

中文图书 ECIP

与自动编目手册

陈源蒸 编著

北京图书馆出版社

中文图书 ECIP 与自动编目手册

陈源蒸 编著

北京图书馆出版社

图书在版编目(CIP)数据

中文图书 ECIP 与自动编目手册/陈源蒸编著. —北京:北京图书馆出版社,2003.3
ISBN 7-5013-2095-0

I. 中... II. 陈... III. 图书编目—自动化—手册 IV. G254.36-62

中国版本图书馆 CIP 数据核字(2003)第 009220 号

书名 中文图书 ECIP 与自动编目手册
著者 陈源蒸 编著

出版 北京图书馆出版社 (100031 北京市西城区文津街7号)

发行 (010)66126153 **传真** (010)66174391

E-mail Btsfxb@publicf.nlc.gov.cn

Website www.nlcpress.com

经销 新华书店

印刷 中特印刷服务部

开本 787×1092 毫米 1/16

印张 28 印张

版次 2003 年 3 月第 1 版 2003 年 3 月第 1 次印刷

字数 665(千字)

印数 1-5000

书号 ISBN 7-5013-2095-0/G·534

定价 60.00 元

目 录

第一部分 ECIP 概述	(1)
一、 CIP 回顾	(1)
二、 ECIP 产生的背景	(8)
三、 ECIP 的具体实施	(12)
第二部分 ECIP 机读格式	(24)
一、 总则	(24)
二、 工作单设定	(31)
三、 ECIP 使用字段、子字段一览表	(38)
四、 DC 元数据—MARC 子字段对照表	(43)
第三部分 细则	(47)
一、 扉页	(47)
二、 版本记录页	(82)
三、 附书名页	(96)
四、 插页	(122)
五、 目次页	(131)
六、 机读数据	(142)
七、 预置数据	(156)
八、 CIP 不用数据	(164)
第四部分 附录	(172)
一、 有关标准	(172)
二、 有关代码表	(196)
三、 标引规则	(198)
第五部分 图例	(249)
一、 几点说明	(249)
二、 图例名称及所述重点数据元素	(250)

第一部分 ECIP 概述

一、CIP 回顾

1. CIP 的发生与发展

图书在版编目数据(Cataloging in Publication Data,简称 CIP Data)是 20 世纪 70 年代初出现在图书版权页上的一个项目。“在版编目的目的是在有关文献被出版前,向出版商提供编目数据,以便使这些数据能够被印刷在该出版物中。通过这种办法,出版物和它的编目数据可以同时被图书馆、书商和其他需要这一数据的人们所利用。”^[1]

在版编目的出现,一方面反映了图书销售战的激烈,促使各出版商希望国家图书馆能尽快地将他们出版的图书编入目录以利扩大推销。另一方面也反映了西方出版商服务工作的深入,出版工业日益与图书馆业、情报业相结合,重视学术交流的需要。^[2]图书馆是出版业的重要市场,尤其是学术著作的主要供应对象,美国出版商为图书馆供货时,都要提供国会图书馆编制的目录卡片(现在是机读目录)。^[3]

此外,CIP 的发生与发展,与 IFLA(国际图书馆协会联合会)倡导 UBC(Universal Bibliographic Control,国际书目控制)计划有很大关系。UBC 的目的是“对世界各国发表的主要出版物以一种国际上可接受的方式,全面而迅速地作出基本目录。”^[4]建立这样一个交换系统,要解决三个关键问题:①书目数据源;②标准化;③组织。这是一个非常艰巨、难以实现的任务,因此,IFLA 将工作重点放在推动各国 NBC(国家书目控制)计划上,帮助各国建立国家书目体系、联合目录体系、专题目录体系、检索刊物体系。而 NBC 目标的实现需要文献信息生产、传递与收藏利用各个环节的协同行动,其书目信息的源头就是图书在版编目。

各国出版商编制图书在版编目数据的过程是:当图书的书名、主编或作、译者最后肯定,目录已完成,国际标准书号已申请下来后,出版商就得尽快填写一张由国家图书馆(美国为国会图书馆)提供的表格。将填好的表格并随附该书的全部校样或该书的书名页、目录、前言、内容简介、版权页等样张,送到国家图书馆。国家图书馆参照出版商填写的数据进行编目,将编好的 CIP 数据返回出版商据以印在图书的版权页上。国家(国会)图书馆同时将此记录在其 MARC 磁带上发布。

早在 1876 年就有人提出了图书附带编目数据的思想,1929—1938 年美国出版业曾试验从国会图书馆订购目录卡片随书配发,因时差太长,费用太高而告终。1958—1959 年国会图书馆进行书源编目试验,又因打乱正常出版秩序,出版商难以操作而中断。1963—1967 年澳大利亚国家图书馆与出版商合作进行“图书在版编目”试验,因各方意见不一而中止。1968 年美国国会图书馆重新提出这个问题,在国家人文科学基金会等机构的支持下,经过反复酝酿,于 1971 年 6 月 20 日正式运行“图书在版编目”,取得成功。受美国影响,澳大利亚在 1972 年 11 月又恢复试验,并转入正式运行。随后,德国、巴西、墨西哥等国在 1974 年,英国、加拿大等国在 1975 年,以及前苏联相继开展“图书在版编目”。

1982 年 8 月,国际图联书目控制办公室在渥太华召开了国际 CIP 会议,就各国 CIP 数据记录

格式不一,影响书目信息交流的问题进行了讨论,决定组成工作组起草国际 CIP 记录标准格式。工作组于 1984 年 3 月提交了初稿,经各方讨论修改,1985 年 6 月完成最终报告,提出国际 CIP 记录的推荐格式和计算机处理的“CIP 数据工作单”。由于 CIP 数据工作单考虑了不同语言的特点,图书印刷格式与机读格式两方面的需求,以及在不同国家出版商与图书馆等不同用户之间的交换,因此所含信息比 CIP 记录要多一些。

1987 年国际标准化组织因 CIP 的实施,决定修订 ISO 1085-75 图书书名页国际标准,并于当年 6 月批准公布了修订后的图书书名页国际标准 ISO 1086-1987。修订后的标准详细规定了图书书名页正面与背面的内容与格式,CIP 数据在图书书名页背面的位置,版本记录相应调整了数据元素,不再重复描写书名与责任者,使图书书名页的文字信息更为清晰,印刷格式更加规范。

由于美国国会图书馆已于 1969 年正式发行 MARC 磁带,因此,美国的 CIP 数据从一开始就应用计算机进行处理,并与 MARC 记录采用同样的机读格式,其机读数据也与 MARC 磁带(现在通过网络)同时向全社会发布,为书商及图书馆界广泛使用。其他国家做法大体相同,只是前苏联由中央出版物登记局承担 CIP 编制工作,并单独发行磁带。

各国 CIP 数据的不断完善,并与国家图书馆的 MARC 记录共同建立联机编目网络,形成效益极高的社会化书目事业,对推动世界范围内的信息资源共享发挥了重要作用。

2. CIP 在我国的实施

我国图书馆界最早是在西文图书编目中知道 CIP 的,利用西文图书上的 CIP 数据,大大提高了编目工作效率,从而联想到在中文图书上予以实施。1979 年开始有人向图书馆界全面介绍了图书在版编目的情况,引起极大反响,图书馆界对此展开了热烈的讨论,提出了种种设想。到 1985 年 5 月,在有关主管部门的推动下,出版界与图书馆界的专家坐到一起,共同商量在我国实现 CIP 的问题,自此出版界也开始对 CIP 进行研究,并有几个图书馆与出版社分别进行合作试验,在部分图书的版权页上加印了书目数据,迈开了推行图书在版编目的第一步。

1986 年 11 月,国家出版局与国家标准局联合召开专家会议,讨论在我国如何实施图书在版编目工作。会上通报了试验工作情况和各方面的希望,就标准局与出版局所提实施方案建议充分交换了意见,认为在我国开展图书在版编目的时机已经成熟,决定由有关部门选派人员组成工作组,制定我国实施图书在版编目工作计划,负责此项工作的组织协调、经费筹集以及有关标准和规程的审查及贯彻执行。会后有关单位分别做了大量工作,并对国外 CIP 实施情况进行了考察。

1987 年 7 月 6 日新闻出版署约请国家标准局等有关单位负责人,举行图书在版编目领导小组成立会议。会议认为首先要制定相关标准,决定由新闻出版署标准室、北京图书馆、中国科学院图书馆、北京大学图书馆、人民出版社、机械工业出版社、新华书店总店和信息分类编码研究所等单位组织起草小组,负责起草《图书在版编目数据》、《图书书名页》两项国家标准。至当年底产生了两个标准的征求意见稿,经出版界及图书馆界广泛讨论修改,至 1989 年 1 月起草小组完成了两个标准的“送审稿”。同年 3 月新闻出版署主持召开了两项国家标准的审定会,形成“报批稿”上报国家技术监督局。1990 年 7 月 30 日国家技术监督局批准发布了《图书在版编目数据》(GB12451-90)、《图书书名页》(GB12450-90)两项国家标准,并要求从 1991 年 3 月起实施。

1991 年 1 月,新闻出版署向各出版单位发出实施图书在版编目的通知,并决定由新闻出版署信息中心负责具体实施工作。1992 年 11 月,信息中心举办“首期图书在版编目业务培训班”,邀请人民出版社、商务印书馆、机械工业出版社等 41 家出版社进行实施 CIP 国家标准的试点工作。1993 年 2 月,参加试点工作的出版社开始向新闻出版署信息中心填送“图书在版编目数据工作单”,图书在版编目在我国正式启动。开始时是在《全国新书目》上刊载 CIP 数据的编制情况,1994

年1月创办《全国发排新书半月报》(内部刊物)发布图书在版编目数据,1999年1月改名为《中国图书在版编目快报》,公开发行,每周发布图书在版编目信息,同时以CNMARC和DBF两种格式发行CIP数据光盘版,便于各单位利用。CIP中心的运行软件也不断改进,逐步完善。1999年3月8日新闻出版署发出通知,从1999年4月1日起全面推行与实施CIP数据标准。

我国图书在版编目(CIP)工作,自1993年开始实施,在整个出版业的共同努力下,到2000年已有了长足的发展,下列数字可以说明其进程:

年 份	在版编目数据	当年图书出版量	CIP占当年出版量的百分比
1993	3200	66313	4.8%
1994	18000	69779	25.8%
1995	25000	59159	42.2%
1996	27000	63647	42.4%
1997	28000	66583	42%
1998	40406	74719	54%
1999	74761	83095	90%
2000	88419	84235	
2001	99000	91416	

2000年编目量较1993年增加85219种,增长2663%;占当年图书出版量的比例,也从1993年的4.8%增长到1999年的90%。《中国图书在版编目快报》周刊已成为社会各界了解新出版图书的重要工具,CIP数据光盘版也已是发行界和图书馆界编制征订与采访机读目录的重要来源。

3. CIP数据的内容与机读形式

《图书在版编目数据》(GB/T 12451—1990)国家标准对著录内容与形式(含标识符号)的规定,和《普通图书著录规则》国家标准保持一致,但由于图书在出版过程中数据元素不全,有些项目没有列入。CIP数据内容分为著录数据与检索数据两个部分。

著录数据 是对图书识别特征的客观描述,包括书名与著作责任者项、版本项、出版项、丛书项、附注项和标准书号项6个著录项目。(“【】”内说明GB/T1245—2001修改情况)

书名与责任者项

正书名(包括交替书名、合订书名) 【修改版无括号中说明文字】

卷数(与中国古籍书名相连的) 【修改版无此项】

章回数(与中国章回小说书名相连的) 【修改版无此项】

并列书名

副书名及说明书名文字 【修改版改为:其他书名信息】

第一著作责任者 【修改版改为:第一作者】

其他著作责任者 【修改版改为:其他作者】

版本项

版次及其他版本形式

与本版有关的著作责任者 【修改版改为:与版本有关的作者】

出版项

出版地

出版者

出版日期 【修改版改为:出版时间】

丛书项

正丛书名

并列丛书名

丛书著作责任者 【修改版改为:丛书主编】

国际标准连续出版物编号(ISSN)

丛书编号

附属丛书名

附属丛书编号 【修改版无此项】

附注项

译著的说明

翻印书的说明 【增:教材及教学参考书的说明】

各项的附加说明

标准书号项

国际标准书号 【修改版改为:按 GB/T 5795《中国标准书号》执行】

检索数据 提供图书的检索途径,包括图书识别特征的检索点和内容主题的检索点。

正书名(包括交替书名、合订书名)

副书名 【修改版改为:其他书名信息】

第一著作责任者 【修改版改为:第一作者】 【增译者】

其他著作责任者 【修改版改为:其他作者】

正丛书名 【修改版无此项】

丛书著作责任者 【修改版无此项】

附属丛书名 【修改版无此项】

主题词

分类号

其中图书识别特征的检索点与著录数据是一致的,而内容主题检索点主题词及分类号需进行赋值标引。

CIP数据的机读格式与CNMARC保持一致。

新闻出版署信息中心在实施CIP标准时,没有要求出版社将CIP数据工作单连同该书的全份校样,或该书的书名页、目次页、前言、内容简介等校样一并上报,只是要求出版社填写一份比CIP数据内容多几个项目(内容简介、价格、责任编辑、印次、页数、印数、字数等)的工作单(见“图书在版编目(CIP)数据工作单”),上报CIP中心,CIP中心据以进行核定。

CIP中心核定后,给予该CIP数据一个核定号,返回给出版社,出版社据以印在图书的版权页上。其样例如下:

图书在版编目(CIP)数据

苔丝/(英)哈代·托马斯著;晓燕译. —哈尔滨:

哈尔滨出版社,2000.10

(世界文学名著经典)

ISBN 7-80639-402-8

I. 苔--- II. ①哈--- ②晓--- III. 长篇小说—英国—近代
IV. I561.44

中国版本图书馆 CIP 数据核字(2000)第 050344 号

CIP 数据 CNMARC 格式:

```
00492nam0#422001932##450#00100130000001000180001302100210003110000410005
21010013000931020015001062000037001212100025001582110012001834610028001956100
0250022369000120024870100250026070200130299▼012000050344▼##$a7-80639-
402-8▼##$aCN$b(2000)050344▼##$a20000926d2000###$ekmy0chiy0120
#####ea▼0#$achi$ceng▼##$aCN$b230100▼1#$a苔丝$f(英)哈代·托
马斯著$g晓燕译▼##$a哈尔滨$c哈尔滨出版社▼##$a2000.10▼#1$12001#
$a世界文学名著经典▼0#$a长篇小说$y英国$z近代▼##$aI561.44▼0#$a
哈代$b托马斯$c英$4著▼0#$a晓燕$4译▼▲
```

4. CIP 的编制过程

实施范围 CIP 中心规定,除下列类型外,所有装订成册的图书均实施在版编目:

中小学课本、教材(中小学教师用书要进行在版编目);

低幼读物(幼儿园教师用书要进行在版编目)

连环画册(大开本精制的连环画册要进行在版编目);

图片、画页以及未装订成册的散页画辑、明信片等;

未装订成册的地图、游览图(交通图),以及教学挂图;

折叠的歌片、曲谱、字帖等;

单独印行的国家标准、行业标准、地方标准、规程、规范(汇编本要进行在版编目)。

操作流程

出版社在图书发稿付排之前,由责任编辑按 CIP 中心编印的《图书在版编目工作手册》^[5]规定,填写“图书在版编目工作单”交本社负责 CIP 工作的组织协调人;

各社负责 CIP 组织协调人将本社责任编辑填写的“图书在版编目工作单”集中寄送新闻出版署信息中心(CIP 中心);

CIP 中心按各社报送的“图书在版编目工作单”编制 CIP 数据,标引中图法分类号和主题词,并给一该 CIP 数据的审定号,加工成可以排印的 CIP 数据标准格式,退回出版社;

出版社按 CIP 中心退回的 CIP 数据印刷格式印在图书的版权页上。

5. CIP 的社会作用

我国出版业全面编制 CIP 数据,提高了出版业的专业水平与学术水平,在标准化道路上迈出了大步,缩小了和发达国家出版业的差距,与国际市场初步接轨,对国际文化交流做出了贡献,使各国学者更方便地了解中国出版的图书,为实现 UBC(国际书目控制)目标,促进资源共享,履行了应尽的责任,受到了国际文化界的好评,也为我国出版业走向世界开拓了广阔的前景。

CIP 数据在图书发行上的作用。随着 CIP 数据复盖面日益增加,逐渐为广大用户重视,1999 年《中国图书在版编目快报》及其光盘版出版以来,已成为发行界了解图书出版情况的重要信息来源。尤其是各地的图书馆供应商,更是以《快报》的光盘版散发给各图书馆作为预订新书的依据。

图书在版编目 (CIP) 数据工作单

1. 书名与责任者项	正书名 包括 交替书名 合订书名		卷 (册) 次	
	并列书名		章回数	
	副书名及说明文字 (包括分卷册书名)		正文语种、 各种文字对照	
	第一责任者 及著作方式			
	其他责任者及著作方式 (包括分卷册责任者)			
2. 版本及出版项	版次	印次	其他版本形式	
	与本版有关的责任者	责任编辑		
	出版地	出版者	出版年、月 重印年、月	
3. 载体形态项	页数或卷册数	图表	开本或尺寸	
	附件	字数	千字	印数 册
4. 丛书项	正丛书名		丛书编号	
	并列丛书名			
	附属丛书名		附属丛书编号	
	丛书责任者		ISSN	
5. 附注项				
	翻译图书的书名原文			
6. 国际标准书号、装订形式、价格	ISBN			
	装订形式		定价或估计	
7. 提要项	内容提要			
8. 排检项	主题词		分类号	

一些网络商也以 CIP 数据作为书目报道的重要信息源。CIP 数据的逐步完善,还为编制《在版书目》(Book in Print, BIP)创造了条件。

图书馆界受益殊多,不仅以 CIP 数据预订新书,用其机读数据建立预订文档,还据以编制图书目录。尤其是中、小型图书馆,基本上是按 CIP 数据进行著录。就是一些大型图书馆,也都利用 CIP 数据建立预编目文档,新书到馆后再进行修改、补充。从而大大节省了编目人力,加快了编目速度,提高了工作效率。

CIP 数据的普遍编制,推动了书目社会化的研究。1990—1997 年关于在版编目的论文就有 81 篇,占同期文献编目研究论文总数的 8%,居第 6 位。⁶⁵各方面均认识到 CIP 数据是所有书目信息的源头,做好 CIP 是建立社会化书目事业、实现更大范围资源共享的关键,所以提出了各种各样的建议,形成一个研究领域,并成为硕士生的论文选题。

6. 存在的问题

图书在版编目在我国经过了 10 多年的酝酿、讨论、准备,方开始实施,人们对其寄以极大期望。但 CIP 实施以后,其功用与效益与人们期望相差甚远,没有解决发行部门与图书馆界重复编目的现象,在书目标准化、规范化方面的影响也不显著,对出版行政管理亦未能起到应有作用。因此社会各界近年来不断指出 CIP 存在的问题,分析原因,提出各种改进建议,仍然对以 CIP 数据为源头,建立社会化书目事业寄以厚望。

(1) 国家标准的问题⁶⁷⁾

国家技术监督局 1990 年 7 月颁布《图书书名页》与《图书在版编目数据》两项标准时,对我国出版界已有规定和习惯做法进行相关协调不够,出版社在执行中产生一些问题:

以往在图书版权页(主书名页背面)上惯有的内容提要、责任编辑、责任校对、装帧设计等内容被排除在外,又未说明这些内容置于何处。

国务院出版口 1972 年 12 月 7 日发布的《关于图书版本记录的规定》,要求版本记录包括书名、作者、出版者、印刷者等;国家标准《图书书名页》规定“版本记录”只提供 CIP 数据未包含的印刷发行、载体形态等记录。在执行国家标准《图书在版编目数据》的通知中,未能就原先关于“版本记录”的规定作出相应说明。

关于各种形态的图书(如竖排本、外文图书、少数民族文字图书)等,缺乏说明;对 CIP 数据的字号、字体,所占的版面大小及位置也无明确规定。

新闻出版署只规定 CIP 中心对 CIP 数据进行审定,而对《图书书名页》标准未作明确说明,基本上一直处于无人监督执行的状况,长期以来形同虚设,失去了应有的技术法规性质。《图书书名页》标准执行不力,对 CIP 数据的编制和图书馆的编目都带来一些负面影响。如译著的书名原文应印在附书名页,但不少译著的书名原文却印在书名页正面(扉页),与并列书名混淆。

(2) 运作方式上的问题

出版社虽然规定由责任编辑填写“图书在版编目工作单”,但工作单填好以后,还有总编室、校对科、出版科等诸多环节,CIP 的许多数据元素直至付印前才能确定,在此前的各个环节上,都有可能发生种种意想不到的情况,造成数据内容改变,而这些责任编辑是无能为力的。这种多头参与和管理的情况是 CIP 数据差错率居高不下的原因之一。⁶⁸⁾

由于出版社只是提交一份工作单,既没有附带其他校样,也没有要求责任编辑填写分类号与主题词(责任编辑认为“把书分入何种类目责任编辑最有发言权”⁶⁹⁾),又未请图书作者帮助标引相关主题内容,因而 CIP 数据分类与主题标引质量都难尽人意,这也是图书馆界意见较多的一个方面。

我国 CIP 工作实施之初,编写了《图书在版编目工作手册》,发挥了很好的作用。随着 CIP 工作的深入开展,实施范围的扩大,工作方式不断改变,图书出版情况也有所发展,手册内容已显得陈旧,要尽快进行修订,以适应当前工作的需要。

在出版社——CIP 中心——书商——图书馆之间,缺乏交流渠道,不能及时吸收各方有益见解,推动 CIP 数据编制与应用的改进,包括有关标准、规范的制定与推广。

在 CIP 数据的审定工作中,对 ISBN 的缺号与重号情况、出版社漏缴样书情况,没有提供相应的统计数据向有关部门汇报,也就没有起到行政管理的辅助作用。

(3) CIP 数据发布上的问题

我国 CIP 数据没有与国家图书馆的 MARC 记录一起发布,图书上所印的 CIP 数据内容又比较简单,不能据以编制完整的 MARC 记录,多数图书馆对 CIP 数据表示了不满意的态度,总的利用率不高。

CIP 数据的发布工作薄弱,印刷版与机读版发行量都很少,既没有成为书商与图书馆订购图书的必备工具,又未能减轻出版社编制书目的负担,更未能为出版社带来明显的经济效益,也就没有成为“联系出版发行部门、图书馆与书目控制部门的纽带”。^[10]

(4) 技术上的问题

虽然我国 CIP 数据从实施起就应用计算机处理,软件系统也逐步完善,但未能建立联机编目网络,手工往返传送,不仅处理速度慢,而且 CIP 中心与出版社沟通不易,责任编辑不能全程参与,数据质量难以保证。人们呼吁“让 CIP 走进互联网”,^[11]用高科技手段推进 CIP 事业。

二、ECIP 产生的背景

1. ECIP 的提出

(1) 数字图书馆时代的编目

计算机在图书馆界的应用,曾经引起一些人对编目工作产生误解,以为使用计算机后,编目工作可以大大省事了。实际情况不然,编目工作比原先的要求更高,工作量更大了。现在启动数字图书馆工程,又有一些人以为有了全文检索,就不需要目录了,这是更大的误解。实际情况是,数字图书馆对书目记录的要求更高,要做的事更多。

全文检索只能在一种图书中查找所需要的数据,如无规范词表控制,同义词都难以找到,即以“电脑”一词检索时,“计算机”的内容就不会出现。若要在几万种、几十万种、上百万种图书中查找所需数据,如无好的书目索引,对每一种书进行全文扫描,即使计算机的运算速度再快,也是很方便的。^[12]因此,书目记录描述的深度就非常重要,书目记录描述到什么程度,检索效果就能达到什么程度。

和传统图书馆一样,图书馆不是图书的杂乱堆放,而是对图书进行有序组织和科学管理,就是编制图书目录,揭示其内容属性,当读者需要时,能为其提供服务。数字图书馆也是要对馆内和网上到处存在的信息资源进行有序组织与科学管理,只是所要做的事情比传统图书馆更加深入,更加广泛,因为它没有馆藏概念的限定。

未来社会,“万千世界统一于数字图书馆的 0 和 1,书籍、期刊、录音带、录像带,乃至古籍、善本、稀世字画甚至 X 光片,都消失了原本的物理形态,只要有相同的属性,就能被同时获取。”^[13]这里的关键是“要有相同的属性”,而信息资源的属性是不能全部由其自然表现的,需要对其进行必要的描述。

(2) “自动编目”的研究

① 国内学者的探索

高文等著《数字图书馆——原理与技术实现》一书的“数字图书馆和传统图书馆主要业务活动的比较”表中^[114]，列出了“自动编目”和“自动标引”两个内容，但只是说“分类和编目这样的技术服务工作将迅速减少”，至于怎样减少以及如何做到自动处理，则语焉不详。此表引自汪冰所著《电子图书馆理论与实践研究》^[115]，汪的解释是“在以用户为中心的电子图书馆中，并且在传统馆藏已完全实现数字化或图书馆不再收藏印刷型资源的情况下，传统的分类业务工作和分类法也就失去其存在的基础。因为此时用户完全可以联机浏览目录，方便地使用各种机读索引从各个角度进行资源检索，甚至还可以浏览全文或进行全文检索，分类号似乎就没有存在的必要了。”“电子图书馆中的印本文献编目工作将在很大程度上实现网络化联机编目，由联机编目中心等机构统一完成，其它图书馆主要通过计算机网络查检所需书目记录，再通过拷贝编目的方式达到避免重复，共享编目成果的目的。”这里，第一段话讲的是分类工作可以不做了，第二段话讲的是统一编目，并不涉及“自动编目”与“自动标引”。

刘炜等著《数字图书馆引论》^[116]到是论及“人工建立索引既慢又贵，数字图书馆中快速变化的海量信息需要另外的方法。有些元数据应该自动生成，有些可以来自专家，也可由经验较少的人士提供，也可要求数字对象的创作者编制适当的元数据，这些元数据最后提交给自动索引程序。”这里的思路是清楚的，即“自动编目”要在信息资源创作过程中实现，图书馆自身是解决不了的。但未提出具体的操作方法。

南京农业大学信息科学系利用新华社电讯稿进行“自动标引”试验，取得一定进展。此前同类工作均由试验者自行录入数据，这次试验利用原始电子文本，是一特色。^[117]

目前还有一些学者用扫描图书印刷版的扉页、版权页等内容的方式，进行数字转换，实现自动编目。^[118]其认识原理是可以的，用于已出版图书的回溯转换尚可，对于新出版图书和某些IT企业舍弃图书印前电子文本不用，而以图书的印刷版进行转换一样，仍然不是完全意义上的“自动编目”。

② 美国学者的论述

William Y. Arms 著《数字图书馆概论》^[119]一书对“自动编目”有较多论述。首先作者说明了建立索引的必要性。“信息检索的任务就是在馆藏中找到符合查询请求的对象。因为计算机没有足够的时间为每一个查询去扫描所有馆藏，进行足够的比较，因此计算机必须有某种形式的索引，便利系统通过查找索引条目不定期检索信息。”“很多信息发现的方法并不检索馆藏中的实际对象，而是检索对象的描述性元数据。”这里回答了分类、编目工作可以减少，甚至不做的说法。

作者接着说明：“描述性元数据通常由专业人员制作。图书馆的目录和科学索引代表着专业人士持续数十年乃至几个世纪以来投入的巨大工作量。这一经济因素对理解当前的发展趋势至关重要。一方面，必须在上述资金投入和专门技术的基础上建立元数据；另一方面，发明既便宜又快捷的元数据制作方法具有很大的诱惑力，这可通过自动标引或提高工作水平的计算机不定期实现。”“自动标引生成的记录一般质量很差，但它们非常便宜。一个功能强大的检索系统，对补偿低质量的个体记录大有帮助。这一点可以通过万维网检索程序来证明，因为它就是自动建立的索引。虽然记录质量不高，但检索服务的成功表明索引是有用的，至少比没有好。”

作者进一步说明：“SGML 和 XML 这样的标记语言的一个潜在用途就是自动标引程序可利用结构标签来建立信息检索的记录。……根据结构标签，字符串可以被识别出并标注出它是什么。这样，标签提供的信息可用于区分特定类型的信息，如作者、标题或日期。”这里已清楚地说明了实现“自动编目”的做法。

(3) ECIP 的提出

① LC 的 ECIP 计划

美国国会图书馆从 1995 年开始实施 ECIP(Electronic Cataloging in Publication)计划,^[20]其内容即是利用印前电子文本加注结构标签方法,进行“自动编目”的尝试。这是该馆进行数字图书馆建设的重要准备工作之一,他们要求出版商提交按下列结构形式加注结构标签的电子文本。^[20]

```
<tp>      beginning of title page
</tp>     end of title page
<sp>      beginning of series page
</sp>     end of series page
<cp>      beginning of copyright page
</cp>     end of copyright page
<toc>     beginning of table of contents
</toc>    end of table of contents
<ch1>     beginning chapter 1
<ch1>     beginning chapter 2
<ch1>     beginning chapter 3
           etc
</ch>     end of last chapter
```

这里只是标明了电子文本中的书名页、丛书页、版权页、目次页和章节,在每一页的内部,还有更详细的标识。以这些电子文本进行自动编目的做法,是实实在在的“自动编目”。

② 我们的认识与所做的工作

我们认为:在相当长的一段时间内,非数字与数字的信息资源将共同存在,并且非数字的信息资源仍将居主流地位。“较为明智的方法是除少数的专门领域外,让数字图书馆与图书馆的传统馆藏继续共存下去,……作为社会公共机构的图书馆继续维护大量的传统馆藏资料,同时提供数字化服务。……在某些特殊领域,数字图书馆可以完全代替传统图书馆的馆藏。”^[22]

“图书馆的读者(用户)既可从书目数据库查找书刊资料线索而获得原书、原刊,也可以直接检索数字对象库,获得多媒体资料。”^[23]因此,在数字图书馆工程建设中,需要研究的,既有数字信息资源的“自动编目”与“自动标引”,又有非数字信息资源的“自动编目”与“自动标引”,两者的数据元素还需要有机进行整合。

去年以来,我们深入分析了我国中文图书在版编目和图书馆界联合编目的现状、取得的成绩与存在的问题,结合《图书书名页》和《图书在版编目数据》两项国家标准的修订,吸收各方对数字文献处理的研究成果,提出了在我国推进 ECIP 计划的构思,进而对中文图书“自动编目”的实现,进行了一系列研究与模拟操作,取得了比较可行的方法。^[24]

2. ECIP 的技术原理

(1) 一次录入,生成两种产品

图书编目实际上是将图书扉页(书名页正面)、版本记录页(书名页背面)等处的有关数据元素,按一定著录规则进行描述,形成书目记录。计算机编目则是用数据库方式把有关数据元素录入相应字段、子字段,通过计算机编目软件进行处理,同时产生目录卡片和机读目录两种产品。

计算机录入排版所产生的图书印前电子文本,本来可产生这两种产品。但即使在信息技术高

度发展的美国,“印刷业已经使用了多年的计算机排版,但是它们的系统完全是为了印刷品的生产而组织的,……它们在应用于数字图书馆之前还需做大量的工作。”^[22]这里有历史的与技术的因素,所以虽然计算机在几十年中产生了无比巨大的信息资源,仅在中国就有几百亿汉字,但却一直不能为计算机所利用。

最早较好地解决这一难题的是美国图书馆自动化专家阿弗拉姆,她设计了可变字段、可变长的数据结构,实现了一次录入,可同时产生目录卡片和机读目录两种产品。虽然卡片打印没有图书印刷对字号、字体那样高的要求,但图书馆界对书目著录的苛刻条件,至今仍然是建立信息检索系统不可忽视的难题,阿弗拉姆很妥善地处理了这一问题。全世界数十万个图书馆利用这一方法,编制了几千万种书目记录,建成了几百个联机编目网络,是计算机信息检索技术最成功的范例。我们应用其方法编制中文图书机读目录,还出版了一些索引/文摘刊物(同时建库)。

全文数据库技术基本上也是应用这一原理,《中国学术期刊(光盘版)》是很好的例子。

(2) 线性文件可进行结构化处理

计算机编目虽然可以产生数字化书目信息,但其制作方法仍然是由人们用手工方式对印刷型图书进行加工而成。“自动编目”的实现,应是利用中文图书印刷版的印前电子文本,从中直接提取书目记录所需要的描述性数据元素,在图书生产过程中自动产生该书的书目记录,其数字化程度较之原先的自动化编目系统就有了质的飞跃。

但印前电子文本是线性文件,要使之成为数据库文件,确定哪些是需要的数据元素及其属性,哪些是不需要的部分,就需要对印前电子文本进行结构化处理。通常有两种做法:一种是以数据库方式录入,定稿后除产生机读数据外,再以数据库文件转换为线性文件,加注排版语句,以产生印前电子文本;一种是以原先的线性方式录入,定稿后在产生印前电子文本后,对有关数据元素加注结构标签,按其结构标签转换为数据库文件。过去已有的此类软件多采用前一种做法,以处理二次文献为多;目前利用OCR技术进行印刷文本全文转换时,则多采用后一种做法。随着文献数字化处理技术的不断发展,操作方法已日趋成熟。但所有这些做法,均未能直接利用印前电子文本产生书目数据。

(3) 属性数据元素在图书物理结构之中

图书馆的书目记录与CIP数据,均包含著录数据与检索数据两部分内容。其中著录数据取自文献本身,也就是书目记录的著录数据存在于该书的印前电子文本之中。通过剖析印刷版图书的物理结构,可确定书目记录所需要的数据元素。

一般图书由外表、开端、正文与结尾四个部分组成。^[23]

外表部分包含:护封(包封、护书纸)、封面(封一、前封面、封皮)、封里(封二)、封底里(封三)、封底(封四、底封)、书脊、勒口。

开端部分包含:衬页、附书名页、扉页、版本记录页、卷首题词、作者图像或其他书影、多作者名单或编辑委员会/组织委员会名单、内容提要/出版说明或编者的话/译者的话、序或作者前言、目录、图版目录、表格一览表、工具书的凡例、使用说明、图例及各类检索表。

正文部分包含:篇、章、节、目或开幕词、讲话、论文、摘录等,工具书则为条目(有时有二级甚至多级条目)。

结尾部分包含:附录、注释或名词解释、英汉词汇对照表、参考文献、索引、跋或后记。

书目记录与CIP数据,主要取开端部分的内容,有时涉及结尾与外表部分。《图书书名页》规定了扉页、版本记录页与附书名页有关事项,《中文普通图书著录条例》规定以扉页与版本记录页为主要信息源,确定这些部分的数据元素并对其进行结构化处理,是实现自动编目的基础。

(4) 实现“自动编目”的根本途径

无论采用哪一种方法对已有印刷型图书进行“自动编目”，都和尼葛洛庞蒂讲的传真机与 Email 的故事是一样的。^[28] 传真方法是用计算机录入，并打印成纸张，而后再用传真机传送至收件人，整个传真过程经历了两次“原子—比特”、“比特—原子”转换。Email 是以数字方式直接传送给收件人。因此，即使以印前电子文本进行编目的二次加工，不仅多了转换过程，而且电子文本与数据库文件之间还不能完全保持一致。

ECIP 是在出版过程中，利用排版印刷的电子文本产生书目数据，这里没有两次“比特—原子”、“原子—比特”转换，也没有线性结构与数据库结构的转换，而是“数字—数字”直接生成。只有实施 ECIP 计划，才是真正意义上的“自动编目”。

ECIP 也是出版数字化的发展趋势。随着信息技术的日新月异，计算机在出版业的应用，已从录入排版进入到出版物自身数字化的阶段，电子书(EBOOK)必然要标识进行检索的数据元素，这与 ECIP 的工作是一致的。

CIP 审定程序还可保证书目质量的可控性，经过作者、责任编辑和 CIP 中心工作人员三道工序，加上联机环境下操作，不仅书目记录的内容与印刷文本完全一致(错也会错得一样)，著录与标引水平都会有很大提高。

三、ECIP 的具体实施

1. 认真执行国家有关标准

信息处理自动化的过程，也就是标准化的过程。信息处理数字化，就需要在更大的范围内实施标准化。

(1) 两个标准同时审定

在制定我国《图书在版编目数据》国家标准时，即考虑到图书的书名页实现标准化是推行 CIP 数据标准化的基础，因此同时制定了《图书书名页》和《图书在版编目数据》两项国家标准。新闻出版署信息中心在实施 CIP 编制工作时，出于当时客观条件的限制，没有要求出版社将图书的书名页(即扉页及版本记录页)、目次页、前言、内容简介等校样与 CIP 数据工作单一起上报，因而出现 CIP 数据与图书实际情况不符的现象，图书书名页不符合国家标准的情况也没有得到纠正。

2001 年，两项国家标准业经修订，结合执行修订后的标准，在实施 ECIP 计划时，可要求出版社将扉页、版本记录页、目次页、内容简介、封面等的电子文本与 CIP 数据的电子文本，一起传送到 CIP 中心，这样既能保证 CIP 数据的质量，又可将书名页等与国家标准不符的地方，及时加以调整。

(2) 相关标准加强协调

ECIP 的实施，除上述两项标准外，还涉及到出版业、图书馆界和计算机领域的有关标准，如《图书流通信息标准》、《普通图书著录规则》、《中国机读目录格式》、《信息技术——通用多八位编码字符集》等。ECIP 的结构处理还要与 EBOOK 的结构保持一致，才能在网络环境下实现更大范围的资源共享。因此，每一工作环节，都要与其他各方进行协调，做到相互兼容。

(3) 统一元数据格式

印前电子文本结构标签的格式，应与目前中文图书的计算机编目格式保持一致，可以文化部发布的 CNMARC 标准为基础，按“自动编目”的特点，作适当调整。

① 由于是用计算机自动处理，关于书目记录著录的范围，一些在手工编目条件下难以做到的事，如内容分析著录^[29]，编制丛书记录，全面反映所有的作者，增加章节目录、作者简介等，只要安排适当结构标签，“自动编目”均可做到，而且还可自动生成篇名条目索引记录(见细则部分)。

② 同理,凡文献中没有出现的内容,则计算机也不能“无中生有”。如 CNMARC 的 1 - - 编目信息块,印前电子文本中是没有的,请作者和责编填写这些字段很困难。同时,在实际应用中,1 - - 块有些字段是为了识别书目记录的唯一性,还是增加检索点,并无有关说明。因此,“自动编目”时对有些字段要作必要调整。

③ 为满足出版界对书目记录的要求,利于出版、发行管理系统的应用,对版本记录页出现的一些数据,如尺寸、印张、版次、印次、字数,需在相应的字段予以描述;出版方面的责任者也要在 7 - - 块反映;此外还要增设一些字段与子字段,以著录分析题名的其他题名信息、主题数据与分类数据;反映并列出版者名称等。

④ 计算机“自动编目”要求唯一性处理,难以介入“人工干预”。所以在字段、字段指示符和子字段的设置中,不宜有选择性的操作。为此,有关字段、字段指示符与子字段需作相应调整。

⑤ 为全面反映图书出版情况,需另行编制丛书/多卷书书目记录。由首先出版的单册印前电子文本的有关数据构成,主要包含丛书/多卷书书名、所有作者、出版说明等描述性元数据,以及丛书/多卷书的检索数据,一般不作下连。随后出版的单册作查重处理。

⑥ 自动编目过程中难以处理规范问题,只是按印前电子文本提取,若有不符合规范要求的数据,由联机编目系统实施后控规范处理。

2. 建立综合书目服务网络

(1) 建立社会化书目服务体制

ECIP 是在图书生产过程中,从图书印前电子文本中以软件方法提取相关数据元素,自动产生数字化的书目记录,因而要在网络环境下运作,方可有更好的效果,书目网络中心的建立势在必行。网络中心一方面为各出版社实施 ECIP 提供技术支撑,例如各出版社不可能配备齐全辅助编目的工具书,中心应在网上提供给各社使用,以保证 CIP 数据的准确制作与及时处理;一方面最大限度地向社会发布,充分发挥其宣传报道新书的作用,扩展图书发行工作。只有建立这样一个社会化的书目服务机构,才能推进 ECIP 的顺利实施。

(2) 全面考虑文化、经济、管理功能

在书目网络建设时,不仅要考虑扩展图书发行的经济效益,还要就金版工程实施和社会数字化建设作出统筹安排。要把 ECIP 作为出版数字化基础建设的一项重要工作,与提高出版业自动化管理水平、推动电子商务模式、促进 EBOOK 制作、扩大中文数字化资源开发等等紧密联系,还要为数字图书馆建设创造良好的条件,使 ECIP 的实施产生最大的社会效益。

(3) CIP-MARC-BIP 综合处理

在图书正式出版前,提取 CIP 数据,也可从最后的图书电子文本中产生完整的 MARC 记录,这是实现中文图书“自动编目”的最佳途径。这一书目记录的充分利用,将产生极大的社会效益,每年减少的重复编目费用就有几千万元。书目网络中心还可用以建立全国性的图书现货目录。在各出版社的配合下,可以逐步建成规范的“可供书目”(Book in Print)。

(4) 提高出版社自动化管理水平

实施 ECIP 计划并自动编制完整的 MARC 记录,还要和出版社建立本社书目信息中心联系起来考虑,实现编辑管理、出版管理、发行管理的高度集成化,全面提高出版社的自动化管理水平。

ECIP 计划与书目网络中心、出版社书目信息中心建设的相互关系见“出版数字化系统处理流程示意图”。