科技部科技基础性工作专项资金重大项目
研究成果

项目名称：我国数字图书馆标准规范建设

子项目名称：我国数字图书馆标准规范发展战略与基本框架

项目编号：2002DEA20018

研究成果类型：研究报告

成果名称：数字图书馆相关领域标准规范现状与发展研究（数字科研）

成果编号：CDLS-S01-002

成果版本：总项目组推荐稿

成果提交日期：2003 年 7 月

撰写人：潘淑春、盛玲玉、牛离平（中国农业科学院科技文献信息中心）

项目版权声明

作者声明

本报告作者谨保证本作品中出现的文字、图片、声音、剪辑和文后参考文献等内容的真实性和可靠性，愿按照《中华人民共和国著作权法》，承担本作品发布过程中的责任和义务。科技部有关管理机构对于本作品内容所引发的版权、署名权的异议、纠纷不承担任何责任。

《我国数字图书馆标准规范建设》课题组网站（http://cdls.nstl.gov.cn）作为本报告的第一发表单位，并可向其他媒体推荐此作品。在不发生重复授权的前提下，报告撰写人保留将经过修改的项目成果向正式学术媒体直接投稿的权利。

数字图书馆相关领域标准规范现状与发展研究（数字科研）

目　录

1. 数字科研领域关键技术概述

1.1 数字科研领域关键技术基本概念

数字科研是近几年随着网络技术、通信技术的发展而兴起的一个全新的领域，在全球信息化浪潮蓬勃发展的形势下，数字科研也以其独具的特色初露端倪。现代化推动数字化，数字化促进现代化，使世界科学研究越发呈现出联合、虚拟、共享的趋势，在这样的大环境下，世界各国竞相开发信息资源，利用信息技术，抢占数字科研领域制高点。

数字科研领域关键技术主要指在进行科学研究中，利用信息技术、通信技术、网络技术收集、整理、加工、存储、传递各种类型的数据和信息，辅助进行科学实验和研究的有关的主要技术。

1.2 数字科研领域信息技术应用的意义

数字科研对科研群体成功的进行开发下一代强有力的科学研究工具，和共享世界主要成果具有重要意义。数字科研环境将会高效地处理和计算每天产生的巨量科研数据，科研人员将要求有效地获得分布在世界各地的领先数据源，以及计算和网络资源，以便及时经济地管理和分析这些数据。E-Science 将为他们建立这一信息基础结构。同时，通过共同的网格基础结构的提供，为他们提供能够进行交叉学科研究的新的可能性，在此方面，英国 CLRC（Central Laboratory of the Research Councils） 努力发挥着全球领先的作用。E-Science 将改变科研人员进行科学研究的方式与方法，在更加全球化的信息大环境中，进行新的科研创新与革命。

参与全球 e-Science 项目的好处是有利于共同利益，可以在为国际群体做出贡献的同时吸取别国的先进经验，实现和促进全球网格发展。

1.3 数字科研（e-Science）领域关键技术发展现状

e-Science指越来越多地通过分布式因特网，利用大型数据资源，全球范围的计算资源和高度形象化等手段，进行全球合作的科学研究。（E-Science means science increasingly done through distributed global collaborations enabled by the Internet，using very large data collections，terrascale computing resources and high performance visualisation）[1]。

典型的数字科研项目是英国 2001-2004 年自然环境研究委员会（Natural Environment Research Council：NERC）开展 e-Science 活动的项目，获得的 9800 万英镑科研经费。另外国家和国际网格项目也是数字科研项目的重要组成部分，

如 DOE 赞助的 Science Grid，NSF 赞助的 PPDataGrid，European Commission 赞助的 GridLab 等 23 个项目，其中美国 14 个，欧盟 9 个。

1.3.1 网格（Grid）技术

（1）网格的定义与概念

有专家认为，网格是把整个因特网整合成一台巨大的超级计算机，实现计算资源、存储资源、数据资源、信息资源、知识资源、专家资源的全面共享。当然，网格并不一定非要这么大，我们也可以构造地区性的网格，如中关村科技园区网格、企事业内部网格、局域网网格、甚至家庭网格和个人网格。事实上，网格的根本特征是资源共享而不是它的规模。由于网格是一种新技术，因此具有新技术的两个特征：其一，不同的群体用不同的名词来称谓它；其二，网格的精确含义和内容还没有固定，而是在不断变化。因此，我们不应该空谈和争论什么是网格，什么不是网格，而应该集中精力解决关键问题。

最“正统”的网格研究来源于美国联邦政府过去 10 年来资助的高性能计算项目。这类研究使用的名词就是“网格”（Grid）或“计算网格”。早期还使用过另一个名词—“元计算”（Metacomputing）。这类研究的目标是将跨地域的多台高性能计算机、大型数据库、贵重科研设备（电子显微镜、雷达阵列、粒子加速器、天文望远镜等）、通信设备、可视化设备和各种传感器整合成一个巨大的超级计算机系统，支持科学计算和科学研究。这方面的代表性研究工作包括美国国家科学基金会资助的 NPACI、“国家技术网格”（NTG）、分布万亿次级计算设施（DTF）、美国宇航总署的 IDG、美国能源部的 ASCI Grid 以及欧盟的 Data Grid 等（有关这些网格研究的信息可从“全球网格论坛”www.gridforum.org 网站查阅）。

也有人把网格看成是未来的互联网技术。国外媒体常用“下一代 Internet”、“Internet2”、“下一代 Web”等词语来称呼与网格相关的技术。要注意的是，“下一代 Internet”（NGI）和“Internet 2”又是美国的两个具体科研项目的名字，它们与网格研究目标相交，但研究内容和重点有很大不同。中国科学院计算所所长李国杰院士认为，网格实际上是继传统因特网、Web 之后的第三个大浪潮，可以称之为第三代因特网。简单地讲，传统因特网实现了计算机硬件的连通，Web 实现了网页的连通，而网格试图实现互联网上所有资源的全面连通，包括计算资源、存储资源、通信资源、软件资源、信息资源、知识资源等。

总之，网格是一个正在出现的信息基础结构，可以基本上改变我们思维和进行计算的方法。网格将连接多国家和区域计算网格，建立全球计算能源。网格可使众多组织拥有和管理的高性能计算机、网络、数据库和科学工具实现综合协调利用。网格应用常需要处理巨量数据和计算，要求组织间安全数据共享，因此当今的Internet和网络基础结构不能胜任这样的任务。从概念上讲，网格可从三个层次来看：最下层是技术和数据网格：计算机硬件和数据网络；中间层是信息网格：

信息数据库，通过硬软件来进行数据处理；最上层是知识网格，经过高级技术处理挖掘数据，产生知识，用于智力决策（网址：http://e-science.ox.ac.uk）。

（2）网格技术发展现状

1）英国

英国 CLRC e-Science Center 的网格开发项目（Grid Development Programme）是为了满足最具挑战性的科学计算问题和解决全球范围实验、计算、数据和视听资源的综合与利用的需求而设立的。项目的主要组成部分是为科学研究和用户群体便于资源共享而开发以网格为基础的信息基础结构。CLRC 在英国国家 e-Science 项目中起着领头的作用，是英国网格支持中心。负责英国所有学科科研群体之间的网格技术应用调度和支持，协调与许多大学和研究中心的计算和数据资源相链接的英国 e-Science 网格的发展[www.e-science.clrc.ac.uk]。

英国 e-Science 网格连接了 CLRC 9 个中心和 2 个试验室，作为试验基地，选择的网格中间件（Grid Middleware）与美国 NASA IPG（Information Power Initiative）相同。英国 e-Science 网格与各类的大学不同的信息技术政策，防火墙等相连，也是对其基本 Globus 基础结构和安全系统的最好检验。

其e-Science核心计划（Core Programme）投入 1100 万英镑，通过各e-Science中心，分配给协作工业网页网格中间件项目，计划的主要任务是由e-Science试验项目中提取对网格基础结构的要求，包括计算，数据存储，网络要求和理想的网格中间件功能要求[2]。

英国 CLRC 正在进行的 6 个试验项目之一“我的网格”（MyGrid）是专门为数字科学家提供支持的项目。该项目可帮助科学家利用纷繁分散的资源，为科学家提供一个数字工作台。这一工作台目的在于支持试验调查、试验结果积累和成果吸收的科研过程；支持科学家对群体信息的利用；以及允许形成动态小组来解决紧急的研究问题的科学合作。数字工作台还可成为资源选择、数据管理和加工规定方面的个性化设施。

其他有关项目还有：NeSC ePortal Demonstration，将建立一个具有视觉吸引力的网格计算数字门户。GridNet 项目，主要是支持网格技术和标准的工作，将其作为英国 e-Science 项目的组成部分。其目标是支持网格技术、网格标准和网格最好利用实践的持续积极的进行，并提供参与研究活动的经费，保证英国 e-Science 网格稳定发展。

2）美国

美国的主要数字科研项目包括 NASA Information Power Grid，Particle Physics Data Grid Project，Condor Project，GriPhyN Project，Legion Project 和 Global Grid Forum 等。

NASA Information Power Grid，在过去的三年，NASA 通过连接其 R&D 试验

室的计算资源，创建了一个新型的基础结构，即信息力量网格 IPG。其主要目的是促进 NASA 对大规模科学与工程问题的解决，为此提供持续的基础结构，实现高性能计算和数据管理服务，满足科研对工作流程管理框架的支持和协调分散的科学与工程问题的解决过程。在 NASA 看来，这样的框架对其组织进行模拟整个系统的问题是必要的。系统不仅对分散的资源进行计算，而且对专门技术和技能也可进行计算处理。为了模拟整个飞机，包括机翼、发动机、起落架等，NASA 还必须建立一种机制，通过这一机制工程人员和科学家可以在不同的地方进行合作。这是其内部网格的宏伟计划。

IPG 中间件是在 Globus Toolkit 的基础上建的，提供了数字认证的安全技术（Grid Security Infrastructure GSI），快速安全文档转换（GridFTP），和计算工作的批量排序资源管理（Globus Resource Allocation Manager GRAM）。

Condor Research Project是在美国维斯康星-麦迪生大学进行的一个专门的计算性工作的工作量管理系统。现今该系统管理着 1000 多台工作站，像其他批处理系统一样，Condor提供了排序机制、规划政策、优选主题、资源监测和管理功能。它可以将组织内的所有计算力进行无缝结合，组成一个资源整体。Condor吸收了许多刚出现的以网格为基础的计算方法和协议，如Condor-G就可和Globus管理的资源互换（interoperable）。在利用率统计中，一个工作日可以向本校研究人员传递 650 多天CPU的工作量[3]。目前该系统也在世界几百个工业、政府、学术组织安装利用，规模由几台工作站和千台以上不等。

GriPhyN Project（Grid Physics Network），是物理学家和信息技术研究人员的一个协作项目，是为了在 21 世纪进行数据精确科学计算而开始建立的 Petabyte-scale 计算环境。GriPhyN 将设立一个叫做 Petascale Virtual Data Grid（PVDGS）的计算环境，满足全球成千上万科学家数据精确计算的要求。此外该项目还着力开发 Toolkit，利用这一 Toolkit 建立 PVDGS，支持 CMS（Compact Muon Solenoid），ATLAS，LIGO（Laser Interferometer Gravitational-wave Observatory）和 SDSS（Sloan Digital Sky Survey）的分析任务。PVDG 系统软件和技术供科学群体利用，以能够使数据分析工作合作进行。在这个过程中，新一代具有专门知识的交叉学科科学家将会受益。该项目将形成一个多种功能、各域独立的虚拟数据系统。PVDG 的开发将提供新的数据分析能力，使之在基础计算机科学和物理学方面产生革命性发现。

Legion Project 也是由 NASA-IPG 支持的项目，是美国佛杰尼亚大学在搞的一个面向目标的，元系统（Meta-system）软件项目。该项目于 1993 年开始，1997 年向公众推出。该项目主要是支持大规模类似应用代码和为用户提供复杂的物理系统管理。

Globus Project是一个研究与发展项目，重点在于使网格概念应用于科学与工程

计算。主要研究内容包括资源管理、数据管理与存取、应用发展环境、信息服务和安全。其软件开发主要是Globus Toolkit，一组服务和图书馆软件，支持网格和网格应用。Toolkit包括网格安全、信息基础结构、资源管理、数据管理、通信、错误检测和可携带软件。特别是建立在开放网格服务结构（Open Grid Service Architecture：OGSA）机制上的Globus Toolkit 3.0（GT3）主要的优势是对数字科研和电子商务都提供了网格协议[4]。

其他还有 Particle Physics Data Grid Project 等，在此不一一赘述。

3）欧洲

欧洲 European DataGrid 是欧洲共同体资助的项目，目的是建立计算和精确数据资源网格，对来自科研的数据进行分析。未来科学将要求协调的资源共享，和各研究机构的许多研究室产生和存储的巨量数据的协作处理与分析。因此 DataGrid 的目的是开发和试验科学协作技术结构，在此基础上，可使各国科学研究活动相互作用，联合进行。

DataGrid是CERN领导的项目，合作范围广泛，有 5 个重要合作伙伴和 15 个合作伙伴，欧洲领先研究机构如European Space Agency（ESA），France‘s Center National de la Recherche Scientifique（CNRS），Italy’s Istituto Nazionale di Fisica Nucleare（INFN），the Dutch National Institute for Nuclear Physics and High Energy Physics（NIKHEF）和 UK’s Particle Physics and Astronomy Research Council（PPARC）都是其重要合作者。15 个合作伙伴来自捷克、芬兰、法国、德国、匈牙利、意大利、荷兰西班牙、瑞典、和英国[5]。

DataGrid 是由物理资源（网络、计算机、数据盘）和支持资源存取与协调利用的中间件软件组成，DataGrid 中间件在不同领域研究中将起到沟通和综合，使之可进行交叉利用的作用，是网格计算的关键环节。项目也会参考现有的、经过长期验证的开放标准，在必要的软件开发时做适当协调。该项目所有的工具和中间件都将作为开放件，开放的方式是重要的，目的在于能够在网格技术发展初期，在国际科技群体中，自由讨论，发表建议；同时，保证项目研究结果能自由获得，以利未来任何组织的商业开发。

该项目面临的竞争是目前网上分散的海量数据共享的问题，DataGrid 项目将依靠刚出现的网格技术，建立一个宏大的计算环境，使分散的大量文档，数据库，计算机，科学仪器与设备都能具此优势集成，共享这个大环境。当前进行的网格中间件研究只是初步的，还有许多不足。但在群体工作经验和对 Globus，Condor 和 SRB 网格中间件的基础上，一些新系统的创建会更加成功。

IVDGL（International Virtual Data Grid Laboratory）是一个全球数据网格项目，主要是为物理天文高尖端试验服务的。其在美国、欧洲和南美的计算、存储、和网络资源提供了一个唯一的试验室，可在国际和全球范围内进行试验和验证工作。

该项目有成员 20 个，其中美国 5 个大学和研究机构参加。该项目和其他一些有关项目如 GriPhyN 和 PPDG 同欧洲 DataTAG（Data TransAtlantic Grid）和 DataGrid 项目联合建立了跨越大西洋的网格试验基地，叫做 WorldGrid。WorldGrid 提供了 VDT 和 EDG 中间件交互作用基础结构，支持 GriPhyN 的四个试验（ATLAS，CMS，LIGO，SDSS），促进了国际合作的可能性。

在该项目中，欧洲原子能研究组织（CERN：European Organization for Nuclear Research）是主要成员之一，它建于 1954 年，是世界最大的粒子物理中心，拥有 20 个成员国。在欧洲和世界网格领域居领先地位。

4）亚洲

有资料表明，在亚洲，日本也参与了全球数字科研网格研究项目。

1.3.2 交互试验室（co-laboratory，Interactive laboratory：I-Lab）技术发展现状

在过去的几年，交互试验室逐步成为网络预报服务创新性的一个亮点，其服务内容包括系统命名生成（Systematic naming generation），生物化学预报（physicochemical prediction），网络数据库结构性检索等，这些服务使其成为最具有预报能力的一站式服务中心。第一个推出的交互试验室是在 1996 年，是以JAVA 语言为基础的网站，提供有限的预报和数据库检索。1998 年，推出其第二代产品（I-Lab II），可执行 50000 个预报和数据库检索，可以通过网络访问工业标准工具，目前已有一些学术机构租用这一系统[6]。

EPSRC（Engineering & Physical Science Research Council，UK）赞助了三项重点在于计算机科学（computer science:CS）的交叉学科协作研究项目（Interdisciplinary Research Collaborations:IRCs），这些重点项目支持了一些大学的关键的计算机科学研究组的长期研究的工作。其中对交叉学科研究协作项目的支持就是主要开发 e-Science 计划。此外还有 1）试验室内外环境 e-Science 高级网格界面项目；2）网格高级知识技术协作项目（CoAKTinG：Collaborative Advanced Knowledge Technologies in the Grid）；3）网络支持的知识服务：医科信息计量学面向解决问题的协作环境研究项目；4）医学图象和信号研究网格（MIAS-Grid：Medical Image and Signal Research Grid）。

1.3.3 虚拟试验室（virtual laboratory）技术发展现状

虚拟实验室的概念已有多年历史，有些虚拟实验室利用模拟方法，有些是采用实时视频/音频的方法。虚拟实验室是在网络环境下，可使用户在虚拟实验系统中进行实验。新加坡国立大学工程教研室的一个试验项目便提供了这样的实例。虚拟实验室可使学生和讲师进行不同尝试，检测各种教学与研究控制方法。电子工程系的学生可在任何时间，任何地方，如家里，登陆学校的虚拟实验室，只要具备性能良好的个人计算机和标准的网络浏览软件即可。新加坡国立大学工程教研

室采用实时视频/音频服务系统，通过若干试验，教授们认为，虚拟试验室有许多优势，首先不用担心实验设备不足或有时间限制；虚拟系统 24 小时可以利用，没有任何时间和空间限制。学生可以在用户界面键入试验参数，就可进行很精确的研究试验。假如在虚拟实验室工作遇到问题，可以随时给他的导师发电子邮件。专家们认为，尽管虚拟实验室优势很多，但虚拟实验室决不会替代实际实验室。专家预测，将来一定会有更多的虚拟实验室，或许 5 年之后，三分之一的试验可能都会以虚拟方式进行。它必定扩大了试验范围与试验时间。

另一个案例是美国 Johns Hopkins 大学化学工程系的虚拟工程/科学试验室，通过 JAVA 和World Wide Web（ WWW）在计算机上进行模拟工程和科学试验项目。在虚拟试验室可以向学生介绍试验内容，解决问题，进行数据收集和科学诠释[7]。

1.3.4 专家系统（Expert System）技术发展现状

在 2002 年北京召开的亚洲信息技术发展大会上亚、非、拉、美各大洲的代表云集北京，就农业信息化和经济全球化 、农业信息资源和数据库建设 、农业专家系统和决策支持系统、信息技术与农业自动化工程、3S 技术（GIS、GPS、RS）在农业中的应用、精准农业、农业和农村政府与企业上网工程、虚拟现实技术和虚拟农业系统、数字化图书馆建设、农业电子商务、农业信息战略和信息管理/知识管理等 11 个领域进行广泛深入的交流和探讨。充分展示了农业信息技术已进入高速发展的新时期。发展中国家的农业信息技术取得了重要的成就；发达国家的农业信息技术在一些重要领域达到了很高水平。

典型的因特网专家系统（Cisco Certified Internet work Expert：CCIE）是一个虚拟实验室数字方案，可进行导航与转换（routing and switching），通信与服务（communications and services）和试验室安全测试等。可为用户提供进行最高级工业技术鉴定需要的所有资源和工具，有 80 多个虚拟试验室数字化方案可供选择。每一个数字化方案都是按标准的拓扑学结构，对各类技术话题进行深入分析。成为用户良好的参考资源。

模拟模型和多媒体辅助试验系统是数字科研中的重要组成部分。专家们可根据物理资源形态和变化，进行模拟研究。如中美专家合作的进行的气象环境检测系统，日本进行的土地数字评价模型项目等。即使是在发展中国家这样的研究也非常普遍。

1.3.5 多媒体数据库管理系统技术发展现状

（1）数据模型技术

用形式化的方法来描述数据的逻辑结构和各种操作即数据模型。它由数据结构、数据操作集合和完整性集合组成。在多媒体数据库系统中，数据模型是核心。

它将用户与存储设备管理及具体的存储结构隔离开，并抽象出数据的静态和动态属性，为建立多媒体数据的使用工具（如编辑器等）提供形式化的基础。

数据库的数据模型可分为三类：一是面向记录的传统数据模型，如网络、层次和关系模型；二是注重描述数据及它们之间语义的语义数据模型；三是面向对象的数据模型。传统的数据模型以规范化关系为基础，但是通信网络多媒体数据库中存在大量的非结构化数据，及大量复杂对象的模拟、操作和推理，所以传统的数据模型不能满足多媒体数据库系统的需要。多媒体数据库通常要通过语义数据模型来更准确地表示数据与数据间的语义关系。主要的语义数据模型包括 E 一 R 模型、事件模型、SDM 模型和函数模型等。面向对象模型是目前最理想的多媒体数据模梨，它吸收了面向对象的编程技术和上述数据模型的优点，能提供对不同媒体的统一的用户界面，具有对复杂对象的描述能力和对象间关系的表示能力。

（2）数据存取技术

编码后的音频和视频数据往往是不定长的，因此，数据库系统需要具备处理大字段不定长数据的能力。为满足视频等连续媒休的时域约束需要．应设法提高磁盘的读写速度，减小读写的延时及抖动。其主要方法：一是使用逻辑卷的方法，将众多的磁盘阵列组合为一个逻辑设备，系统提供接口将读写命令分解到各物理设备，以并行的方式来提高速度；二是使用应用级存储方法，即应用程序具有数据存储的智能，能越过操作系统直接对数据定位；三是使用交叉存储方法，减少硬盘磁头的搜索和定位时间；四是根据数据在磁盘上的物理位置来调度数据提取次序，减少延时。

（3）多媒体的集成和编辑

在通信管理信息检索系统中，有时需要同时显示视频、图像、文本和播放音频等信息。这就需要定义好描述时间关系的时态规范，并利用规范对媒体间的时间关系进行描述，以确保多媒体数据在存储和传输过程中实现同步。媒体的集成可由多媒体编辑器来定义。由底层机制来实现。由干多媒体数据一般要经过压缩所以在对它们进行编辑时，数据的长度会引起变化，这不仅会给数据重新存入带来问题，而且还会引起媒体间的不同步，故应采取适当的方式，对被编辑的媒体进行局部的重编码，以保持原来的大小。

（4）基于内容的多媒体索引、查询和检索

多媒体信息可以通过形式化的方法，如标识符、属性、关键字等来检索。这种方法只与数据类型和数据结构有关，无需对内容作任何分析。但在通信网络管理智能化系统中，许多多媒体应用并不满足于这种简单的检索方式，它们需要对媒体进行语义内容的分析，以达到更深层次的检索。

1.3.6 专家论坛（list serve，discussion group，web forum…）技术发展现状

专家论坛是科学家进行信息交流的便捷方式。如英国自然环境研究委员会主办的全球网格论坛，就充分反映了国际数字科研的动态信息。全球网格论坛建立了ACE 网格研究组开发的先进的 e-Science 协作环境。这一项目将对如何进行建立虚拟协作环境，如何支持科学家在共享科学数据的大环境中共同工作进行研讨，促进这项工作的进展。利用交叉学科研究协作在虚拟协作环境中的经验和无线及移动技术，该项目将开发与当前网格服务与未来协作界面交互作用的基础结构。

2. 数字科研领域标准规范概述

2.1 数字科研领域标准规范主要范围

数字科研标准规范除通用网络有关标准规范，如系统标准规范、数据和多媒体数据存储与交换、通信协议、搜索引擎、检索标准规范外，专门的标准规范主要是软件构件技术的标准，即中间件标准，包括：CORBA 标准，ActiveX 标准，Java Beans 标准，和最新推出的网格技术标准如 Globus，Condor，Toolkit 和 SRB 网格中间件等。

2.2 数字科研领域标准规范制定宗旨

数字科研标准规范的主要作用是建立广泛的超越网络的数字平台，满足巨量信息计算、传递、共享的需要，实现为科研创造虚拟试验室的良好条件，全球共享科学知识与成果。为实现科研领域资源共享， 使信息基础结构更趋科学、合理。因此，数字科研领域标准规范制定的宗旨应是：以促进科学研究和共享知识为目的，积极推进标准规范建立，建立数字网格，最大范围的加快全球数字科研发展进程。

2.3 数字科研领域标准规范特点

数字科研是近几年发展起来的新事物，90 年代，以美英为代表的研究机构提出的 e-Science 的出现，为世界科学研究打开了新的研究天地。使数字科研领域标准规范问题日趋重要。要实现全球合作，数字科研，必须在统一的平台和系统中工作，利用一致的标准进行知识与数据互换和共享。许多研究机构除执行国际标准化组织的通用标准外，还积极参与标准的指定与推广，特别是在网络发展方面，对数字科研所及标准规范最多的组织，如 1994 年建立的 W3C（World Wide Web Consortium）拥有来自全世界的约 500 多成员组织，30 多个 W3C 工作组，推出了一系列规范标准，包括 XHTML，CSS，XML，DOM， PNG， CGM 和 CAD 等。这些标准规范的主要特点是：

（1）重视专门技术人员参与数字科研标准规范的研究

不管是英国还是美国，在数字科研领域都有一批专家专门从事数字科研和网格规范化研究和有关活动，包括设立数字科研论坛来广开思路，收集意见与建议，完善这一庞大复杂的系统工程的规范化建设。英国设有专门资金，鼓励数字科研群体与国际群体积极合作，这样可有效地向本国的群体传递了网格技术，促进了自身的发展，使其在全球网格论坛（Global Grid Forum）中对国际认同的网格协议的开发发挥着积极的作用。其“网格网”（GridNet）网络项目就鼓励其英国专家参与有关标准组织如全球网格论坛，IETF（Internet Engineering Task Force）和 W3C（World Wide Web Consortium）。以此推动本国和国际的标准化数字科研发展。

（2）谋求建立网格海量计算和虚拟试验环境

由于进行数字科研要求快速计算和数据处理的精确性的特点，数字科研标准规范是直接为科学研究的全球共享与合作服务的，任何中间件的选择与规定都是紧密围绕这一中心主题工作。它是从实践中来，到实践中去，任何时候都没有离开过实践的土壤。建立适宜的网格环境，制定相应的构件标准，可促进这一目标的实现。

（3）强化了数字科研标准规范的动态性更新性

和当前发展变化的信息技术一样，数字科研标准规范不断随着网格的发展而发展，不断随着科研的深入而变化，新的更加适合数字科研发展的规范建议层出不穷，使这一年轻的领域充满着朝气。不断推陈出新，标准规范越来越完善。

信息领域大环境的带动和影响，使数字科研标准规范随之进化。它即是信息大环境中的新事物，又将推动信息大环境的发展，特别是对未来 Internet 的作用和影响将是不可低估的。在这种情况下，数字科研标准规范将会不断改进，不断有所创新，成为保证数字科研发展的必要手段。国内外对数字科研标准规范的重视，充分说明了其不可替代的重要作用。

（4）注重数字科研标准规范的开放性

为了达到共享资源，联合研究的目的，明智的专家们在促进数字科研标准规范的形成与建立方面，表现了极大的合作热忱，开放使用，广开言路，吸收一切有益的东西，完善数字科研标准规范成了这一群体共同的愿望。在此基础上，他们尽可能开放中间件技术规范和建议，以求获得建议和合作共享的机会，使这方面的标准规范发展更快，推动数字科研的进展。

3. 数字科研领域标准规范发展状况

数字科研领域标专门标准规范的形成与发展，是同该领域学科专家和信息技术专家对此的高度重视分不开的。负责其标准建议、讨论、修订和发布等的主要标准组织是“全球网格论坛”，IETF（Internet Engineering Task Force）和 W3C（World Wide Web Consortium）。

3.1 数字科研领域网格计算现有标准规范的种类、主要内容与适用范围

高性能计算的应用需求使计算能力不可能在单一计算机上获得，因此，必须通过构建“网络虚拟超级计算机”或“元计算机”来获得超强的计算能力。20 世纪 90 年代初，根据 Internet 上主机大量增加但利用率并不高的状况，美国国家科学基金会（NFS）将其四个超级计算中心构筑成一个元计算机，逐渐发展到利用它研究解决具有重大挑战性的并行问题。它提供统一的管理、单一的分配机制和协调应用程序，使任务可以透明地按需要分配到系统内的各种结构的计算机中，包括向量机、标量机、SIMD 和 MIMD 型的各类计算机。NFS 元计算环境主要包括高速的互联通信链路、全局的文件系统、普通用户接口和信息、视频电话系统、支持分布并行的软件系统等。

元计算被定义为“通过网络连接强力计算资源，形成对 用户透明的超级计算环境”，目前用得较多的术语“网格计算（grid computing） ”更系统化地发展了最初元计算的概念，它通过网络连接地理上分布的各类计算机（包括机群）、数据库、各类设备和存储设备等，形成对用户相对透明的虚拟的高性能计算环境，应用包括了分布式计算、高吞吐量计算、协同工程和数据查询等诸多功能。网格计算被定义为一个广域范围的“无缝的集成和协同计算环境”。网格计算模式已经发展为连接和统一各类不同远程资源的一种基础结构。在进行网格计算和数字科研中主要涉及的标准规范有：

XHTML（Extensible Hyper Text Markup Language，扩展超文本标记语言）是 Web 世界的描述语言，是所谓的可扩展超媒体标记语言，是在 HTML4.01 的基础上于 2000 年 1 月成为建议规范的，修改版 XHTML1.1 于 2001 年 5 月推出，可具此开发不同 XHTML 文档，适宜各类 device 或用户群体。如 XHTML Basic（2000，12 推出），就是为移动手机，PDAs，Pagers 和 settop boxes 的网络用户开发的 XHTML 文档。XHTML 正在变得日益普及而且大有取代 HTML 之势。

CSS（Cascading Stylesheets，层叠样式表）是一种制作网页的新技术，现在已经为大多数的浏览器所支持，可控制 HTML 以及任何可扩展置标语言内容。成为网页设计必不可少的工具之一。使用 CSS 能够简化网页的格式代码，加快下载显示的速度，也减少了需要上传的代码数量，大大减少了重复劳动的工作量。

DOM（Document Object Model，文档对象模型）。DOM 是“可扩展标记语言”或 XML 的基础。XML 文档具有称为节点的信息单元的层次结构；DOM 是描述那些节点和节点间关系的方式。

CGM（Computer Graphics Metaware）计算机图形元文件是一种图形交换标准。ISO :8632-1986。面向图形的输出文件格式。

CAD（Computer Aided Design）计算机辅助设计。

SVG（Scalable Vector Graphics，可缩放矢量图形）。它是用来描述二维图形的

XML 语言，最重要的特点是它不是一个私有格式。SVG 是开放标准，由 W3C（World Wide Web Consortium）建议。SVG 图形是可交互的和动态的，可以在 SVG 文件中嵌入动画元素或通过脚本来定义动画。

PNG（Portable Network Graphics）是可携式网络图像。作为 GIF 和 JPEG 的替换格式 — 它支持这两种格式的所有优点，但它似乎在使 Web 浏览器更大、更好和更健壮的过程中被遗忘了。仅在最近才有浏览器可以不需外挂程序地支持它。

RDF（Resource Description Framework，资源描述框架），在 W3C（World Wide Web Consortium）主持下发展起来的资源描述框架， 它是一种基于 XML 元数据语言。是一种能实现元数据编码、交换和再使用的基础结构。该基础结构通过设 计一种能支持语义、语法、结构的通用惯例的机制来实现元数据的互操作。

3.2 数字科研领域专家系统现有标准规范的种类、主要内容与适用范围

专家系统是基于知识的系统（Knowledge-Based System），一个完整的专家系统应由知识库（ Knowledge-Base）、数据库（Bata Base）、推理机（Inference Engine）、知识获取模块（ Knowledge-Acquisition Module）和解释接口（Explanatory Interface）组成。知识库中存放系统求解问题所需求的加识，数据库用来存储初始证据和推理过程中得到的各种中间信息，推理机是一组程序，用来控制和协调整个系统，它通过输入的数据，利用加识库原有知识按一定的推理策略解决所提出的问题，加识获取模块就是学习模块，它为修改和扩充加识库存的原有加识提供相应的手段、解释接口是用户与专家系统交互的环节，负责对推理给出必要的解释，便于用户了解推理过程，为用户向系统学习和所作所为系统提供方便，具有解释功能是专家系统区别于其它计算机程序的标志。

3.2.1 系统平台标准与规范

（1）软件平台

操作系统：

网络：遵循 TCP（Transmission Control Protocol）/IP（Internet Protocol）协议。

单机：Windows98、Unix、Linux

服务器：WindowsNT、Windows2000、Unix、Linux

客户机：Windows98、Unix、Linux

数据库应用系统：SQL Sever；Foxpro/Access for Windows 等。

（2）硬件平台

单机：PII 以上的 PC 机（内存 64M、主频 166、硬盘 2G）。

网络（基于 PC 机的网络系统）：服务器、客户机、集线器、（网卡/Modem）、网线。

服务器：PII 及以上的 PC 服务器。

客户机：PII 及以上的 PC 机。

（3）编程工具：Visual C++；Visual Basic; Power Builder；Delphi6; Java；Html；Asp 等。

浏览器：Netscape；IE 等。

地理信息系统：Mapinfor；Arcview；Autocad；Mapgis；Arc/Info，SICAD 等。

动画开发工具：Authorware；ToolBook；3DMAX 等。

通用系统标准规范在本项目中有专门介绍，本文不予详述。

3.2.2 软构件开发类的三个标准接口

（1）使用 Windows 操作系统进行构件开发时应遵循 COM/DCOM 接口标准。

COM（Component Object Model）即组件对象模型，是一种平台独立的、分布式的、面向对象的可创建交互式二进制软件组件的系统。它不是一种编程语言，而是一种编程规范。它是建立在 OLE 基础上的“对象模型”，COM 允许使用一些定义良好的软件模块在系统中共存并相互作用，从而构成更大和更复杂的系统。其基本思想是面向对象和软件重用，这也是目前软件编程思想的主要发展方向。

（2）使用 Unix 操作系统进行构件开发时应遵循 CORBA 构件标准。

CORBA（Common Object Request Broker Architecture 公共对象请求代理体系结构）是由 OMG 组织制订的一种标准的面向对象应用程序体系规范。或者说 CORBA 体系结构是对象管理组织（OMG）为解决分布式处理环境（DCE）中，硬件和软件系统的互连而提出的一种解决方案；OMG 组织是一个国际性的非盈利组织，其职责是为应用开发提供一个公共框架，制订工业指南和对象管理规范，

目前 corba 主要适合于分布式跨平台的信息管理应用；比较典型的应用是大规模的企业信息管理和事务处理应用以及行业解决方案（数据库中的中间件技术）和在网管系统中多厂商互连互通解决方案

（3）进行垮平台开发时应遵循 Java Beans 构件标准。它非常重要，允许开发人员由软件组件构造复杂系统。这些组件可以自己开发，也可以由一个或多个不同厂商提供。Java Beans 定义了组件块互操作行为的体系结构。

3.2.3 数据库接口标准与规范

ODBC（ ODBC，Open DataBase Connectivity）开放数据互连是 Microsoft Windows 开放服务体系结构（WOSA，Windows Open Service Architecture）的一部分，是一个数据库访问的标准接口。它提供了对关系数据库访问的统一接口，实现了对异构数据源的一致访问。ODBC 技术的推出为异构数据访问提供了解决方案。引起了数据访问方面的巨大变革，ODBC 成为最通用的数据访问技术。随着异购数据访问技术的不断发展，逐渐开发出了 OLE DB 和 ADO 等技术，但 ODBC 仍然扮演着重要的角色。ODBC 可以在 Windows、Unix 等多操作系统上使用。

OLE DB（Object Link and Embedding Database）是一个组件数据库访问接口，它把对数据源的操作过程分为客户和服务器两个方面，提供了比传统数据库更高的效率。它可以访问任何格式的文件系统（不管是关系的还是非关系的），只要这些数据源提供自己的数据提供程序，即使是用户自己定义的文件格式甚至提供数据的硬件等都可以通过该接口来访问。

ADO（ActiveX Data Object，ActiveX 数据对象）是 Microsoft 提供的一种面向对象、与语言无关的（ Language-Neutral）数据访问应用编程接口。

3.2.4 知识库接口标准与规范

目前还没有一个标准的接口，但是有一些知识的获取的方法。

（1）知识类型（types of knowledge）

领域是千变万化的，因此知识的类型也是纷繁复杂的，总结起来有以下几种，如表 1-1 所示。

①过程性知识，描述一个问题是怎样被解决的，此类型知识对如何做提供指导。规则、策略、过程等是专家系统中使用过程性知识的典型类型。

②描述性知识，描述对于一个问题我们知道些什么，以及问题求解当前的状况。包括简单陈述，如真假断言等。

③元知识，关于知识的知识。元知识常被用来使用（检索）知识库中最适合求解当前问题的知识，专家通常利用元知识，使问题求解向着最有希望的方向进行，以提高系统求解效率。

④启发性知识，主要用来描述经验规则，这些经验规则引导推理过程。启发性知识也称为浅层知识，它是经验的，表示专家用于求解问题的经验。专家有时也把基本知识或称为深层知识，如基本原理、基本关系等总结成启发性知识辅助问题求解。

⑤结构化知识，描述知识结构。这种类型知识描述专家关于一个问题的总体认知模型。

表 1 知识类型

知识类型	描述对象
过程性知识	规则、过程等
描述性知识	概念、对象、事实等
元知识	关于知识的知识，如如何使用知识
启发性知识	经验规则
结构化知识	规则集、概念、关系

人工智能学者利用认知心理学的成果，提出了一些知识表示的实用技术，领域专家和知识工程师的任务就是选择最适合领域问题的知识表示技术。

（2）知识获取方法（the way of knowledge acquisition）

①手工知识获取（manual knowledge acquisition） 开发人员与领域专家通过交谈或问卷 的形式获取知识，是获取知识的基本方法。

②半自动化知识获取

知识编辑器：知识录入、词法语法检查、查询、修改等功能。

知识库求精器：对知识进行求精。

③自动知识获取（automatic knowledge acquisition） 使用机器学习、知识发现、数据采掘（Data Mining）等方法从大量事例或数据中抽取知识。

（3）推理机构件

推理机构件属于功能构件，不可再分割。推理机构件的分类依据：控制策略和知识表示。可以有以下种类的推理机构件：确定性正向推理机构件，确定性反向推理机构件，MYCIN 正向推理机构件，MYCIN 反向推理机构件，面向凸函数证据理论模型的正向推理机构件，面向凸函数证据理论模型的反向推理机构件，面向加权模糊逻辑的正向推理机构件，面向加权模糊逻辑的反向推理机构件，支持两级不确定性的正向推理机构件，支持两级不确定性的反向推理机构件。可用表 2 描述：

表 2 推理机构件

控制策略 / 知识表示	正向推理	反向推理
确定性知识表示	确定性正向推理机构件	确定性反向推理机构件
MYCIN 知识表示	MYCIN 正向推理机构件	MYCIN 反向推理机构件
凸函数证据理论模型	面向凸函数证据理论模型的正向推理机构件	面向凸函数证据理论模型的反向推理机构件
加权模糊逻辑	面向加权模糊逻辑的正向推理机构件	面向加权模糊逻辑的反向推理机构件
两级不确定性知识表示	支持两级不确定性的正向推理机构件	支持两级不确定性的反向推理机构件

在构件结构中最重要的概念是接口，接口是集合在同一个名称下相关方法的集合，构件之间的通讯正是基于接口的，它可以说是构件和其客户之间严格类型化的契约。目前有人开发标准的推理机构造接口，即标准的“即插即用”接口，目的是提高软件开发的效率，同时使应用程序的定制更为灵活、更易维护。这样，可使用任何自己熟悉的开发工具从外部访问接口公布的对象，轻松组装应用系统。

3.3 数字科研领域多媒体数据库系统现有标准规范的种类、主要内容与适用范围

3.3.1 系统平台标准规范

（1）软件平台

操作系统：

网络：遵循 TCP（Transmission Control Protocol）/IP（Internet Protocol）协议。

单机：Windows98、Unix、Linux

服务器：WindowsNT、Windows2000、Unix、Linux

客户机：Windows98、Unix、Linux

数据库应用系统：SQL Sever；Foxpro/Access for Windows、Sybase、Oracle、Informix 等。

（2）硬件平台

单机：PII 以上的 PC 机（内存 64M、主频 166、硬盘 2G）。

网络（基于 PC 机的网络系统）：服务器、客户机、集线器、（网卡/Modem）、网线。

服务器：PII 及以上的 PC 服务器。

客户机：PII 及以上的 PC 机。

能实现超大容量存储的介质主要有 RAM、硬盘（阵列）、光盘（库）、磁带（库）等。

（3）编程工具：Visual C++；Visual Basic；

浏览器：Netscape；IE 等。

动画开发工具：Authorware；ToolBook；3DMAX 等。

3.3.2 软构件开发类的标准接口

Windows 下的 Direct X 可对计算机的硬件驱动进行优化，对硬件起到加速作用.其中 DirectDraw 管理视频输出，DirectSound 管理声音输出，DirectPlay 管理网络通信，Direct3D 管理三维图形。

具有多媒体的接口和交互功能

不同的媒体与数据库有不同的交互接口；每种媒体均有自己的存取和表现方法；用户对同一种媒体的表现形式可能有各种不同的要求，如不同的图像尺寸。不同的播放帧率等。所以系统必须具有良好的用户界面和交互功能及容易使用的数据库语言，以方便用户对多媒体数据的操作。

MPEG（moving picture experts group）是 ISO/IEC 标准。MPEG__7 被正式命名为“多媒体内容描述接口”，目的是为描述多媒体数据制定一个标准，这些数据将

支持一定程度的信息意义的解释，它能够被一个设备或计算机代码传递或存取。该标准不仅提供描述多媒体信息（静止图像、图形、3D模型、音频、语音、视频以及有关这些元素在多媒体表现上是如何组合成的），还能在许多不同环境下支持不同的应用，即可提供一种灵活的可扩展的视听数据描述框架。

多媒体数据库的建立还涉及数据压缩与存储。并且这种存储方式能够为以后的数据处理带来最大的方便。由于 ASP／ADO 技术提供丁高效率的 ODBC 或 OLE－DB 数据来源的连接功能，因此凡是支持 ODBC 的数据库，如：Access、SQL ServerO、racle、Informix 都可以用来存储多媒体数据。

3.3.3 压缩／还原标准

目前常用的图象、视频和音频处理软件都可使用行业或标准格式的压缩编码技术以产生或读取显示相应的文件。如采用 LZW 压缩算法的 TIF 和 GIF 格式，采用 JPEG 和 GIF 标准的静态图象格式，MPEG1 标准的活动图象格式（VCD），MPEG2 标准的活动图象格式（DVD），MP3 的声音压缩格式等。但另一方面，大多数工程领域的 CAD 软件由于在图形数据处理方面存在的交互性、实时性、相关性等特性和差异，大多是以非压缩格式输出相关文件。其中最典型的如 AutoCAD 的 DWG、DXF 及 IGES 格式等。

DWG--AutoCAD 的 DWG 格式成为了二维工程图事实上的一种标准。

DXF--数字交换格式。用 ASCII 码或二进制文件存储矢量数据的格式，AutoCAD 和其他一些 CAD 软件用其来进行数据交换，DXF 文件可以转换成 ARC/INFO 图幅。

IGES（Initial Graphics Exchange Specification）--1980 年，由美国国家标准局主持成立了由波音公司和通用电气公司参加的技术委员会，制订了基本图形交换规范 IGES。作为较早颁布的标准，IGES 被许多 CAD/CAM 系统接受，成为应用最广泛的数据交换标准。

4. 数字科研领域与数字图书馆领域标准规范的相互联系和可能影响

全球信息化的发展趋势，冲击人类社会各领域，首当其冲的是图书馆界，世界数字图书馆建设项目日新月异。与此同时，数字科研悄然兴起，而且是站在信息技术最高起点上。但是，无论是数字图书馆领域还是数字科研领域，其研究和运作的对象均是数字化数据，数字图书馆环境是规范化建设数字化、网络化信息资源与服务，数字科研环境更侧重高效地处理和计算网络资源与共享，数字科研有必要需求数字图书馆的信息服务，两者的标准规范均是要研究建立相应的信息基础结构，这两者之间有相通之处又有区别。因此，数字科研领域标准规范对数字图书馆领域数据信息加工、存储和提供服务有重要参考价值，有些可作为数字图书馆标准规范使用。

4.1 数字图书馆领域关键技术与数字科研领域标准规范的通用性和共识点

4.1.1 数字图书馆领域标准规范可为数字科研提供借鉴

数字图书馆领域标准规范如元数据著录和可扩展置标语言是帮助查找、存取、使用和管理信息资源的信息，既适合于电子资源、又适合于非电子资源，不仅包括编目信息，也包括其他管理和存取资源的信息。XML 描述信息资源或数据对象，其目的在于使用户能够发现资源，识别资源，评价资源，而且对相关的信息资源进行选择、定位和调用，追踪资源在使用过程中的变化，实现信息资源的整合、有效管理和长期保存。

面对网络海量信息资源的整序需求，资源发现已成为 Internet 应用的瓶颈与焦点。虽说网络上已有多种搜索引擎，并辅之以布尔逻辑检索等方法，但同样不能满足用户对信息检索的需求，尤其不能满足用户对特定信息的准确检索。例如 Yahoo、Lycos、Altavista 等，这些搜索引擎的工作方式，是通过自动搜索程序来抓取网页信息，然后以自动拆字 （词 ）做索引的方式建立数据库，不能有效地过滤资源，造成检索结果数量大而有用信息少的弊病。

因此，在以网络资源建设与服务为主体的数字图书馆中，元数据和 XML 已成为数字图书馆建设的关键技术之一，被数字图书馆当作进行知识组织和资源发现的工具，为网络信息资源的组织提供了重要手段。同时，世界上已开发出并付诸使用的元数据有多种，例如：美国联邦地理数据委员会的地理元数据项目（FGDC，Federal Geographic Data Committee）、编码文档描述 （EAD，Encoded Archival Description）、频道定义格式 （CDF，Channel Definition Format）、教育管理系统（IMS，Instructional Management System）、全球信息定位服务 （GILS，Global Information Locator Service）、博物馆信息计算机交换标准框架 （CIMI，Computer Interchange of Museum Information）、互联网内容选择平台 （PICS，Platform for Internet Content Selection）和已成为美国国家标准的都柏林核心元数据 （DC，Dublin Core）等等。

在数字图书馆中，支持数字资源长期保存的数据即保存元数据，其框架结构体现较突出的，如 OAIS（Open Archive Information System），具有接收信息、保存信息并提供信息服务等主要功能。OAIS 中的 PDI（Preservation Description Information），描述内容信息的特征以保证内容信息的长期保存，它包括下列内容：出处信息（provenance），描述内容信息的来源，产生以后的监管人、加工处理历史等；上下文信息（context），记录内容信息与信息包以外其他信息的联系；参考信息（reference），包括对资源描述的附加信息和资源标识符 （用来标识内容信息的惟一性，例如一本书的 ISBN 号 ）；固定信息（fixity），包括用于认证的信息，

例如数字签名等，以保护内容信息不受篡改。关于 OAIS，国际元数据界已逐渐达成共识。例如中国国家图书馆制订的《中文元数据标准方案》，就以 OAIS 作为总体框架；美国 OCLC/RLG 已经正式提出保存元数据的概念并企图要在 OAIS 信息模型的基础上制订一个保存元数据的标准框架；澳大利亚国家图书馆也在保存元数据的研究方面做了许多努力。保存元数据这一概念的提出对数字化资源的长期保存具有深远意义，相信在对数字科研领域服务中也将形成共性操作。

4.1.2 信息系统之间的互操作“中间件”，为数字科研利用数字图书馆资源提供优化途径

目前，数字图书馆领域对不兼容的数据格式和数据结构问题，通过元数据的互操作即主要是通过语义互操作和结构与语法的互操作来实现。各种元数据间的互操作，能够真正支持最广泛范围的资源发现、检索和使用，并直接影响信息的共享、互换以及透过系统、语言和地理位置的界限而访问的可能性。通常包括以下几种途径：（1）开放档案协议 OAI（Open archive initiative），此协议始于电子出版（E-Print），目标是通过电子出版团体内部系统的互操作来达到团体内的信息共享，后来将目标扩大为寻求一种简便的方法来实现不同的数字资源系统间的开放检索（也就是跨系统检索 ），使不同元数据方案下相等或近似相等的元数据元素相互映射，以实现语义上的互操作。（2）资源描述框架 RDF 与可扩展标记语言 XML：由 W3C 推出的 RDF 是一套描述资源及其属性和属性值的模型，其制定的目的主要是为元数据在 Web 上的应用提供一个基础结构，以方便不同元数据间的互操作。可扩展标记语言 XML 作为元数据的编码标准，提供了元数据在语法层次上的互通性，使它跨越特定平台、特定系统的限制。使用 RDF/XML 命名域的概念，在创建一个元数据格式时，借用其他元数据集的某些元素，可以减少重复劳动并增强元数据格式间语义互通性，方便互操作的实现。

4.1.3 数字图书馆建设与数字科研领域通过分布式网络利用大型数据资源和计算资源及高度形象化数据之间趋于标准共用

数字图书馆的运作，无论是数据的加工、存取，信息的浏览、检索，还是资源的整合与长期保存都是以元数据为基础实现的。数字科研正在出现的网格（Grid）技术是为科学研究中涉及全球范围实验、计算、数据和视听资源的综合与利用需求而开发的信息基础结构。数据网格是最基础的信息结构，其命名的透明性、定位的透明性、协议的透明性和时间的透明性以及提供的目录服务、注册与发布、信息发现、存储资源代理服务、身份认证与访问控制、调度和方法执行等，对数字图书馆和数字科研均可构成互通。因此，数据网格的关键技术即元数据目录和存储资源元文档，对数字图书馆和数字科研两者之间有着密切相互联系和影响。

4.1.4 数字科研对多媒体信息的处理标准也适用于数字图书馆

如数字科研在处理运动图像数据时的定位基于 MPEG_7 标准，数字图书馆对该类数据处理也是基于此标准。MPEG__7 标准的应用，为数字科研和数字图书馆获得高效率的运动图像存储和传输、智能化的管理和保护提供了基础。其主要应用方面有：（1）数字图书馆和数字科研中有关教学和教学管理内容，MPEG__7 可以促进教育领域采用音频和视频等媒体资料进行形象直观的教学和培训。（2）数字图书馆的图像目录、音乐词典、影片视频和音频档案等。（3）数字科研档案的档案资料、历史图片、视频录像和原始档案电子版等。电视和电影档案中保存有大量的各种格式的多媒体资料，如数字、模拟磁带和胶片、CD 等，需要按照国际标准格式进行存储和交换。另外，要对大量旧的模拟视听资料进行数字化，在数字化和压缩阶段，可以在数据库中使其包含基于内容的索引特征。对于新的视听媒体，在视频生产的各个阶段可以把描述信息附加在视频流上，从而极大地提高了用手工进行有限词汇注释的质量和生产率。（4）数字科研中专题研究如建筑结构设计、船体模型、内部结构、工艺表现等媒体的搜索、储存和表象。（5）视听应用中对多媒体视听领域信息的采集、内容处理、视频编辑、新闻检索、语音和图像识别管理等。（6）科研服务中人类特征的识别、以及电子商务。

4.2 数字科研领域与数字图书馆领域标准规范在发展策略方面可相互借鉴

4.2.1 数字科研领域标准规范起步晚，发展快，拥有技术和财政实力

数字科研领域标准规范建设一般采用政府和项目支持的做法，因此无论是美国、英国还是欧盟都拥有技术和财政实力支持，因此虽然起步晚，但是发展快。数字图书馆建设应充分利用已有和现行国际主流标准规范，借鉴数字科研领域标准规范形成和发展的经验，积极参与已有和现行标准规范的更新和修订，制定数字图书馆领域尚未成型或有待改进的专门标准规范。数字图书馆综合了文本、图像、声音等信息的数据库即多媒体数据库，可处理的信息类型发生了重大变化。可采用数字科研多媒体系统中标准规范，并参与对这些标准的完善和改进工作，使未来数字图书馆与数字科研的多特征检索技术标准规范形成共同发展。

4.2.2 数字图书馆建设同数字科研领域标准规范的不同点在于其不同的研究对象

数字图书馆建设的研究对象是数字化资源和用户，目的是整合资源，为用户服务。数字科研的目的是建立宽松的网络环境，为科学家提供共同研究的快捷条件。因此，数字科研领域标准规范着眼于更大范围的网络环境的建立和高速计算能力的实现，以辅助科研的进展。数字图书馆领域标准规范则重点在于网络环境

下，经过整合的现成的信息资源产品的发布与提供，即不但注重信息的管理，也注重信息的发布和共享。为此，它们的资源管理系统中间件肯定有着许多的不同，数字科研领域所需要的某些数据，应该是数字化图书馆提供的，包括多媒体信息，从而，促进数字科研的发展。因此，数字科研的标准规范和数字图书馆标准规范在数据管理上应该是一致的，在系统中间件的设置上是有所不同的。

4.2.3 数字科研领域标准规范的开放性机制是数字图书馆标准规范建设应仿效的

数字科研领域标准规范建设的共享研究方式可做为我国数字图书馆标准规范建设的参照，采取开放性的标准规范眼法机制，有利于标准规范的广泛应用和用户参与，有利于标准规范的形成和执行。同时，数字科研领域集中体现的网格技术应用调度和支持，是不分地域地获取世界范围的数字科研信息资源的非常重要的一种信息基础结构。对数字图书馆提供信息服务有很好的借鉴作用。

总之，数字科研是以分布式海量数据计算为基础，基于智能技术的大型、开放、分布式网络环境，欲为全球科学家提供数字科研理想的基础结构。这与数字图书馆的目标有着本质上区别，但数字图书馆网络信息获取、管理和共享，又与数字科研领域对分布在世界各地的科研数据源共享有着相通之处，这两个领域在制定各自发展策略时应充分互相参考，择优利用，资源共享。

参考文献

[1] E-science overview， Natural environment research council. E-science in NERC. www.Nerc.ac.uk/funding/escience （检索日期：2003-03-15）

[2] Promotion of Grid Middleware Development. http://www.escience-grid.org.uk （检索日期：2003-03-20）

[3] Condor research project， Condor high throughput computing， http://www.cs.wisc.edu/condor，（检索日期：2003-03-20）

[4] The globus project. Information Services in the Globus Toolkit 3.0 Release. http://www.globus.org/toolkit/gt3-factsheet.html （检索日期：2003-03-15）

[5] Project overview. Datagrid. http://web.datagrid.cnr.it （检索日期：2003-03-15）

[6] ACD/I-Lab. www.acdlabs.com/ilab （检索日期：2003-03-15）

[7] Michael Karweit，A virtual engineering/science laboratory course，Dept Chemical Engineering Johns Hopkins University

[8] Mei Fangquan，Asian agricultural information technology and management: Proceedings of the 3rd Asian conference for information technology in agriculture Oct.26-28，2002，Beijing，China，China Agricultural Scitech Press，605p.

[9] 信息技术标准目录总汇. 中国电子技术标准化研究所等编. 北京：中国标准出版社，2001，210 页

[10] Extensible Markup Language. W3C architecture domain. www.w3.org （检索日期：2003-03-22）

[11] 刘刀桂，孟繁晶. Visual C++ 实践与提高--数据库篇. 北京，中国铁道出版社，2001

[12] 马小虎，张明敏 严华明等. 多媒体数据压缩标准及实现. 北京，清华大学出版社，1996

[13] 顾景文，张桦，黄晓生. B /S 模式多媒体数据库应用的文件处理. 计算机工程，第 27 卷，第 2 期