

ICS 35.040
A 24



中华人民共和国国家标准

GB/T 20530—2006

文献档案资料数字化工作导则

Guide for the digitize processing of document and archives

2006-10-09 发布

2007-03-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布

目 次

前言	I
引言	II
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 数字化工作标准化指导原则	3
5 数字化工作的组织、人员	3
6 数字化对象的确定原则	3
7 数字化项目的一般过程	4
8 数字化过程中适用技术的选择	5
9 数字化成果的存储、管理与使用要求	7
10 数据利用和检索报导体系	9
11 数字化成果的测试指标	11
附录 A(资料性附录) 现行有关法律、法规目录	12
附录 B(资料性附录) 现行分类法目录	13
参考文献	14

前 言

本标准为首次制定。

本标准的附录 A 和附录 B 为资料性附录。

本标准由中国标准化研究院提出并归口。

本标准起草单位：中国标准化研究院、国家档案局、西安大东国际数据股份有限公司。

本标准主要起草人：刘植婷、董静平、常建宏、房庆、史立武。

引 言

本标准是为规范我国的文献档案资料数字化工作,建立、健全我国统一的数字化文献档案资料信息检索报导体系,指导我国信息数字化市场的发展,开展国际目录情报交流,促进文献档案资料信息资源的开发,促进我国国民经济发展与科学技术研究而制定的。

本标准是开展和制定各种文献档案资料数字化工作及数字化产品标准的重要依据。



文献档案资料数字化工作导则

1 范围

本标准规定了文献档案资料数字化过程中涉及的标准与一般管理；数字化对象的确定原则；数字化工作的一般过程；数字化过程中适用技术的选择；数字化成果的管理；数据利用和检索报导体系；数字化成果的测试指标等。

本标准适用于各级政府机构、企事业单位以及其他社会组织 and 个人的文献档案资料数字化过程。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件，其随后所有的修改单（不包括勘误的内容）或修订版均不适用于本标准，然而，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。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 9415(所有部分) 信息处理 数据存储设备用磁盘盘片

GB/T 16971 信息技术 信息交换用 130 mm 可重写盒式光盘(GB/T 16971—1997, idt ISO/IEC 10089:1991)

GB/T 17825.7—1999 CAD 文件管理 签署规则

GB/T 17825.10—1999 CAD 文件管理 存储与维护

GJB/J 3865 光盘存储档案信息技术与管理规范

电子出版物管理规定

互联网出版管理暂行规定

中华人民共和国档案法

中华人民共和国著作权法

关于制作数字化制品的著作权规定(国家版权局 1999 年 12 月 9 日发布)

3 术语和定义

下列术语和定义适用于本标准。

3.1

文献 document

记录有知识的一切载体。(GB/T 3792.1—1983)

3.2

档案 archives

国家机构、社会组织和个人从事政治、经济、科学、文化等社会实践活动直接形成，具有保存和利用价值的文字、图表、声像等形态的历史记录。(GB/T 3792.5—1985)

3.3

资料 article

为工作、学习或研究的需要而收集的各种材料。

3.4

分类 classification

按照选定的属性(或特征)区分分类对象，将具有某种共同属性(或特征)的分类对象集合在一起的过程。(GB/T 10113—2003)

3.5

数字化工作 digitize processing

按科学的方法分门别类地以电子格式加工、处理、存储文献档案资料,并能对这些信息资源进行高效的插入、删除、修改、检索、提供访问接口和信息保护等操作的过程。

3.6

数据 data

对事实、概念或指令的一种形式化表示,适于人工或自动方式进行通信、解释或处理。(GB/T 18391)

3.7

电子文件 electronic records

指在数字设备及环境中生成,以数码形式存储于磁带、磁盘、光盘等载体,依赖计算机等数字设备阅读、处理,并可在通信网络上传送的文件。(GB/T 18894—2002,3.1)

3.8

电子档案 electronic archives

具有保存价值的已归档的电子文件及相应的支持软件产品和软、硬件说明。(GB/T 17678.1—1999)

3.9

电子出版物 electronic publication

以数字代码方式将图文声像等信息存储在磁、光、电介质上并可以复制发行的大众传播媒体。供通过计算机或者类似功能的设备阅读使用,以表达思想、普及知识和积累文化。(GB/T 17933—1999)

3.10

网络出版物 network publication

将数字信息以一定的编排方式存储在网络服务器上,通过计算机和类似功能的联网设备调阅使用,以表达思想、普及知识和传播文化的大众传播媒体。(GB/T 17933—1999)

3.11

多媒体 multimedia

综合表现音频、视频、图形、图像、动画和文本的信息组合。(GB/T 17933—1999)

3.12

数字签名 digital signature

附加到数据单元上的或是对数据单元作密码变换后的数据,以使数据单元的接收方确认数据单元的来源和数据单元的完整性,并保护数据,防止被人(例如接收方)伪造。(GB/T 18811—2002)

3.13

密钥 key

控制加密与解密操作的一序列符号。

3.14

数据审计 data audit

基于对国际互联网(Internet)和企业内部互联网(Intranet)以及远程网(Telenet)之上建立的数据传输和信息发布活动中,进行的版权审计、数据安全审计、数据质量审计和效益审计(含数据过滤、安全屏蔽、漏洞扫描)等。

3.15

社会组织 community

以共同的物质生产活动为基础而相互联系的人们,按一定的目的和系统组织起来的团体。

3.16

报导 publishing report

是指对数字化成果及数字化检索体系向外界公开发表的过程或行为。

3.17

逻辑卷 logic volume

是一个区别于“物理卷”的概念,是指一组内容相互关联的按照一定规律组合起来的数字化成果的存储体。

4 数字化工作标准化指导原则**4.1 优先采用国家标准**

在数字化工作中,凡是已有相应国家标准的,应优先采用国家标准。

4.2 积极采用国际标准

在制修订数字化工作相关的国家标准过程中,应积极参照和采用 ISO/IEC 等先进的国际标准;在采标过程中,当国家标准尚未制定时,可参照和采用相应的国际标准。

4.3 数据兼容与共享**4.3.1 数据兼容**

在兼容未来发展技术的基础上,遵循普遍性、权威性、合理性等原则建立的行业标准或企业内部标准,可过渡性的保留使用。但其信息发布和数据传输部分的设计,应采用模块化堆叠设计,以保证在国家标准或国际标准颁布实施时,信息发布和数据传输可根据新的标准体系方便地升级。

4.3.2 数据共享

为了使采用新标准对原系统的影响最小化,减少投入风险,凡采用内部标准设计的系统,均须考虑建立在二次检索基础上的数据共享接口设计。并至少为数据的重复使用和管理建立联机 and 脱机使用的两种模式。

5 数字化工作的组织、人员**5.1 数字化工作组织认定及标准审查**

5.1.1 组织认定:从事数字化工作的组织机构应具有数字化工作能力。

5.1.2 实施技术环境标准审查:应具备有数字化工作专业知识和能力的人员、技术和设备,能熟知数字化工作的内涵、规则及国际、国家的有关标准,所做的数字化工作能与数字化工作目标相一致。

5.2 数字化工作的人员

5.2.1 数字化工作项目负责人应当具有较高的调查研究水平和良好的组织领导能力,全面掌握与任务有关的标准、方法和技术要求,熟练运用与数字化工作有关的软、硬件相关知识和技术。

5.2.2 数字化工作各岗位人员应掌握一定的数字化工作技术和基础知识,熟悉与本职工作有关的标准、方法和技术,胜任数字化工作相关岗位的任务。

6 数字化对象的确定原则**6.1 合法性确定原则**

只有通过正规渠道获取的(如档案、图书、资料部门保存的文献档案资料,政府机关与企事业单位的现行文件,证券与金融部门的票据,社会组织和个人保存的文献档案资料,正式出版物等),并且一般不会引起版权争议的文献档案资料才可以作为数字化的对象。

6.2 实用性认定原则

国家新闻出版总署、国家版权局、文化部、国家档案局等及经国家授权的组织机构认定的文献档案资料可以作为数字化的对象。

数据的产生具有稳定的过程;数据的公布具有稳定的渠道;连续数据的内容具有可比性;这样的文献资料已经为社会公众所接受,在社会上已经具有一定的权威性,也可以作为数字化的对象,例如:名人的手稿;能说明某一历史事实的地图、照片、录音、录像等。

任何进行数字化的文献档案资料,应符合《电子出版物管理规定》、《互联网出版管理暂行规定》等有关的规定。

6.3 完整性原则

应当选取内容、结构和文体上完整的文献档案资料进行数字化。

7 数字化项目的一般过程

一个完整的数字化项目一般应经过如下过程:

7.1 需求调研

编制项目的市场调研报告,参与调研人员应在报告上署名。

7.2 数字化工作计划的编制

计划编制应当符合本标准的有关规定。计划内容应当至少包含以下内容:

- 技术设计;
- 人员组织及分工;
- 时间安排与项目进度;
- 原始资料、文献、档案数字化过程中,出入库的安全与保密管理措施;
- 经费预算;
- 招投标等项内容。

7.3 法律事务

文献档案的数字化应在保证遵守现有法律、不侵犯他人合法权益的前提下进行。要完成以下(但不限于)工作:

- 明确得到文献档案权利人的著作权等权利的授权或许可;
- 文献档案权利人的权利状况复杂的,数字化行为人可进行著作权等知识产权的分析与评估,以保证数字化的合法性;
- 知识产权等法律的分析与评估依据中华人民共和国现有法律进行;
- 数字化行为人承担对文献档案进行数字化的法律责任。

7.4 方案设计

方案设计应给出项目的蓝图,描述其制作流程,各种媒体的组合方式,表现的层次,提供的交互功能等。方案设计阶段重点是给出项目的总体结构。方案细化是在方案总体设计的基础上对项目的细节作出规划。

7.5 技术工艺准备

技术工艺准备阶段的主要工作是:

- 依据项目任务书(或合同)及项目计划采集相关数据;
- 对数字化对象进行分析、分类;
- 编制作业的指导性技术文件;
- 准备数字化过程中需要的基本软硬件系统,这些软硬件应当符合本标准的相关规定,经试用合格后方可投入数据生产与利用环境。

7.6 制作与编辑

项目作业人员应根据作业的指导性技术文件,按照项目进度计划,贯彻项目的质量控制要求,并对数据处理过程中产生的中间以及结果数据,实行数据移交过程签收制。

7.7 成品的审校与验收

7.7.1 项目承担单位应组织专人,对项目作业人员提交的成品进行全面、系统的测试和验收。

7.7.2 成品测试需要的文档(含数字化产品本身)

由专业技术部门出具测试报告,并组织专家验收,出具验收报告。

7.7.3 成品的标准化审查

在对成品进行审查验收的同时应安排对成品进行标准化审查。经标准化审查的数字化成品从格式、存储、分类与代码、传送安全,到所采用的软件、平台等均应符合本标准及其他相关标准的规定。审查结束,应形成标准化审查报告。标准化审查报告是该项目成果报告的重要组成部分。

7.8 整理项目的作业文档

完整的作业归档文件至少应包括如下九大部分:

- 项目设计要求书;
- 数字化工程标书;
- 数字化工程项目投标书(含所有投标单位);
- 数字化工程施工合同;
- 数字化工程作业登记表;
- 数字化工程勘误表;
- 数字化工程项目测试报告;
- 数字化工程项目验收报告;
- 标准化审查报告等。

7.9 项目成果的鉴定和验收、归档

7.9.1 成果鉴定

7.9.1.1 鉴定内容主要包括:成果报告;成果图表;资料统计表;经费决算。

7.9.1.2 鉴定依据:任务书或合同书、数字化工作计划以及本标准规定的各项技术指标。

7.9.1.3 鉴定要求:按照有关规定进行鉴定。鉴定通过后要填写科学技术成果鉴定证书,鉴定未获通过,应限期补充修改,再次报请重新鉴定。未完成归档的项目成果,不能鉴定或验收。

7.9.1.4 鉴定时间:应在归档完成之后及时进行。

7.9.2 成果验收

7.9.2.1 凡不需要进行鉴定的成果,在归档完成后即可进行成果验收,验收内容及依据同 7.9.1.1 和 7.9.1.2。

7.9.2.2 验收办法:由项目任务下达单位或委托单位派人组织验收,形成由验收人签字和验收单位盖章的书面验收结论。与验收依据有明显差距的成果不予验收。

7.9.3 成果归档

7.9.3.1 归档内容主要包括:任务书或合同书及其相关文件;项目计划;电子数据;采用的软件系统、压缩格式、数据库定义;数字化成果的数量和质量报告;成果鉴定证书和验收结论(待鉴定、验收后归档);经费决算报告。

7.9.3.2 归档要求:按照档案法规的有关规定,将归档材料系统整理,经项目负责人审查签字,由档案管理部门验收后保存。档案材料应内容齐全,编排合理,装订整齐;按项目文件保密规定合理划分密级,妥善保管。

7.9.3.3 归档时间:应在数字化项目成果报告完成后二个月内。

8 数字化过程中适用技术的选择

8.1 分类与标引技术

8.1.1 分类的基本原则

数字化过程中对信息的分类应当遵循科学性、系统性、简明性、可扩展性、兼容性和综合实用性的原则。

8.1.2 采用的分类标准

对于数字化信息的分类在遵循上述原则的基础上,应根据信息内容情况参照国家和行业有关标准

和规范。(现行分类法目录参见附录 B)

8.1.3 标引

标引就是对文献的内容和形式特征进行分析,用科学、规范的人工语言(信息检索语言)来代替分析结果中的自然语言的过程。最常用的标引是分类标引和主题标引两种,前者依据分析结果对文献赋予分类号,后者依据分析结果对文献赋予主题词。

8.1.4 标引技术

对要进行数字化的文献档案资料可以采用人工标引技术,也可以采用如下自动标引技术:无标引技术、自动抽词标引技术、人机结合抽词标引技术、自动赋词标引技术和自由标引技术。

8.2 录入与扫描技术

下述几种录入与扫描技术都各有利弊,数字化的组织者应当根据项目要求进行准确选择。

8.2.1 规则录入技术

按照数字化项目要求,编制一套符合要求的录入规则,数字化工作人员按照规则进行录入,然后再由程序编制人员编制相关程序,将规则录入结果交由程序转化为项目要求的数字化成果。

8.2.2 格式录入技术

按照数字化项目要求,由程序编制人员编制相关程序,制成既定格式,最后由数字化工作人员按照既定格式直接录入成项目要求的数字化成果。

8.2.3 扫描技术

如果采用扫描技术,则应根据数字化项目要求,对扫描的分辨率、扫描格式和扫描色彩位深等相关参数作出严格的规定。

8.3 数据合成技术

对文本、图片等文献资料的数据合成,一般可以采用以下几种技术:

8.3.1 纯图像技术

技术简单,只需将数字化对象扫描成图像,经图像压缩后存储在介质上即可,但由于不能提供方便的检索手段,一般只能当作文献的数字化拷贝。

8.3.2 目录文本、正文图像技术

目录全部用文本方式存储,正文部分用压缩的图像存储,并在目录与正文之间建立相应的索引关系,用户可以通过这样的索引关系,根据目录找到正文的图像。

8.3.3 全文本技术

文献中所有文字全部采用文本方式存储,文献中的图像也用一定的代码与之建立固定的联系,这样可以实现全文检索。

8.3.4 全文索引技术

是文本方式和图像方式的混合体,文本和图像之间建立索引关系,用户既可以进行全文检索,又可以很方便地在文本和图像之间进行切换。

8.4 数据加工质量控制技术

8.4.1 数据质量精度的确定

8.4.1.1 优质:误码率低于 0.25/10 000。

8.4.1.2 良好:误码率超过 0.25/10 000,未超过 0.5/10 000 的。

8.4.1.3 合格:误码率超过 0.5/10 000,未超过 1/10 000 的。

8.4.1.4 不合格:误码率超过 1/10 000。

8.4.1.5 误码率的计算方法可参照《图书质量管理规定》的有关条款执行。

8.4.2 识别字典技术

对扫描生成的数据而言,识别字典技术是保证数据质量的重要技术手段。识别字典是厂商对每个汉字的多个扫描或手写输入样本进行特征抽取,并对各个结果进行一定的甄选,以模板的形式存在于系

统中。单个扫描或手写样本特征的抽取算法和过程与用户使用时的处理过程完全一致,只是增加了对同一汉字多个样本的综合,最终形成模板并存入识别字典。只有这样,识别字典才能最大可能地适应不同的扫描结果或不同人的不同写法。

8.4.3 校对技术

8.4.3.1 人工校对技术:只需录入一遍,但需反复校对多遍,效率较低而且难以保证较高的数据质量。

8.4.3.2 程序自动校对技术:依据概率论的数学方法模型:两个不同的人在同一个地方,犯同一错误的概率几乎为零的原理,建立的自动校对程序,校对效率高、数据误码可控制在标准范围内。

9 数字化成果的存储、管理与使用要求

9.1 数字化成果的存储和保管技术

9.1.1 数字化成果存储介质的要求

9.1.1.1 存储介质应当是标准化程度高、性能稳定、有发展前途、容易检测、显示还原方便的不可擦写的光盘载体。但也可采用其他介质作为容错备份手段。

9.1.1.2 存储介质必须进行杀毒处理,不得带有任何计算机病毒。存储介质在保存前,所规定的签署及日期等应填写齐全。按照需要,在数字化成果的存储介质中可存放运行环境与系统等有关的软件程序。

9.1.2 数字化成果的管理要求

9.1.2.1 作为数字化成果的存储介质应作写保护处理,将其置于只读状态。

9.1.2.2 建立数字化成果管理的记录系统。应为每一份数字化成果建立必要的记录,记载数字化成果的形成、管理和使用情况,用这些记录来证实数字化成果内容的真实性。

9.1.2.3 数字化成果部门应及时对数字化成果进行登记、建账,同时建立机读目录。应对数字化成果定期检查,按照数字化成果保管环境的要求,严格执行管理制度。

9.1.3 数字化成果的保管环境

9.1.3.1 存储数字化成果的光盘不得擦洗、划刻、触摸盘片裸露处,不得弯曲、挤压、摔打盘片、防止盘片沾染灰尘和污垢。

9.1.3.2 存储数字化成果的光盘应避免阳光直接照射,并远离热源、酸碱等有害气体和磁场。执行 GB/T 17825.10 的规定。

9.1.3.3 存储数字化成果的光盘,其保管场所的环境温度为 14℃~24℃,相对湿度 45%~60%。

9.1.4 数字化成果的检测与维护

9.1.4.1 每年定期检测一次。以一个逻辑卷为单位,采用等距或随机抽样的方式进行,样品数量以不少于 10%为宜。首先进行外观检查,确认载体表面是否有物理损坏或变形,外表涂层是否清洁及有无霉斑出现等。然后进行逻辑检测,采用专用或自行编制检测软件对载体上的信息进行读写校验。通过检测发现有出错的载体,须进行有效的修正或更新。

9.1.4.2 原则上每四年拷贝一次,且原载体至少继续保留四年。

9.1.4.3 对于数字化成果的检测与维护,应当建立相应的维护管理档案,对数字化成果的检测、维护、拷贝等操作过程进行记录,避免人为的误操作或不必要的重复劳动。

9.1.4.4 对数字化成果进行有效性、安全性检查时,如发现光盘损坏或出现问题,应及时拷贝。如软、硬件平台发生改变,则应及时转换。检查、拷贝、转换等情况要登记。检查登记表格式及填写方法见 GB/T 17678.1—1999 附录 B 中表 B4。

9.1.4.5 存储数字化成果的光盘应放入专用防火防爆盒中。存储数字化成果的光盘应制作三套, A 套封存, B 套提供利用, C 套异地保存。

9.2 数字化成果的数据更新与备份审查

9.2.1 数字化成果如果要进行变更,应当经过数字化成果所有人(或所有单位领导人)的签字同意,并

且办理相应的变更手续,所有变更都应该记录在案。同时应保持 A、B、C 三套内容一致,须及时填写更改登记表,登记表见 GB/T 17678.1—1999 附录 B 中的表 B3。

9.2.2 数字化成果的备份,应当经数字化成果所有单位的领导签字同意,并将备份情况记录在案。

9.3 电子档案成品数据的出借

9.3.1 已经归档的电子档案数据原载体不外借,只能以拷贝或网上传输的形式提供利用,并应登记。使用者不得私自复制、拷贝、修改。

9.3.2 电子档案可在归档与管理系统的终端上查阅,但查阅人员只能查阅本人权限之内的电子档案,如查阅超越权限的电子档案,需按本单位有关规定执行。

9.3.3 使用权限的审核应由提供利用的决策者执行。要根据各种人员级别、层次进行使用权限的认定,并依此向利用系统注册登录。在利用中,由系统自动判定当前使用者身份的合法性及其所使用功能的范围,并由系统自动对其使用各种功能操作的路径进行跟踪与记录。对涉及使用未经授权的功能,应能拒绝响应并给予警告提示。

9.3.4 对拷贝的提供与回收应当进行严格管理。应当避免把载体上存储的数据全部拷贝,并通过技术手段防止所提供拷贝的再复制。同时对那些提供利用的拷贝须进行回收。应有完善的提供拷贝手续,提供者和利用者双方应对提供拷贝的内容进行确认,并对使用载体的类型、数量、使用时间、最后回收期限及双方责任人等情况进行登记。对回收来的拷贝,在确认未被利用者替换后,应作信息内容的消除处理。

9.3.5 批准为“使用权限保护”的电子档案项目,要在台账、目录、载体上标识。档案部门应严格执行电子档案的“使用权限保护”和密钥管理规定。

9.3.6 采用通信传输或直接利用等方式时,对有密级的信息内容要作加密处理,对所使用的密钥进行定期或不定期的更换。

9.3.7 系统应对利用的全过程进行有效的跟踪监控,并自动进行相关记录,作为对利用工作查证的依据。利用的系统应有较强容错能力,避免由于误操作带来不可挽回的损失。

9.4 保障电子档案真实性的基本要求

为保证电子档案具有与相应纸质档案相同的法律与鉴证作用,电子文件从产生、运行到文件归档直至电子档案的整个管理过程都应处于可控制的环境中,排除可以修改它的可能因素。

9.4.1 确保电子档案及其副本的真实性

9.4.1.1 具有法律鉴证作用的电子档案应有防写措施。如果要拷贝,则应使用不可逆式记录介质,以保持电子档案文件的原始性和真实性。

9.4.1.2 纸质档案在数字化过程中一般采用扫描方式,对原文字、印章、签署、落款等形式和颜色不能作任何改动。

9.4.1.3 对纸质档案进行数字化的过程应做详细登记。登记内容包括:日期、纸质档案的档号、电子档案的编号、表格形式、电子存储方式、压缩格式、数量、操作人员签名。

9.4.1.4 数字化操作人员应对电子档案与相应纸质档案的一致性负法律责任。当需要提供电子档案副本时,电子档案副本的提供部门应对此副本与原件的一致性负责。

9.4.2 电子文件的签署规定

9.4.2.1 利用数字化手段直接形成的电子文件按照有关国家标准和规定进行签署。

9.4.2.2 电子文件的签署必须完整。签署应采取分级负责的原则,签署者应承担相应的签署责任。签署时,一般一人只能签署一项,而且字迹要清晰。签署者姓名不能省略,日期应当年、月、日完整。

9.4.2.3 电子文件的签署采用证书式数字签名或手写式数字签名。采用证书式数字签名者需要向专门的技术管理机构注册登记。在没有光笔或数字化仪的情况下,应在签署单的签名栏中设置口令进行授权签署。

10 数据利用和检索报导体系

10.1 数字化出版技术基本规定

10.1.1 出版方法

数字化成果按照国家有关规定经过一定的审批手续后可以出版并向公众发行。出版发行方法可以选择:数据库在线出版、光盘发布、电子邮递、超级连接等方法。

10.1.2 防伪技术

数字化成果在出版时可采用一定的防伪技术,以起到版权保护、秘密通信、数据文件的真伪鉴别和产品标志等作用。可以根据数字化出版物的不同情况,采用:信息伪装、数字水印、信息隐藏、数字指纹或标识和可见水印等不同技术。

10.1.3 软件和著作权

数字化成品是一种智力活动的成果,按照《中华人民共和国著作权法》,数字化的成果应当依法获得知识产权的保护,享有版权(著作权)、工业产权和邻接权(但对原文献档案资料的信息进行加工或直接转化而行使著作权时不得侵犯原作品的著作权)。

数字化成果的知识产权受法律保护。

10.2 数据审计和网络共享标准

10.2.1 数据审计

数据审计包括:版权审计、数据安全审计、数据质量审计和效益审计。数据审计工作应当由国家指定机关进行,审计结果应有书面的审计报告。

10.2.1.1 版权审计

数字化成品在出版或发布前,应按照《中华人民共和国著作权法》和中华人民共和国版权局在1999年12月9日发布的《关于制作数字化制品的著作权规定》的有关条款,到国家指定的版权管理机构对该成品的著作权予以审计。审计时要提交相关数字化成品的著作权记录。

10.2.1.2 安全审计

数据内容的安全审计应根据国家《互联网出版管理暂行规定》和《电子出版物管理规定》以及有关行业特种要求进行。数据安全审计至少应当包括数据内容的安全审计和数据传输的安全审计。

数据的内容应符合上述两项规定中的有关条款。凡是数字化成品的内容不符合国家规定的,将被审计为数据内容不符合国家规定,将被确定为内容不安全。

数据传输安全审计的内容包括:数据在传输过程中可能出现的数据丢失、数据变形、数据泄密、数据失密等有关数据传输过程中的安全漏洞问题。

安全审计工作应由国家指定的专业机构进行,并提供安全审计报告。

10.2.1.3 质量审计

数字化成品的数据质量审计应当参照本标准中的有关条款,对数据内容质量、数据编辑加工质量和成品的封装设计质量进行审计。

质量审计划分为:优良、良好、合格或不合格四种审计结果。

质量审计工作应在国家指定的专业机构或具体相关资质技术委员会指导下进行。

10.2.1.4 效益审计

数字化成品的经济和社会效益审计应由具体相关资质的社会机构进行。

该项审计主要依据数字化成品项目的计划、合同、招投标书等有关项目文件,按照科学的计算方法计算项目的投入、产出以及所产生的社会效益来进行审计和评价。

10.2.2 安全漏洞扫描

在数据利用和检索报导体系中,应当设置相应的安全漏洞扫描程序。安全漏洞扫描程序应当具备

如下四点要求：

- 漏洞检测的完整性；
- 漏洞检测的精确性；
- 漏洞检测的合适范围；
- 漏洞检测功能的及时更新。

10.2.3 安全日志

鉴于数据利用和数据传输过程中的安全要求，因此在数据利用和检索报导体系中应当设置安全日志。该安全日志系统应能实时地监测网络上和用户系统中发生的与安全相关的事件，并将这些情况真实、详尽、完善地进行记录，在必要时能提供宝贵的数据，并具有防销毁和篡改的功能，同时该系统应提供完善的日志审计功能，可以从不同的角度对日志进行查询统计，并将结果以不同的方式显示给审计员。审计员可以随时查看被该系统记录和被禁止的访问控制信息，分析其中的问题，与安全管理员一起细化网络的访问控制策略，更好地保护数据利用和检索报导系统。

10.3 检索报导体系

10.3.1 检索要件的设置

根据数字化项目的要求，数字化成果在提供利用和报导过程中可以采用如下检索要件：

- 10.3.1.1 一级检索：根据用户输入的关键字，在所有数字化成果中检索到相关信息的过程。
- 10.3.1.2 二级检索：在一级检索结果的基础上，再次根据用户输入的关键字检索到相关信息的过程。
- 10.3.1.3 关联性检索：根据用户输入的关键字，在所有数字化成果中检索到相关信息后，还能进一步提供与此信息相关的信息的过程。

10.3.2 检索方法的设置

- 10.3.2.1 精确检索：根据用户输入的关键字，在所有数字化成果中检索到完全匹配信息的检索途径。
- 10.3.2.2 模糊检索：根据用户输入的关键字，在所有数字化成果中检索到包含相关信息的检索途径。
- 10.3.2.3 组合检索：根据用户从不同角度提出的多个检索要求（例如同时提出时间、内容、作者等要求）而检索到能同时满足上述检索要求的相关信息的途径。

10.3.3 检索报导层次设置

数字化成果的检索报导层次分为：标题报导、题录报导、主件报导、主件和附件完全报导四个层次。其中题录报导应当符合 GB/T 3792 的规定。

10.3.3.1 标题报导

标题报导应包括三大类基本标识要素：类别、主题（含引题、标题、副题）、时间。

10.3.3.2 题录报导

题录报导除包括 10.3.3.1 外，还应包括资料来源、作者、主题词、内容索引和报导制作单位五大要素。

10.3.3.3 主件报导

主件报导一般泛指报导的正文或主体内容部分。

10.3.3.4 主件和附件完全报导

除 10.3.3.3 外的所有可能的附件以及超级链接。

10.3.4 归并数据再发布的基本要求

- 10.3.4.1 数据版权没有争议。
- 10.3.4.2 发布的数据与归并编辑前的数据完全一致，无篡改，并可回溯。
- 10.3.4.3 数据结构完整，不能断章取义。
- 10.3.4.4 数据的发布是安全的。

11 数字化成果的测试指标

11.1 测试指标值

11.1.1 数据存储介质质量指标:要符合 GB/T 9415、GB/T 16971、GJB/J 3865 等标准中有关条款的质量指标。

11.1.2 数据兼容性质量指标:数字化成果的数据在利用和对外报导的过程中可能遇到各种各样的软件或硬件使用环境。质量好的数据应当有比较大的兼容性,并能在各种不同的环境中进行数据转换。

11.1.3 数据一致性指标:在利用和对外报导数字化成品数据过程中,对数据的操作应保持成品数据与源数据的完全一致性和完整性。

11.1.4 数据查词效率指标

11.1.4.1 查全率 $P = \text{正确分到类中的文章数} / \text{属于该类的文章总数} \times 100\%$,应当 $\geq 90\%$ 。

11.1.4.2 查准率 $R = \text{正确分到类中的文章数} / \text{分到该类的文章总数} \times 100\%$,应当 $\geq 95\%$ 。

11.1.4.3 查词速度 = 查词时间 / 10 万条信息,应当 $\leq 0.1 \text{ s}$ 。

11.1.5 数据封装质量指标

数字化的成品出版时,其数据封装质量指标可参照国家有关电子出版物的规定执行。

11.1.6 经济效益指标

数字化成果的经济效益指标,应当依据数字化项目的计划、合同、招投标书等有关项目文件,按照国家或国际通行惯例,采用科学的计算方法,计算项目的投入、产出以及所产生的经济效益来进行评价,也可从纯会计学原理和法则来评估。

11.2 测试过程

11.2.1 数据误码控制测试

遵照科学的抽样方法对数字化成品进行抽样测试,随机抽样率应不低于 5%。

11.2.2 系统功能测试

按照测试大纲要求,将分组数据加载到真实环境中,通过预设的各项系统功能对数字化成品进行现场测试。

11.2.3 安装环境测试

安装环境测试包括:用户系统不同配置的测试、软硬件资源使用环境自动检测、提示帮助信息等。

11.2.4 执行效率测试

对大量数据的检索查询、动画的声画时基配合、数据的装载时间、对鼠标击键的反应等。

11.2.5 相容性测试

跨平台的数字化产品,需要此项测试,以保证在 Windows、DOS、MAC、UNIX、LINUX 等主流操作系统下均可使用。

11.2.6 存储光盘片的测试

由于光盘和硬盘读取速度的差异,在把制作好的成品从硬盘转移到光盘上时,把集成好的成品写到光盘上后,应至少在三种不同的光驱上进行读出测试,以保证刻录过程的正确性。

存储光盘一般只能选择一次性写入光盘,如:CD-R、DVD-R。

11.2.7 内部人员/外部人员测试

测试应分别由数据产品的制作者和一般用户分别进行,汇集两者的意见,不忽略任何问题以得到全面的测试结果。测试后要对产品进行必要的修改,甚至修改总体设计。必须在理想的设计、有限的投入、投机的市场之间做出折衷。

11.2.8 商品化包装、手册、广告、售后服务测试

商品化包装包括光盘片表面的印刷设计、光盘盒中彩页的设计和印刷、用户手册的撰写及广告宣传、售后服务。

附 录 A

(资料性附录)

现行有关法律、法规目录

1. 《关于制作数字化制品的著作权规定》(国家版权局 1999 年 12 月 9 日发布)
2. 《中华人民共和国著作权法》(2001 年 10 月 27 日第九届全国人民代表大会常务委员会第二十四次会议通过修订)
3. 《中华人民共和国著作权法实施条例》(自 2002 年 9 月 15 日起施行)
4. 《中华人民共和国档案法》
5. 《电子出版物管理规定》
6. 《互联网出版管理暂行规定》(2002 年 8 月 1 日起施行)
7. 《互联网信息服务管理办法》(已经 2000 年 9 月 20 日国务院第 31 次常务会议通过)
8. 《互联网电子公告服务管理规定》
9. 《互联网站从事登载新闻业务管理暂行规定》
10. 《中华人民共和国计算机信息网络国际联网管理暂行规定》
11. 《实施国际著作权条约的规定》
12. 《计算机软件保护条例》



附 录 B
(资料性附录)
现行分类法目录

1. 《中国图书馆分类法》(简称《中图法》)
2. 《中国科学院图书馆图书分类法》(简称《科图法》)
3. 《中国人民大学图书分类法》(简称《人大法》)
4. 《中国档案分类法》
5. 《标准文献分类法》
6. 《国际标准分类法》(ICS)
7. 《杜威十进分类法》(DDC)
8. 《国际十进位分类法》(UDC)
9. 《国会图书馆分类法》(LCC)
10. 《国际专利分类法》(IPC)



参 考 文 献

- [1] GB/T 3792.1—1983 文献著录总则
 - [2] GB/T 3792.5—1985 档案著录规则
 - [3] GB/T 10113—2003 分类编码通用术语
 - [4] GB/T 17678.1—1999 CAD电子文件光盘存储、归档与档案管理要求 第一部分:电子文件归档与档案管理
 - [5] GB/T 17933—1999 电子出版物 术语
 - [6] GB/T 18391.4—2001 信息技术 数据元的规范与标准化 第4部分:数据定义的编写规则与指南
 - [7] GB/T 18811—2002 电子商务基本术语
 - [8] GB/T 18894—2002 电子文件归档与管理规范
 - [9] 图书质量管理规定
-



中 华 人 民 共 和 国
国 家 标 准
文 献 档 案 资 料 数 字 化 工 作 导 则

GB/T 20530—2006

*

中国标准出版社出版发行
北京西城区复兴门外三里河北街16号
邮政编码:100045

<http://www.spc.net.cn>

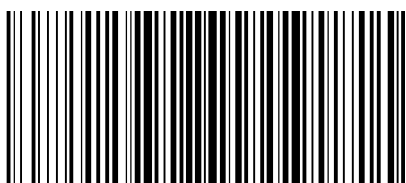
电话:(010)51299090、68522006

2007年3月第一版

*

书号:155066·1-28953

版权专有 侵权必究
举报电话:(010)68522006



GB/T 20530-2006