的逻辑集成, 支持以标准形式跨资源系统进行搜寻、检索、转换和整合;
- c. 支持基于整个分布资源体系的集成服务和 Service 管理机制, 支持对分布式第三方工具或服务系统进行动态和无缝调用, 形成逻辑整体服务系统;

- d. 支持对资源体系进行逻辑重组,以构建适应不同用户群的虚拟信息资源和服务系统;
- e. 支持整个机制的开放性、可伸缩性和可扩展性,能方便接入和动态组合任意数量或类型的资源与服务系统(包括新的资源形式、系统形式和服务工具)。

集成信息服务体系的核心问题是分布、异构系统的互操作,可能的实现形式包括:

- a. 基于中间协调与转换代理的联邦式系统(Federated Systems),例如 NCSTRL<sup>[12]</sup>、基于 Z39.50/ILL 协议的集成检索系统、英国 MIA 结构<sup>[13]</sup>及 DNER<sup>[4]</sup>系统,可通过全面和复杂的代理机制支持强有力的集成、转换服务;
- b. 基于标准搜寻协议的开放资源体系(Open Harvesting Systems),如 OAI 协议机制<sup>[14]</sup>,资源系统(作为数据提供者)可通过简单开放机制,提供基础性元数据和读取功能,从而支持服务集成者搜寻和提取元数据、建立元数据库、读取数字对象及提供其它第三方服务;
- c. 基于整合检索协议的跨系统搜集整合机制(Gathering/Integration Systems),在诸如 LDAP、WHOIS++、SDLIP 等协议支持下,这些机制可直接利用各数字资源系统的可公开获取信息进行整合检索,例如 CrossRoads<sup>[15]</sup>、Isaac Network<sup>[16]</sup>、Imesh<sup>[17]</sup>、LFDL<sup>[18]</sup>等。

图 4 给出了一个集成信息服务体系结构,通过各种协议和机制来支持对分布式数字资源的检索和获取,通过一系列转换、整合、服务调用、工具调用和管理机制来提供逻辑集成和服务管理,并按照不同需要定制和集成不同的资源与服务(包括不同的集成服务体系)来形成不同的虚拟集成信息服务体系。

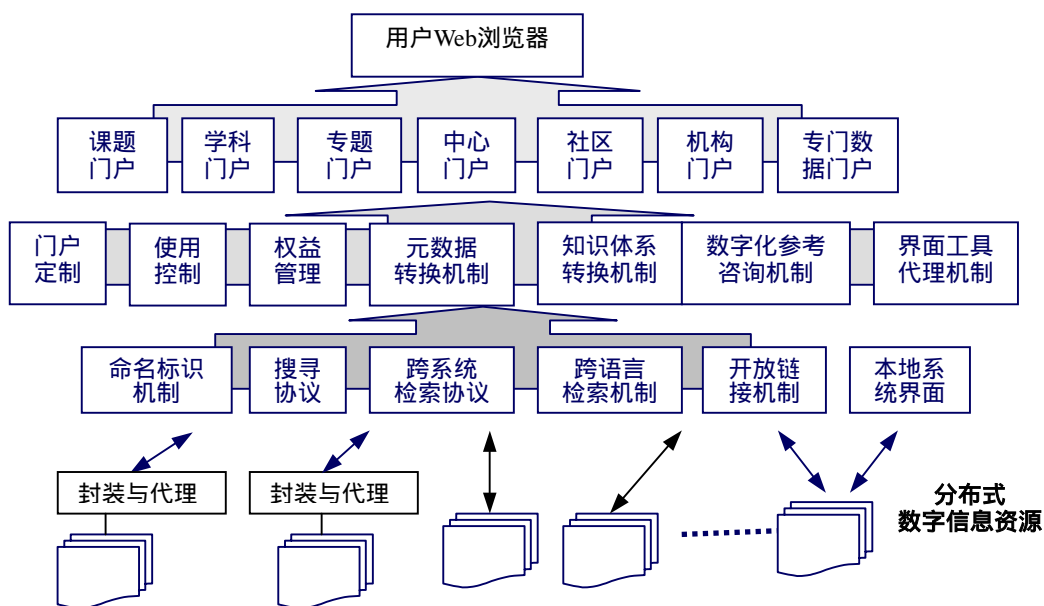


图 4

#### 4. 基于用户信息活动的数字图书馆模式分析

作为围绕用户信息活动和用户信息系统来组织、集成、嵌入数字信息资源与信息服务的系统机制,第三代数字图书馆的模式本身将是个性化和动态的<sup>[19-22]</sup>。根据目前的思考和研究,可能包括以下形态:

- a. 个性化定制和聚合的数字图书馆(Personally focused DL),即按照用户个人需求和知识体系定制的聚合了分布式多元化资源、工具和服务的数字信息体系。
- b. 按照用户群组或项目定制和集成的数字图书馆(Project/group focused DL),即按照用户群组或项目需求、知识体系、组织方式和活动形式定制的集成了分布式多元化资源、工具和服务的数字信息体系。
- c. 动态文献与动态数字图书馆机制(Active Document/Active DL),即可按照不同的用户需求、不同的知识组织方式、不同的视图机制或结构、不同的内容析取方式和链接结构、

以及这些需求或机制的动态变化来灵活组织和集成的文献或数字图书馆系统, 以支持根据用户需求灵活地析取、注解、链接、集成、组织和呈现信息内容。

d. 基于数字图书馆的学术交流机制 (DL-based scholarship), 即有机融合到数字资源系统或集成数字信息服务体系中的用户信息交流和协作机制, 支持用户对数字图书馆中信息内容的注解、析取、链接、补遗、再集成、传递、讨论、协作处理和合作写作等活动, 形成数字图书馆基础上的用户个人或群体工作空间。

e. 嵌入到用户信息使用环境中的数字图书馆 (User system imbedded DL), 即直接和有机嵌入到用户网站、网络学习系统、计算机辅助协作系统、 workflow 管理系统、决策支持系统、知识管理体系等中的数字信息服务机制, 支持动态捕获需求、动态定制、面向用户工作和知识结构的链接与嵌入机制、基于智能和基于内容的主动匹配与推送等功能。

作为一个预期使用场景和范例, 图 5 给出了一个嵌入到课程学习系统中的数字图书馆机制, 它在支持开放嵌入的课程学习系统(左圆)和分布式数字资源系统及其集成服务体系(右圆)的基础上, 提供以下可能功能:

a. 根据具体课程内容及教学要求, 在课程概念集支持下定义课程知识组织体系 (Course Knowledge Organization System, CKOS) 和课程信息需求;

b. 根据 CKOS、课程信息需求和开放门户描述机制, 定制和组织课程信息门户;

c. 通过搜寻推送机制搜寻相关资源, 利用开放链接机制在课程信息门户、课程内容、讨论、作业等中无缝嵌入相关资源及辅助处理工具, 并自动组织课程数字资源库; d. 基于课程内容、讨论和作业等动态解析课程知识集合和学习热点难点, 动态修改 CKOS 和信息需求, 并动态重组课程信息门户;

e. 对用户提供的课程资源 (例如参考文献) 可通过关联信息搜寻, 自动搜寻相关信息 (例如引用被引用、同作者、同课题、权威文献、枢纽文献、语义关联、使用关联等) 及辅助处理工具, 并将结果传递给提供者 and 推送到课程学习系统、课程信息门户和资源库。

学习者进入课程学习系统时, 这个机制将自动、动态地汇集和嵌入相关信息资源、工具和服务。类似的机制还可嵌入到研究协作系统、 workflow 管理系统、决策支持系统等中。

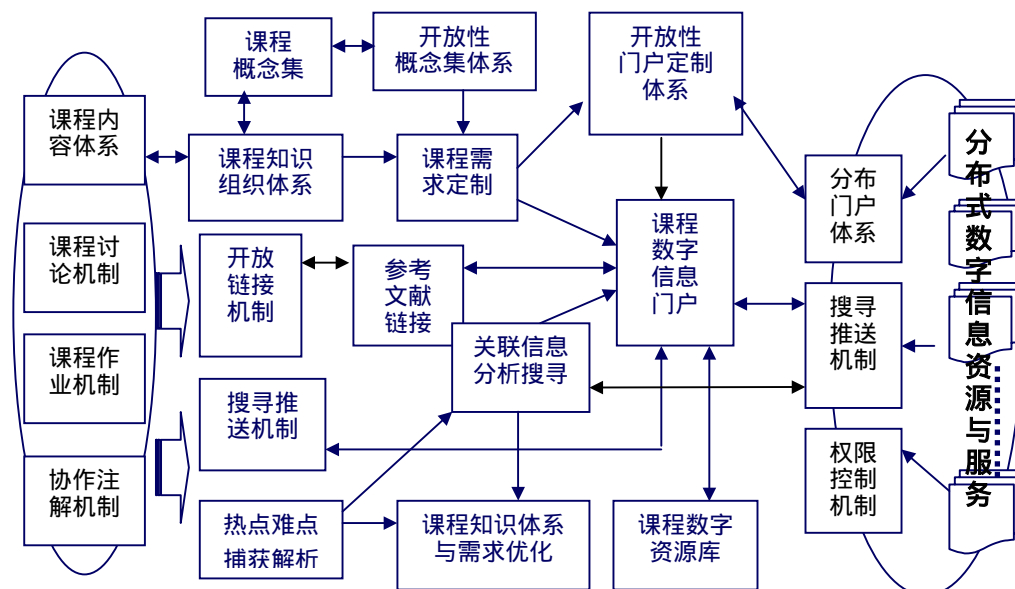


图 5

### 5. 数字图书馆范式演变挑战

上述数字图书馆范式演变的挑战至少体现在两个层面: 一是对数字图书馆建设的影响, 一是信息服务根本理念和模式的挑战。

### (1) 范式演变对数字图书馆建设的影响

数字图书馆建设是一个逐步深化的过程,前一代系统为后一代机制提供发展基础,而后一代机制又对前一代系统提出宏观参照框架和长期发展制约。因此,尽管我们现在可能仍然着重于数字信息资源系统或集成机制的建设,但必须充分考虑后续系统机制的影响。

首先,第一代数字图书馆作为基础数字资源系统,将被集成到第二代集成服务体系中,这种集成体系反过来将对这些资源系统建设提出技术和管理等方面的要求。例如,按照集成服务体系,基础数字资源系统应该在内容标记与标识、元数据、知识组织体系、检索与传输协议、界面机制、用户使用控制、知识产权管理、长期保存保护等方面采用标准化和开放性技术,以支持集成和与其它系统的互操作;应该置于一个集成体系框架中、基于合作和集成来规划和实施数字资源系统建设,包括与整个信息生产链上的其它所有参与者—例如出版商和用户资源系统的合作和集成。

进而,从第三代数字图书馆角度,第一代、第二代数字图书馆系统都应是能被灵活和动态地嵌入到用户信息处理环境中的信息资源与服务机制,这就要求这些资源和服务机制在更广泛基础上实现标准化、开放性和互操作(例如主要不再基于 MARC 和 Z59.50 等标准,而更多地基于 WWW/XML 标准和诸如教育、电子商务、政府信息管理等应用领域标准),支持基于内容和知识组织体系的开放性的信息析取、转换、注解、链接、重组和传递,支持信息内容、资源系统和信息服务过程的开放性定义、描述、解析、链接、构建、嵌入和集成,支持对信息内容、功能模块和使用机制的可伸缩和分布式管理控制,从而保障数字图书馆系统及其任意内容或功能部分能被灵活、有机、可靠和可控制地嵌入用户信息系统和用户信息活动过程中。

### (2) 范式演变对信息服务理念和模式的挑战

新型数字图书馆形态本身体现了关于信息服务和信息系统的新理念,并要求在信息结构、组织体系、服务功能和运营机制等方面有新的模式来支持和保障其有效发展,这些都对我们长期熟悉的许多根本理念和模式提出了挑战。例如:

a. 系统模式。数字图书馆范式演变体现了一个系统研究开发的基点从信息资源逐步过渡到信息服务机制再到用户信息活动的过程。在这种背景下,信息系统(或者“图书馆”、“数字图书馆”)将不再仅仅是、甚至主要不是一个资源系统或者一个基于资源的服务系统,而是一种以用户为中心来聚合资源、服务、信息利用活动的动态机制,其目标和功能都着力于支持用户利用信息、提炼知识、解决问题,成为用户工作环境和流程的有机部分。实际上,这种机制恰好体现了信息服务系统的根本目标和根本形态,它的挑战以及它的生命力也正是在于以用户为中心来认识、设计、组织和不断发展信息服务系统。

b. 信息模式。我们习惯于认为信息内容和信息系统都是静态和独立的,仅仅是信息检索和处理过程中的被动、僵固的对象。但第二代和第三代数字图书馆机制都反映出信息模式各层次(包括文献、元数据、知识组织体系、信息处理流程、信息系统架构)开放性可描述、可解析、可抽取、可转换、可链接、可嵌入、可互操作、可重组、可扩展的趋势,文献成为一种有机系统,知识体系和信息系统也成为动态变化、联系和组合的有机体,可随着用户的信息需要、利用过程和新信息产生而以不同形式来呈现、组合和变化,从而支持丰富的信息处理和服务功能。这种趋势不但要求我们以开放标准来定义各层信息模式,而且需要重新认识元数据和信息组织概念,引入诸如数字对象、开放链接、分布式代理、开放系统框架等概念和技术,建立开放灵活的信息组织技术机制,真正根据用户在其信息活动中的动态需求来快速和动态地构建、集成、嵌入信息资源与服务机制<sup>[23-24]</sup>。

c. 信息服务模式。传统信息服务模式主要是基于资源的服务,以提供固定组合的有形信息单元为核心。这当然是、而且仍将是信息服务的一个重要方面,但第二代、第三代数字图书馆机制却更多地直接面向用户对信息进行感知、捕获、分析、重组、传递、应用的全过



程及其多样化需求,更多地关心“是否通过我的服务解决了您的问题”(而不仅仅是“我是否提供了您索要的文献”),嵌入到和贯穿于用户信息利用过程,通过资源与工具的链接、集成和嵌入来动态构建恰好符合用户需要的服务机制,通过信息的析取、链接、嵌入和重组来形成恰好符合用户需要的知识产品。这种服务不再仅依靠单一物理资源系统或某种固有服务模式,不再游离于用户活动和用户环境之外,而要求围绕用户个性化信息活动和环境,充分集成和动态组合各种资源、工具和服务,动态设计、组织和协调有关服务模式与系统形态。实际上,信息技术发展已经并将继续为实现新的服务形态提供有效的手段和工具,但我们迫切需要在思想观念、人员素质、系统组织、工作方式等进行改革来适应新的服务形态<sup>[25]</sup>。

#### 参考文献:

- [1] 张晓林, 分布式数字图书馆机制. 情报学报, 2001 (待发)
- [2] 张晓林, 虚拟信息服务体系的资源建设. 21 世纪图书馆发展与变革. 北京图书馆出版社, 2000 年 7 月
- [3] 曾蕾等, 数字图书馆: 路在何方? 情报学报, 19(1)64-73, 2000.1
- [4] Adding Value to the UK's Learning, Teaching and Research Resources: the Distributed National Electronic Resource (DNER). ([http://www.jisc.ac.uk/pub99/dner\\_vision.html](http://www.jisc.ac.uk/pub99/dner_vision.html))
- [5] The Making of America II Testbed Project White Paper, V2.0, Sept. 15, 1998 (<http://sunsite.berkeley.edu/moa2/wp-v2.html>)
- [6] Zia, L. L. The NSF National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) Program. D-Lib Magazine, Oct 2000. (<http://www.dlib.org/dlib/october00/zia/10zia.html>)
- [7] Agogino, A., et al. The Development of a National Science, Mathematics, Engineering and Technology Educational Library. In Global Digital Library Development, Proceedings of 9<sup>th</sup> NIT Conference, pp.1-12, Tsinghua University Press, 2001.5
- [8] 5th European Conference on Research and Advanced Technology for Digital Libraries <http://www.ecdl2001.org>
- [9] Larsen, R. Information Dynamics; An Information-Centric Approach to Digital Library Interoperability. In Global Digital Library Development, Proceedings of 9<sup>th</sup> NIT Conference, pp.137-144, Tsinghua University Press, 2001.5
- [10] Maurer, H. Beyond Classical Digital Libraries. In Global Digital Library Development, Proceedings of 9<sup>th</sup> NIT Conference, pp. 165-174, Tsinghua University Press, 2001.5
- [11] Ed Fox. 与作者在第九届 New Information Technologies 国际学术研讨会上的讨论, 2001.5
- [12] Leiner, B. The NCSTRL Approach to Open Architecture for the Confederated Digital Library. D Lib Magazine, Dec. 1998 (<http://www.dlib.org/dlib/december98/leiner/12leiner.html>)
- [13] MIA Logical Architecture. Setp. 1999. (<http://www.ukolm.ac.uk/dlis/models/requirements/arch/>)
- [14] Open Archive Initiatives. (<http://www.openarchives.org/>)
- [15] CrossROADS <http://www.ukoln.ac.uk/metadata/roads/crossroads/>
- [16] Isaac Network <http://scout.cs.wisc.edu/research/isaac/index.html>
- [17] The IMesh Toolkit: An architecture and toolkit for distributed subject gateways <http://www.imesh.org/toolkit/>
- [18] Shi, R. et al. Interoperable Federated Digital Library Using XML and LDAP. In Global Digital Library Development, Proceedings of 9<sup>th</sup> NIT Conference, pp.277-286, Tsinghua University Press, 2001.5
- [19] 张晓林等, 基于网络的个性化图书情报服务机制. 现代图书情报技术, 2001.1
- [20] 张晓林等, 基于 Web 的专业化课题化信息服务平台. 图书情报工作, 2001.2
- [21] Atkins, D. E. et al. Toward Inquiry-based Education Through Interacting Software Agents. IEEE Computer, May 1996 (<http://www.computer.org/computer/dli.r50069/r50069.html>)
- [22] Poley, J. Digital Learning and Libraries. In Global Digital Library Development, Tsinghua University Press, Proceedings of 9<sup>th</sup> NIT Conference, pp.227-234, 2001.5
- [23] 张晓林, 信息系统的数字对象与扩展文献技术. 情报理论与实践, 2001(待发)
- [24] 张晓林, 数字化信息组织的结构与技术(一)(二). 大学图书馆学报, 2001.4/2001.5
- [25] 张晓林, 走向知识服务. 中国图书馆学报, 2000.5

本文最初发表在《中国图书馆学报》2001 年第 6 期第 3-8 页, 续 17 页。