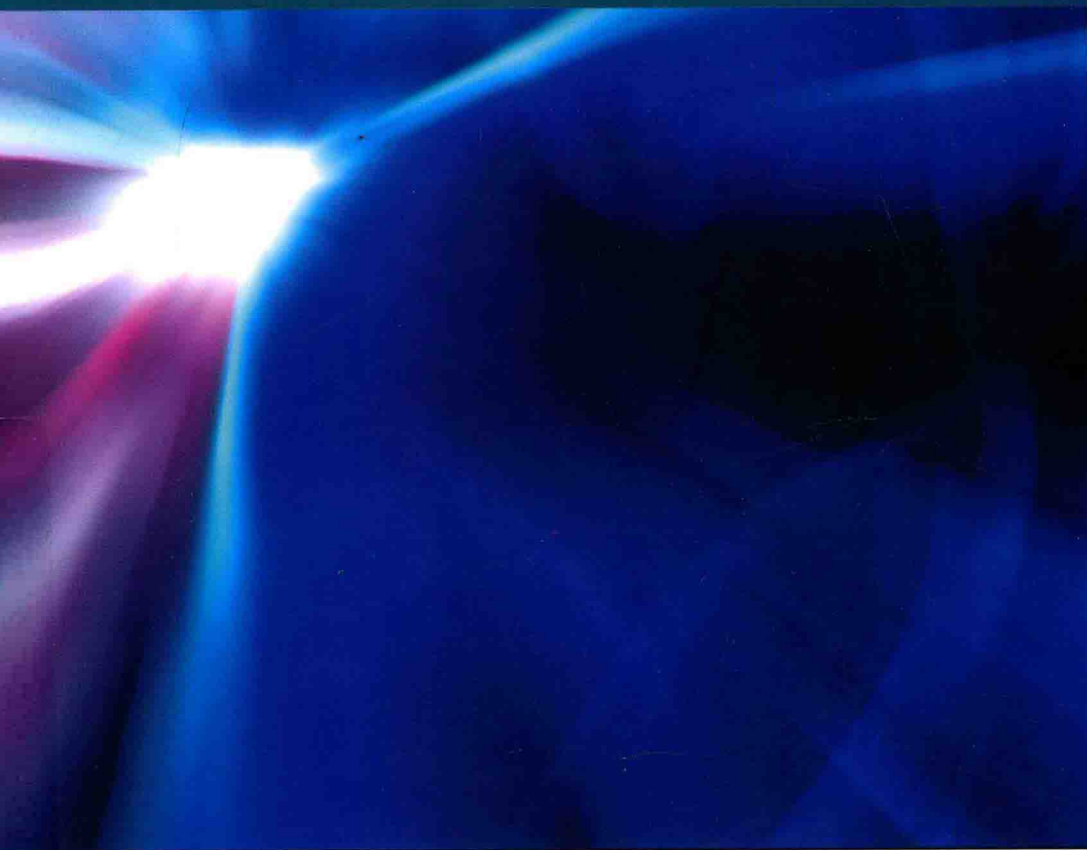


国家数字图书馆工程标准规范成果

图书馆数字资源统计 标准和应用指南

吕淑萍 罗云川 主编



國家圖書館出版社

图书馆数字资源统计标准和应用指南

吕淑萍 罗云川 主编

国家图书馆出版社

图书在版编目(CIP)数据

图书馆数字资源统计标准和应用指南/吕淑萍,罗云川主编. —北京:国家图书馆出版社, 2010. 8

(国家数字图书馆工程标准规范成果)

ISBN 978-7-5013-4394-2

I. ①图… II. ①吕…②罗… III. ①数字技术—应用—图书馆管理:资源管理—统计—标准—指南 IV. ①G250.76-65

中国版本图书馆 CIP 数据核字 (2010) 第 134275 号

责任编辑:金丽萍 王涛 高爽

书名 图书馆数字资源统计标准和应用指南

著者 吕淑萍 罗云川 主编

出版 国家图书馆出版社(原北京图书馆出版社)

(100034 北京市西城区文津街7号)

发行 010-66139745 66151313 66175620 66126153

66174391(传真) 66126156(门市部)

E-mail btsfxb@nlc.gov.cn(邮购)

Website www.nlcpress.com→投稿中心

经销 新华书店

印刷 河北三河弘翰印务有限公司

开本 787×1092(毫米) 1/16

印张 12.25

版次 2010年8月第1版 2010年8月第1次印刷

字数 110(千字)

书号 ISBN 978-7-5013-4394-2

定价 58.00 元

丛书编委会

主 编：国家图书馆

编委会：

主 任：周和平

执行副主任：詹福瑞

副主任：陈 力 魏大威

成 员(按姓氏拼音排名)：卜书庆 贺 燕 蒋宇弘

梁蕙玮 龙 伟 吕淑萍 申晓娟 苏品红

汪东波 王文玲 王 洋 杨东波 翟喜奎

赵 悦 周 晨

本书编委会

主 编：吕淑萍 罗云川

编 委：韩新月 申晓娟 韩 超 冷 熠 赵保颖 琚存华 王丽华
范 雪 王 荟 李志尧 宋仁霞 李 杉 曾 燕 刘婷婷
赵 悦 胡 洁

总 序

数字图书馆涵盖多个分布式、超大规模、可互操作的异构多媒体资源库群,面向社会公众提供全方位的知识服务。它既是知识网络,又是知识中心,同时也是一套完整的知识定位系统,并将成为未来社会公共信息的中心和枢纽。数字图书馆建设的最终目标是实现对人类知识的普遍存取,使任何群体、任何个人都能与人类知识宝库近在咫尺,随时随地从中受益,从而最终消除人们在信息获取方面的不平等。“国家图书馆二期工程暨国家数字图书馆工程”是国家“十五”重点文化建设项目,由国家图书馆主持建设,其中国家数字图书馆工程的建设内容主要包括硬件基础平台、数字图书馆应用系统和数字图书馆标准规范体系。

标准规范作为数字图书馆建设的基础,是开发利用与共建共享资源的基本保障,是保证数字图书馆的资源和服务在整个数字信息环境中可利用、可互操作和可持续发展的基础。因此,在数字图书馆建设中,应坚持标准规范建设先行的原则。国家数字图书馆标准规范体系建设围绕数字资源生命周期为主线进行构建,涉及数字图书馆建设过程中所需要的主要标准,涵盖数字内容创建、数字对象描述、数字资源组织管理、数字资源服务、数字资源长期保存五个环节,共计三十余项标准。

在国家数字图书馆标准规范建设中,国家图书馆本着合作、开放、共建的原则,引入有相关标准研制及实施经验的文献信息机构、科研机构以及企业单位承担标准规范的研制工作,这就使得国家数字图书馆标准规范的研制能够充分依托国家图书馆及各研制单位数字图书馆建设的实践与研究,使国家数字图书馆的标准规范成果具有广泛的开放性与适用性。本次出版的系列成果均经过国家图书馆验收、网上公开质询以及业界专家验收等多个验收环节,确保了标准规范成果的科学性及实用性。

目前,国内数字图书馆标准规范尚处于研究与探索性应用阶段,国家图书馆担负的职责与任务决定了我们在数字图书馆标准规范建设方面具有的责任。此次将

国家数字图书馆工程标准规范研制成果付梓出版,将为其他图书馆、数字图书馆建设及相关行业数字资源建设与服务提供建设规范依据,对于推广国家数字图书馆建设成果,提高我国数字图书馆建设标准化水平,促进数字资源与服务的共建共享具有重要意义。

国家图书馆馆长 周和平
2010 年 8 月

目 录

图书馆数字资源统计标准	(1)
1 引言	(3)
2 范围	(3)
3 规范性引用文件	(4)
4 术语及定义	(4)
5 数字资源统计指标体系	(9)
6 数字资源统计数据采集	(28)
7 数字资源统计管理	(35)
附录 A(规范性附录) 数字资源描述指标	(37)
附录 B(规范性附录) 用于进一步统计分析的类目	(39)
附录 C(规范性附录) 主要的内容单元类型及计量单位	(45)
附录 D(资料性附录) 数字资源统计报表示例	(47)
图书馆数字资源统计标准应用指南	(61)
1 指南概述	(63)
2 数字馆藏统计指南	(63)
3 数字资源服务统计指南	(87)
4 数字资源相关设施统计指南	(91)
5 数字资源经费统计指南	(95)
附录 1(资料性附录) 数字资源总况统计清单	(101)
附录 2(资料性附录) 数字资源多维度统计报表	(140)
附录 3(资料性附录) 自建数字资源业务管理环节的统计(推荐)	(175)
附录 4(资料性附录) 自建数字资源业务管理环节统计清单(推荐)	(178)
后 记	(181)

图书馆数字资源统计标准

1 引言

随着信息环境的不断发展,用户信息需求和信息利用习惯都发生了巨大转变,数字资源在信息服务中发挥着越来越显著的作用,并已成为图书馆信息资源的重要组成部分,引发了图书馆对数字资源统计的迫切需求。同时,由于数字资源类型多元化、数据结构多样化和复杂化、资源整合与服务整合快速发展等各种因素综合影响,数字资源管理中的统计问题逐渐突显出来。

关于图书馆统计,国际上已经开展了若干相关标准规范的研制,其中产生重要影响的成果有:ISO2789:2006 Information and documentation—International library statistics、ANSI/NISO Z39.7:2004 Information Services and Use: Metrics & statistics for libraries and information providers—Data Dictionary 和 Measures for Electronic Resources(E-Metrics) Complete Set 等。从总体上看,上述研究大多着眼于图书馆宏观统计,其中有关数字资源的统计规范远不能满足我国图书馆界的实际需求,而国内尚未形成图书馆界共同认可的数字资源统计标准,各馆所采用的统计指标体系和统计方法均存在相当大的差异,降低了数据的比较、分析和决策价值。

在国内外已有研究的基础上,遵循适用性和科学性的原则制定本标准:

(1) 适用性

国家图书馆二期工程暨国家数字图书馆工程拥有大量数字资源及各类应用平台,制定图书馆数字资源统计标准,满足标准化、规范化的管理需求,同时,兼顾国际、国家和行业同类标准,并符合国家数字图书馆工程初步设计和详细设计需求,与该项工程其他子项目的相关成果保持一致性。此外,考虑到统计数据的通用性和可比较性,数字资源统计标准具备与我国数字图书馆标准与规范建设项目(CDSL)、中国高等教育数字图书馆(CADLIS)及其他大型数字图书馆建设项目(如:NSTL)等相关标准规范之间的兼容性。

(2) 科学性

根据“基本遵循国际标准,立足实际业务需求”的适用性要求,本课题组在 GB/T 13191—2009、ISO 2789:2006《信息与文献—图书馆统计》的统计指标体系下,针对数字资源的统计特性进行了扩展研究,细化数字资源相关的统计指标和统计方法,形成一套完整的标准规范。本标准充分考虑未来数字资源的发展,能够方便地实现现实应用和未来扩展应用。

本标准由国家图书馆主持研发并负责维护,国家图书馆将密切跟踪和研究相关国际标准、国家标准和行业标准的发展动态,及时修订本标准。

2 范围

本标准构建了图书馆数字资源统计指标体系,对数字资源相关术语、数字资源体系、统计指标和统计方法做出了明确界定,适用于图书馆数字资源的馆藏、服务、设施和经费统计。

本标准基于一般图书馆数字资源统计的业务需求制定,因此,根据本标准所采集的全部统计数据并非适用于所有类型图书馆,也并非适用于所有目的的统计。但是,本标准确立的统计指标体系涵盖了数字资源的基本属性,具有较强的适用性,可将其应用领域扩展至图书馆以外的其他信息服务机构和信息资源管理领域,使用者可根据需要直接应用本标准或进行适当修订后应用。直接应用本标准的单位,可以根据统计目的和实际需求,选择或扩展部分指标进行统计。如果需对本标准进行修订,建议修订时遵循本标准确立的基本统计指标体系,在深度和广度上进行适当扩展。

本标准不涉及数字资源评估的指标体系和方法,但为数字资源的绩效评估提供部分基础数据。根据本标准收集的统计数据,不仅能够满足图书馆数字资源日常管理需求,为数字资源比较和分析提供基础数据,同时,还能够基本满足国际主流标准中数字资源评估相关指标的数据需求。

3 规范性引用文件

下列文件中的术语通过本标准的引用而成为本标准的术语^①。凡是注明日期的引用文件,其随后所有的修改(不包括勘误的内容)或修订版均不适用于本标准,但鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是未注明日期的引用文件,其最新版本(包括任何增补文件)适用于本标准。

GB/T 13191—2009/ISO 2789:2006 信息与文献—图书馆统计

GB/T 4894—2009/ISO 5127:2001 信息与文献—术语

GB/T 18391.3—2009/ISO/IEC 11179—3:2003(E) 信息技术元数据注册系统(MDR)第3部分

ISO 14721:2003(E) Space data and information transfer systems—Open archival information system—Reference model

Release 3 of the COUNTER Code of Practice for Journals and Databases:2008

4 术语及定义

下列术语和定义适用于本标准。

4.1 馆藏

4.1.1 数字资源(digital resources)

^① 本标准中提及但未在“术语及定义”章节加以解释的术语,参照 GB/T 4894—2009/ISO 5127:2001 和 GB/T 13191—2009/ISO 2789:2006 规定的定义。

经过选择、组织和加工处理,以数字格式存在的各种媒介信息。

4.1.2 数字馆藏(digital collection)

图书馆馆藏中所有的数字资源,包括图书馆本地拥有的和获得一定期限使用权的数据库和数字文献。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.22。

注2:包括购买、许可授权、受缴、捐赠、交换、数字化、网络信息采集等方式获得使用权或保存权的数字资源,未签订购买或许可授权合同的试用数字资源不包括在内。

注3:本馆承建的联机目录作为独立的数据库属于数字馆藏。

注4:不包括对网络资源的链接,因为图书馆没有通过合法的协议(例如法定呈缴权)、许可或者其他合约或合作协议来确保其使用权。经图书馆编目、整合并纳入其数据库或数字文献中的免费网络资源单独统计,其他免费网络资源不计入数字馆藏。

注5:开放获取的资源被视为免费网络资源(见 4.1.18)。

4.1.3 新增(addition)

统计周期内新增加的数字馆藏。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.3。

注2:可通过购买、许可授权、数字化、捐赠或者交换等方式获得。

4.1.4 剔除(withdraw)

统计周期内注销的数字馆藏。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.40。

注2:注销可通过删除数据、取消许可授权、剔除数字资源的物理载体等方式实现。

4.1.5 描述性记录(descriptive record)

由计算机处理的书目或其他标准格式的单个记录,用于标明、描述任何物理形式或某一内容单元中的文献。

[GB/T 13191—2009/ISO 2789:2006,定义 3.2.11]。

注1:描述性记录的集合通常以数据库的形式发布。

注2:记录包括的元素有:题名、作者、主题、摘要、原始日期等。

4.1.6 内容单元(content unit)

计算机处理的已出版作品的唯一可识别文本、图片或视听单元,这些已出版的作品可能是其他已出版作品的全文或者摘要形式。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.9。

注2:如果数字资源专门以描述性记录为主要内容,则其内容单元为相应的描述性记录。除此以外,描述性记录一般不作为内容单元进行统计。

4.1.7 元数据(metadata)

定义或描述其他数据的数据。

[GB/T 18391.3—2009/ISO/IEC 11179—3:2003(E),定义 3.2.18]。

4.1.8 对象数据(object data)

各种类型文献的数字化文件,如:文本文件、图像文件、音频文件、视频文件等。

4.1.9 数据库(database)

数字资源的描述性记录或者内容单元的集合(包括文本、图片、音频、视频、工具软件等),并带有统一的用户界面及检索、处理数据的软件。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.10。

注2:内容单元或记录通常根据某一特定目的收集并与某一特定主题相关。

4.1.10 全文数据库(full-text database)

包含文本型一次文献的完整或基本部分的数据库,其内容单元为非数值/事实形式的一次文本文献单元,如图书、期刊、报纸、档案等。

注:改写 GB/T 4894—2009/ISO 5127:2001,定义 4.1.1.4.12。

4.1.11 文摘/索引数据库(abstract and indexing database)

包含文摘、索引、目录、题录等二次文献的数据库,用于分析和揭示关于某一学科或主题的参考文献。主要由全文组成的数据库不包括在内。

注:改写 GB/T 4894—2009/ISO 5127:2001,定义 4.1.1.4.13。

4.1.12 数值/事实数据库(numeric/factual database)

包含描述性、事实性信息或者数值数据的数据库,能够直接提供可用的事实和数据,如工具书、表谱、图录、科学数据、统计资料、网络导航等。

4.1.13 多媒体数据库(multimedia database)

包含由文本、图片、音频、视频等组合而成的文献,并提供交互利用的数据库。

注:以移动图像或声音为主要特征、包含少量全文和其他内容的数据库,属于多媒体数据库。

4.1.14 工具型数据库(reference database)

不仅提供数字资源内容,而且提供交互式 and 个性化的技术支持,供检索、存储、分析或管理用户个体相关研究信息的数据库。

4.1.15 复合型数据库(hybrid database)

以文本、图片为主,包含多种类型内容单元,且各种内容单元权重均等或无法明确分割的数据库。

4.1.16 数字文献(digital document)

由图书馆购买、数字化或通过其他途径获得的,以数字形式存在的带有特定内容的信息单元。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.12。

注2:不包括数据库。

注3:不包括整合在数据库中的数字资源。

注4:一份数字文献可以构成一个或者多个文件。

注5:一份数字文献包含一个或多个内容单元。

4.1.17 源文献(original document)

经过扫描、转换和录入等方式数字化加工前的原始文献或原生数字文献。

4.1.18 免费网络资源(free internet resources)

没有限制获取的网络资源。

[GB/T 13191—2009/ISO 2789:2006,定义 3.2.18]。

4.1.19 数字馆藏量(stock)

本地拥有的和已经获得使用权、至少有一定使用期的某一特定类型数字资源的量。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.37。

注2:按统计周期结束的时间进行统计。

注3:可通过数字馆藏的数量体现,也可通过数字馆藏所占用物理存储空间的容量体现。

4.1.20 数字馆藏发布(published)

通过特定的渠道将数字馆藏提供给用户予以利用。

注1:发布者可以是图书馆本身,也可以是资源提供商。

注2:发布方式包括:局域网、互联网、数字电视、卫星通信、移动通讯、专用网络、光盘等。

4.1.21 许可(licence)

在指定条件下对知识作品的利用所作的授权。

[GB/T 4894—2009/ISO 5127:2001,定义 4.7.2.1.10]。

4.1.22 使用权(access rights)

获得或利用图书馆馆藏的权利。

注1:改写 GB/T 13191—2009/ISO 2789:2006,定义 3.2.2。

注2:对数字馆藏而言,这意味着图书馆需要依据法律、许可或者其他契约或合作协议,确保其用户永久或暂时的获取权利。

4.1.23 长期保存(long term preservation)

以适当的、能够被独立理解的格式进行长期保存信息的行为。

注1:改写 ISO 14721:2003(E),定义 Long Term Preservation。

注2:长期是指存储于知识库中的信息其保存期限足以应付技术变化(包括支持新媒体和数据格式)的影响和用户群体变化的影响。

注3:保存期限可无限延长。

4.2 服务

4.2.1 注册用户(registered user)

为了在图书馆内或图书馆以外使用图书馆的馆藏或服务而在该图书馆注册的个人或组织。

[GB/T 13191—2009/ISO 2789:2006,定义 3.3.20]。

注:用户可根据要求进行注册或者在加入某机构时自动注册。

4.2.2 当前用户(active user)

统计周期内访问过或者使用过图书馆设施或接受过服务的注册用户。

[GB/T 13191—2009/ISO 2789:2006,定义 3.3.2]。

注:可包括在馆内外利用数字馆藏服务。

4.2.3 访问(session)

一次成功的对在线服务的请求。它是用户行为的一个循环,一般从用户建立与服务或数据库的连接开始,到终止行为发生时结束。终止行为可以是显式的(通过退出或注销离开服务)或隐式的(由于用户不反应而导致的超时)。

[COUNTER:2008,定义3.1.4.2]。

4.2.4 被拒访问(rejected session)

由于超过并发用户数限制而导致对数据库或者联机目录的失败请求。

[GB/T 13191—2009/ISO 2789:2006,定义3.3.21]。

注:由于密码错误或网络故障导致的失败请求不包括在内。

4.2.5 检索(search)

特定的求知性查询,通常是指对数据库或者联机目录的服务器提交一个检索式。

[GB/T 13191—2009/ISO 2789:2006,定义3.3.24]。

注1:每一次检索请求提交给服务器时就会被记录下来算作一次检索。

注2:当场重复的反复检索、双击或者其他证据表明是用户无意识的检索行为的,不包括在内。

4.2.6 下载(download)

成功请求了描述性记录或内容单元,例如显示、打印、保存或发送电子邮件。

[GB/T 13191—2009/ISO 2789:2006,定义3.3.4]。

注:对网络服务器日志的成功请求是获得那些特定返回代码,遵循国家超级计算应用中心(NCSA, National Center for Supercomputing Applications)的定义。

4.2.7 远程使用(remote use)

由图书馆在网络上提供、用户从图书馆的局域网外通过网络对服务的利用。

注:改写 GB/T 4894—2009/ISO 5127:2001,定义4.5.4.2。

4.3 设施

4.3.1 终端(terminal)

终端是指由键盘和监视器或屏幕组成,用于发送和接收信息或指令的输入输出装置。

[GB/T 4894—2009/ISO 5127:2001,定义4.4.4.5]。

4.3.2 工作站(workstation)

独自工作或联网的计算机或不具有智能的终端。

[GB/T 4894—2009/ISO 5127:2001,定义4.4.4.4]。

4.4 经费

4.4.1 常规经费(regular funds)

用于维持图书馆日常业务运行的经常性资金。

4.4.2 专项经费(special funds)

用于图书馆开展特定项目的非经常性资金。

5 数字资源统计指标体系

5.1 指标体系构建原则

5.1.1 规范性原则

本标准规定了通用、规范的统计指标,所构建的指标体系符合国家数字图书馆工程其他子项目成果的相关要求,在一定条件下,可实现与国际国内相关标准规范指标体系之间的相互转换。

5.1.2 适用性原则

本标准符合国家数字图书馆工程设计需求,指标体系具有很强的可操作性和实践性,能够满足图书馆数字资源日常统计管理需求,并为数字资源评估提供部分基础数据。

5.1.3 扩展性原则

本标准构建的指标体系具有可扩展性,属于核心指标的一级指标作为必备统计指标,二级指标及扩展指标可选择性统计,在此基础上,使用者可对指标体系进行横向和纵向的适当扩展,以适应不断变化的数字资源环境和深入统计需求。扩展时应遵循以下原则:

(1) 向上兼容原则

对指标体系进行扩展时,增加指标的涵义不能超过其上级指标的涵义范围,不能与原有指标的涵义有任何重复。

(2) 一致性原则

如果直接增加其他标准规范的指标,应选择与本标准原有指标范围不重复且涵义不矛盾的指标,并说明指标来源,使用时严格遵守其原有涵义和范围,如果对其涵义和范围进行改写,应在标准中予以说明。

5.2 数字馆藏统计指标

图书馆数字馆藏包括数据库和数字文献,本标准分别规定了馆藏数据库量和馆藏数字文献量统计指标。由于图书馆的数字馆藏一般以数据库形式面向用户发布并提供服务,因此,本标准仅规定了数据库发布量统计指标,数字文献发布量统计指标可根据需要参照数据库发布量统计指标进行设定。

数字馆藏的数据存储格式多样,应用目的各不相同,除统计数字馆藏的容量和发布容量以外,数字馆藏数据的实际占用存储容量同样具有重要的统计意义。考虑到指标体系的系统性和完整性要求,本标准将实际占用存储容量指标纳入数字馆藏量统计指标体系中予以呈现。