

我国信息资源开发利用研究

《我国信息资源开发利用研究》课题组

一九九二年九月

课题组成员:

姓 名	所在单位	职 务	职 称
徐如镜	交通部情报所	副所长	研究员
吴荣荣	中国科技情报所	主 任	高 工
稽溥礼	中国科技情报所	主 任	副研究员
张凤楼	国家科委情报司	总 工	研究员
陈通宝	国家科委情报司	处 长	高 工
曾民族	国防科工委情报所	总 工	研究员
王友彭	北京市情报所	所 长	研究员
王允贞	商业部情报所	副所长	高 工
陈道泉	国防科工委情报所		高 工
吴光义	中国科学院	主 任	高 工
刘德力	北京市情报所		副研究员
鞠昌整	机电部机械情报所	副总工	高 工
蔡绍莹	国际电子报		高 工
黄国俊	能源部核情报所	主 任	高 工
孙伯庆	化工部情报所	副主任	高 工
梁素珍	农科院文献信息中心		工 程 师
李迪颖	新华社	副处长	高 工
张 行	中国科技情报所	副主任	工 程 师
黄小虎	中国科技情报所		工 程 师
杨 琪	交通部情报所		工 程 师
傅 强	国家科委情报司		工 程 师

目 录

总报告：我国信息资源开发利用研究	1
附件一：世界文献资源的发展概况	30
附件二：国外数据库信息资源的开发利用	41
附件三：我国农业信息资源开发利用研究	53
附件四：我国机械电子工业信息资源开发利用研究	64
附件五：我国化学工业信息资源开发利用研究	71
附件六：我国运输信息资源开发利用研究	81
附件七：我国商业信息资源开发利用研究	89
附件八：中国科学院信息资源开发利用研究	113
附件九：我国科技情报资源开发利用研究	117
附件十：国家经济信息自动化管理系统的信息开发与利用	137
附件十一：北京市城市服务信息资源估算	141
附件十二：大众媒介信息资源量的估算统计表	144
附件十三：全国自建数据库情况	146

《我国信息资源开发利用研究》

总 报 告

一九九一年七月底，国务院电子信息系统推广应用办公室国电办字（1991）第032号文下达电子信息应用专题研究计划，其中专题之五《我国信息资源开发利用研究》下达国家科委情报司承担。按照国务院电子办的安排和要求，八月初，情报司组织有关人员研究了课题组的组成和工作大纲；八月中旬成立了由国家科委情报司主持，有机械部、电子部、化工部、交通部、能源部、农业部、商业部、科学院、国家科委、国防科工委、新华社、北京市等单位的情报和技术部门的领导及技术骨干组成的课题组；根据初步所做的前期准备和预研工作中了解的情况，对这一课题如何在没有国内外的直接经验可资借鉴，工作量很大而时间又很紧的实际条件下完成研究任务的技术路线和工作方法进行了认真研究，并据此研究制定了研究工作方案和调查参考提纲。经过一个多月的工作，在试点的部门摸索出了一些规律和方法，及时组织了交流。到十月底各部门均进行了抽样调查，提出了初步的材料。在向电子办作了阶段性汇报，听取了电子办和有关专家的意见后，一面按照实际情况修改调整了调查和研究的方法，并按此反馈给各抽样点补充抽样调查的材料，同时，根据抽样调查材料的不足，补充进行了部分专家调查。到十二月底，调查报告一段落，一月初，汇总写出此报告的第一稿。经课题组的讨论修改和听取了各方面的意见，二月修改出了第二稿。经电子办组织专家审查，又根据专题评审小组审查提出的意见作了修改。下面就是经这几次修改后所形成的第三稿。

1 信息资源开发利用的研究是规划和发展我国信息产业的必要基础

1.1 基本概念的讨论和研究范围的界定

1.1.1 信息

进行信息资源开发利用的研究，当然必须首先明确信息资源的含义，界定所要研究的信息资源的范围；而要作到这一点，按理又必须首先明确信息的含义。

对信息的定义极多，而且很不一致。例如，《辞海》（1979年版566页）定义：“信息是指对消息接受者来说预先不知道的报道。”；乌家培提出：“信息的属性表明：信息是事物运动的状态及其变化的影射。”[3]前者是从信息的接收方来定义信息，而后者则是从信息的发送方来定义信息，目的和角度不同，就有不同的但都是有用的定义。正如大英百科全书（1989年版552页）中所指出的：“信息现象渗透于物理的和精神的世界之中，信息的多种多样性使得迄今为止所有为信息作一统一定义的企图全都落空了。”考虑到这一实际情况，需要根据课题研究的实际需要合理界定信息这一基本概念的范围。根据本课题研究的需要，在研究信息资源时自然是侧重考虑信息的发送方，但在研究其开发特别是在研究信息资源的利用时，又必然要适当考虑信息的接收方即其效果。因而，在本课题研究中，信息范围的界定侧重从信息的发送方出发，兼顾接收方，以此原则界定本课题研究中信息的范围为：对可能的使用者来说，可能具有使用和保存价值的事物运动的状态及其变化的映射。

在信息的定义和范围界定中还有一个值得注意的问题是：信息和情报这两个概念的关系。由于在我国的文字中情报一词由来已久，原来的含义与谍报类似。现在，《辞海》(1979年版 1991页)定义情报为：“以侦察手段或其他方法获得的有关敌人军事、政治、经济等各方面的情况，以及对这些情况进行分析研究的成果。”和乌家培定义情报为：“情报是指特定(如科学、军事、商业等)活动中从外部获得的用特殊方式传递的涉及有关各方利害关系并有一定时效性的社会经济信息。”[3] 就都是保留了情报的谍报意义。但不能不注意到信息和情报两词在英文中同为 Information，而在日文中又都为“情报”。而这在国际上已通用而且是同义词在我国的不同行业中却习惯的被分别翻译为信息和情报。由于情报一词在中文原有的和引入的概念上的差别，已在实际上造成了一定的混乱，例如，在部分部委的信息机构设置中，信息中心与情报中心并存等等。因而，在规划我国信息产业的发展中，应该适时明确和统一信息与情报这两个词的概念。

1.1.2 信息资源

对于信息资源的定义国内外也有多种说法。

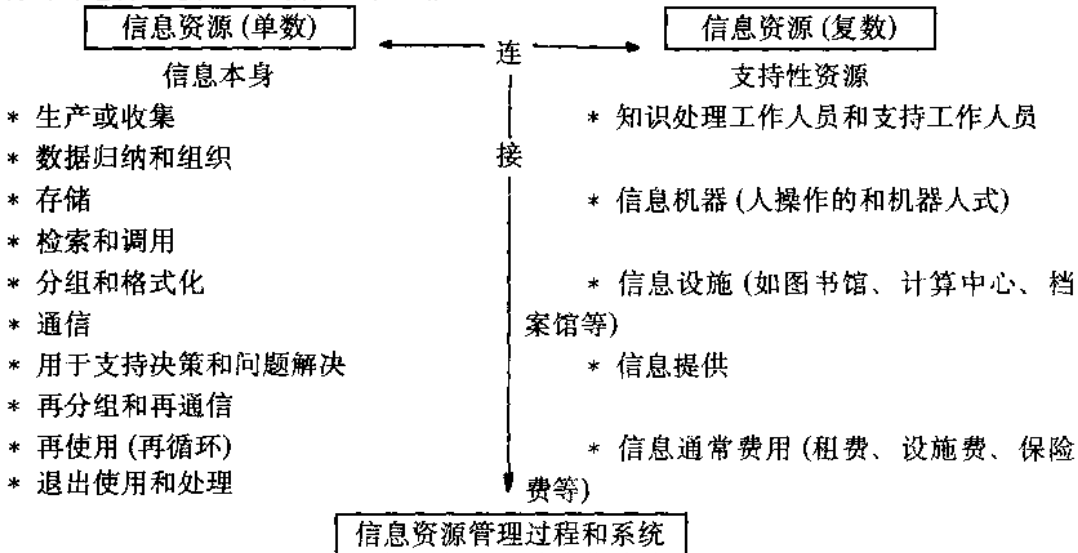
美国 F.W.Horton, Jr. 在其《Information Resources Management》一书中对信息资源提出了以下的概念：

信息资源 (Information Resource) 当英文资源一词使用单数时，信息资源意为其自身，即其内容。例如一个文件、报告、记录或出版物等信息产品或信息服务中的信息，而不是物理的文件或版式等本身。

信息资产或信息财富 (Information Asset)：是指在任一种介质如文献、书、计算机数据库、通讯等等上的信息记录的内容和含义。这一名词的意义和上面所介绍的使用英文单数时的信息资源一词一样。

信息资源 (Information Resources) 当英文资源一词使用复数时，信息资源意为操作信息的所有工具，如设备、供应、物理设施、人和其他资源以及提供上述支持资源的资本、运行投资与费用等。

信息资源管理 (IRM - Information Resources Management)：这一名词同时包括了上面两个名词所包含的意义，即信息内容和信息工具两者。其关系如下图所示：



参照上述概念,我们认为将信息资源的概念分为狭义的和广义的两种比较清晰和易用。狭义的信息资源与 F.W.Horton 所提出的使用英文单数的信息资源的概念相同,即仅指信息的内容本身;而广义的信息资源则为 F.W.Horton 所提出的用单数和用复数两种信息资源所包含内容的总和,即既包括信息内容本身,也包括处理(广义的处理)信息所用的组织、技术、设施、资金、人和其他支持性资源。如下式所示:

狭义的信息资源 = F.W.Horton 的用单数的信息资源

广义的信息资源 = F.W.Horton 用单数的信息资源 + 用复数的信息资源

根据这次各专题中的分工,首先,需将信息资源的研究范围界定于信息(内容)本身,而不包括处理(广义的处理)信息的组织、技术、设施、资金、人和其他支持性资源。

其次,信息作为资源正如美国总统在图书情报白宫会议上所说的就像水、空气、阳光一样。这不仅是指信息对于人类的重要意义,也说明了信息作为资源的一些重要特性:

* 信息作为资源的特性就像空气、阳光和水一样,弥漫于空间和时间的整个参照系之中;

* 信息作为资源像水资源一样,有需要不断探明的一面(如地下水),人们掌握的资源量只是已探明的部分;

* 信息作为资源也像水资源一样,有需要根据明确的使用目的(如饮用水、工业用水等)限定其范围(如领海内的海水不计入饮用水资源范围等);

* 信息作为资源就像水一样,作为原始资源(如天然水)是免费的,但只要经过处理的(如自来水)和某些特殊的(如矿泉水),就是有价的。

因而,信息资源的范围可以参照水资源的情况来界定。以业已探明的、经过处理(包括收集、传递等广义的处理)的,反映经济、社会、科技、政治各方面的活动与发展的信息为研究的范围。

再次,信息的内容总是离于一定的介质上;而且,信息的不同介质还在一定条件下反映信息资源的开发程度(如印刷型、机读型等);而作为信息主要介质的人脑中的信息,尽管往往是个人信息资源的主体,但不论是对具体的一个人还是对人类的总体而言,以脑为介质的信息资源的统计,在目前的科学技术的条件下还均难以实现。因而,在目前的研究中,信息资源概念中的信息内容需要连同包含载信息的介质,而其范围目前则尚只能界定在人脑以外的各种介质。

综合以上分析,这次研究中将信息资源的范围界定为业已探明的、经过处理的、对国家各级和各行业管理和各类生产活动(含知识产品的生产活动)以及对个人和家庭活动有用的、已记录或已交流的信息内容。

1.1.3 信息资源的开发利用

信息资源的开发程度包括多个方面,其中主要的应反映在收集齐全程度、有序化的程度、可获取的程度等方面,现提出下列指标作为反映信息资源开发利用程度的初步指标:

* 信息资源的收全率

$$IRC \text{ 信息资源的收全率} = \frac{\text{已收集的信息资源量}}{\text{信息价值权值}} \times 100\%$$

* 信息资源的有序化指数

$$IRO \text{ 信息资源的有序化指数} = \sum_{i=1}^n \frac{\text{有序化信息资源量}_i}{\text{收集的信息资源总量}} \times \text{有序化系数}_i$$

式中有序化系数表示不同排序方法(如仅按收集时序排、按未规范化的键值排、按规范化键值排、按多个键值排等等有序化程度不同的排序方法)和不同的有序化工具(如书本式目录、卡片式目录、机读目录、数据库等等有序化程度不同的工具)的不同有序化程度的系数。

* 信息资源可获取指数

$$IRA \text{ 信息资源可获取指数} = \sum_{i=1}^n \frac{\text{服务信息资源量}_i}{\text{收集的信息资源总量}} \times \text{可获取系数}_i$$

式中可获取系数表示不同的提供服务方式在可获取时间、空间以及保密(保密级别确定的准确率和解密的及时率)等方面受限制程度系数的倒数。

除了以上各项外,提供服务的信息资源中信息的准确率、提供服务的时差、更新周期等等也反映信息资源的开发程度。

信息资源的利用程度则主要反映在信息资源的利用率、用户使用率以及产生的效益等指标中:

* 信息资源利用率

$$IRS \text{ 信息资源利用率} = \frac{\text{每年被利用的信息资源量}}{\text{收集的信息资源总量}} \times 100\%$$

* 用户使用率

$$IRU \text{ 用户使用率} = \frac{\text{每年实际使用信息资源的用户数}}{\text{应使用信息资源的用户总数}} \times 100\%$$

反映信息资源开发利用情况的指数即为所有上列各项的函数。但是,必须指出,虽然按照课题的分工,以上讨论的信息资源的有关概念和指标中,都只考虑了信息内容本身,但由于信息资源在开发利用时所用的就不仅仅只是作为狭义的信息资源的信息本身,而必然动用了全部广义的信息资源,即动用了为开发利用信息资源所用的设备、资金、人力等,因而反映信息资源开发利用情况的指数也必然地应该包括设备、资金、人力等投入的情况。这样,这一指数在反映信息资源开发利用程度的同时,也反映了信息资源开发利用投入/产出的情况。

信息资源开发利用指数 IRDUF

$$IRDUF = F(IRC, IRO, IRA, \dots, IRS, IRU, \dots, IRBF)$$

式中 IRBF 为反映效益和投入的信息资源开发利用效益系数

1.1.4 信息资源的分类和构成

信息资源根据不同的用途和不同的角度有多种不同的分类方法和构成。现择要介绍如下:

* F.W.Horton Jr. 从信息资源管理出发,将信息资源分别归入源、服务和系统三类之中。以企业为例,其具体的分类构成如下所示:

信息源	信息服务	信息系统
广告/推销宣传	自动/电子数据处理	书目
顾问/经济人	视-听	经济/商业
消费者调查	草图/制图/图表	财务/会计
信用机构	编辑/书写/翻译	资产/地产
国际	教育/训练/会议	法律/立法
市场研究	设施管理	专利/版权/商标
新闻媒介	表格和出版物	人事/工资
联机/文件传递	图书馆/信息中心	购置/合同
出版物订购	邮政和传递	质量控制
美国政府出版物	缩微/计算机输出缩微品	科学/技术
其他	摄影	保安
	印刷/复印	旅行/运输
	记录/档案	管理信息系统
	统计/分析	其他
	电讯/通信	
	分时	
	文字处理	
	其他	

* 按信息产生源的范畴则可分为：人类社会产生的信息资源：包括人口、经济、科技、政治、文化、教育、卫生保健等方面的信息资源；以及自然界产生的信息资源：包括生物产生的和非生物产生的信息。

* 按信息产生的部门分，则 Marcu Porat 《信息经济》中提出的第一和第二信息部门的拓扑结构提供了比较详细的构成。本文不再赘述。

* 按信息的性质则可分为：逻辑思维信息和形象情感信息；也有将其分为情绪信息、知识信息和控制信息。

* 按信息的加工程度可分为：

一次信息资源：包括文献、报告、统计、专利、报刊等等；

二次信息资源：由一次信息摘要或浓缩而得，包括目录、文摘、索引、评论、年鉴等等。也有将评述等分为二次信息资源。

* 按信息资源载体的介质可分为：

口头的：包括面对面交流、通过电话交流、会议、教育和训练等等；

书面的：包括手写的，印刷的，正式出版物等等；

视听的：包括唱片、录音带、激光唱片、相片、录象带、录象盘、电影等等；

缩微的：包括缩微平片、缩微胶卷等等；

计算机可读的：包括联机数据库、可视图文、电子邮政、电子新闻、电子杂志和以磁带、磁盘、CD-ROM 和其他光介质为载体的信息资源等等。

* 按信息媒体的类型则通常把信息资源分为：文字(文献型、事实型、数值型等)，声音，图象，活动图象和最近发展起来的多媒体等等。

* 按信息资源的可取得性和可保存性，一般可分为：

正式记录的：包括机读和非机读的文件、资料、档案和出版物等一般具有保存价值的材料；

非正式记录的：包括过程性材料、半成品材料和私人材料等；

无记录的：包括会议、电话、口头交流信息和教育培训等各类。

* 按信息资源的服务范围则一般可分为：自用的，可共享的，公用的等。

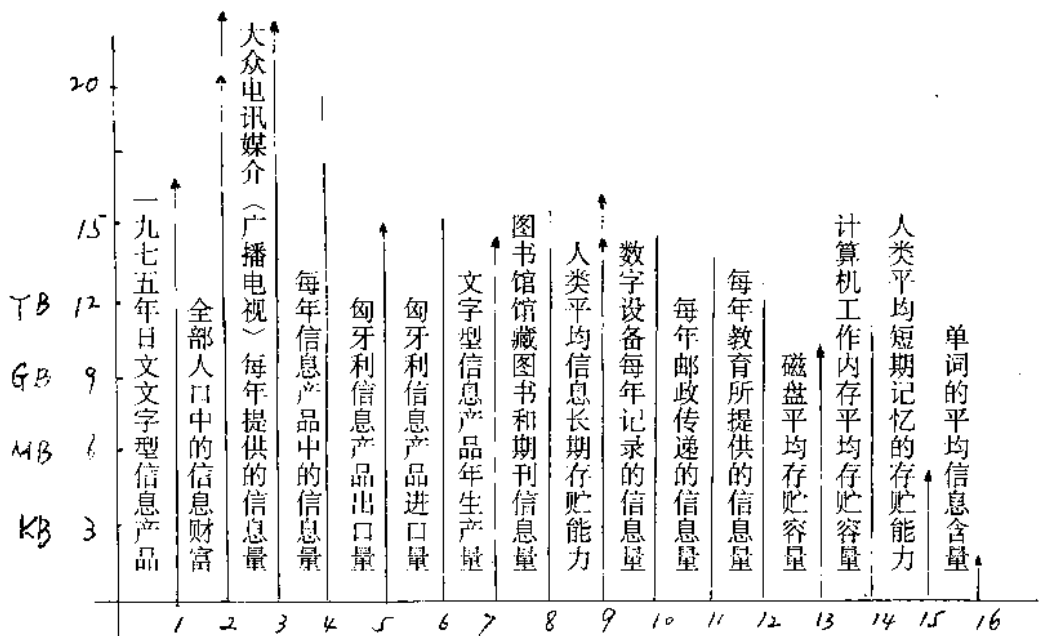
参照以上介绍的多种分类构成，根据此次研究的实际需要和可能，除了从我国的实际情况出发，采用按行政部门进行调查统计外，拟主要采用最后两种分类方法，即按正式记录的、非正式记录的和无记录的来统计信息资源；同时，按自用、共享和公用三类统计不同服务范围的信息资源。

1.2 国外信息资源开发利用概况

1.2.1 世界信息资源总量估计

世界信息资源量：根据国外一些资料介绍，全世界每年约产生 720 亿（另一说为三百多亿）条各种介质的信息。[8] 若按每条平均一千字节估算，则全世界每年产生的信息资源量约在数十 TB（万亿字节）的量级。但据大英百科全书“信息处理与信息系系统”条目中的估计，仅美国的商业和事务处理每年就约处理四千亿张以纸为介质的文件。如以每页一、二千字符估算，则年信息量就已达数百 TB。由于以上所列数字都没有说明界定的范围和估算的方法，难以具体分析造成这些明显矛盾的原因，但可以大致估计界定范围的不同（例如服务范围、保存价值等等的不同）是造成这些估算数字不同的重要原因。匈牙利曾有人对信息资源量的量级作过估算，其初步估算结果示于下图。按其估算，每年生产的文字型信息产品量为近百 TB 量级[19]；日本计划部门统计：1975 年日文文字型信息产品量达百万 TB 量级[19]，按平均重复发行数为一万计，则每年实际产生的文字型信息产品的信息资源量为百 TB 量级。因而，可以这样估计，每年世界产生的文字型信息资源总量可能达到千 TB 甚至更高的量级。

世界信息流通量：根据国外资料介绍，每天约有四十亿比特（另一说法是每天约有四百万组）的信息量向世界各地发送。按这两个基本一致的数字计算，每年世界相互发送的信息量约为近二百 GB（十亿字节）。



1.2.2 世界文献资源的总量和分布

文献是人类创造的知识财富的载体，是信息资源中数量很大、价值很高的一个部分。目前全世界仅发表的科技论文一项，估计每年就约达六百万篇，累计总数已达十亿这一量级，累积了人类几千年来所创造的知识财富；而且有关统计资料比较齐全和准确，可比性比较高。因而，分析文献资源的量和分布对了解、比较和估算信息资源的情况有很大用处。

文献资源主要包括以下十大门类(详见附件一)：

(1) 期刊：据估算，目前世界出版各类期刊约二十万种，而且根据专家测算，期刊种数还以平均每年增长 5—10% 的速度稳定增长，大约每 15 年翻一番。

据 ISDS (International Serial Data System) 统计，截止到 1991 年 4 月份，先后共有 193 个国家的 518.962 种连续出版物(包括期刊、报纸、年报、不定期连续出版物等)进行了登记注册。1991 年 1 月 1 日世界连续出版物按国别的分布情况是：

美国 19.76%， 法国 17.80%， 加拿大 11.51%，
 德国 6.35%， 英国 5.91%， 澳大利亚 4.35%，
 日本 3.57%， 瑞典 2.08%；

以下所占比例在 1% 以上的依次为：匈牙利、芬兰、荷兰、丹麦、比利时、意大利、西班牙和新西兰，中国为 0.46%。按语种统计，使用的文种多达 144 种，常用的有十几种，按文种的分布情况是：

英文占 42.89%， 法文占 21.41%， 德文占 6.15%，

· 西班牙文占 2.82%，日文占 2.47%；

以下依次为：荷兰文、瑞典文、匈牙利文、丹麦文、芬兰文、意大利文、俄文，其所占比例均在 1% 以上；同时使用多文种的占 8.25%；其他所占比例在 1% 以下的文种共占 5.79%。国外有人对科技期刊的利用情况以引文进行了统计，结果发现有 36% 的论文从未被引用过，49% 的论文仅被引用过一次，只有 15% 的论文被引用一次以上。

(2) 图书：根据联合国教科文组织 (UNESCO) 1961 的建议，篇幅不少于 49 页装订成册的出版物定义为书。《美国在版书目》(1990 年光盘版) 收录了二万多家出版商出版的 77 万种图书；《国际在版书目》(1990 年光盘版) 收录了英美以外的 103 个国家，5220 家出版商出版的 17 万种图书；考虑到相互的交叉重复，估计世界每年出版的图书约为 80-90 万种。

据美国和日本的有关资料统计：1989 年，美国共出书 53,446 种，其中，社会科学的占 35.29%，自然科学的占 22.41%，文学艺术的占 21.72%，少年读物占 10.13%；1990 年，日本共出书 40,576 种，其中，社会科学的占 34.5%，自然科学的占 15.8%，文学艺术的占 31.7%，儿童读物占 7.4%。

(3) 报告：美国《公司与工业研究报告文摘累积索引》(1990 年光盘版) 报道：美国公司与工业研究报告每年约出版十万种；前苏联出版的科研报告和实验、设计报告每年也约为十万种。按此估算，全世界每年发表的科技报告约七十余万件。

(4) 会议文献：《世界会议预报》预报两年内将召开的国际会议，其报道的会议数为：1989 年为 3123 个，1990 年为 2986 个，1991 年为 3757 个。《会议论文索引》报道每年一千多个会议的会议论文，1985 年报道量为 72,608 篇，1989 年报道量为 60,484 篇。由这些数字估算，全世界每年的国际会议论文约为十几万篇。如将各国的国内会议也考虑在内，则会议文献的总量至少还将再提高一个量级。

(5) 政府出版物：《美国政府出版物目录》(1990 年光盘版) 报道了美国政府出版物 98,000 篇。英国、法国、日本等国的政府出版物每年大约也达几万种。按此估算，全世界每年的政府出版物可达百万篇的量级。

(6) 学位论文：美国的综合学位论文资料库报道：已收存了自 1861 年以来世界 37 个主要国家九百所大学的六十多万篇博士与硕士学位论文，平均每年约递增三万五千篇。

(7) 标准文献：据统计，国际和各国的标准化组织已颁布了国际和各国标准七十五万件。仅美国根据有关数据库的统计，现行专业和军用等标准已达十万件。

(8) 产品样本：据估算，全世界每年出版工业产品样本约五十万件。

(9) 专利文献：《世界专利索引》每年收录三十三个单位公开的专利五十多万件。美国每年公布新专利约五万到七万五千件，累计已超过四百万件。据估算，全世界每年公布专利约一百余万件。

(10) 技术档案：每年世界各国收藏的技术档案估计约为数百万件。

1.2.3 国外数据库的总量和分布 (详见附件二)

八十年代末，世界提供服务的数据库已近五千个，记录总量约五十亿条，每条记录长度在二百到二千字节之间。服务用机读信息资源总量在 1-10TB 的量级。各年的具体数据列于下表：

年份	数据库 (个)	制造商 (个)	服务机构 (个)	记录量 (百万条)
1979	560	300	290	148
1982	800	450	320	270
1985	3100	1200	670	1680
1988	4062	1820	790	2255
1989	4786	1950	868	4989

数据库生产按国别分布的基本情况是，美国自 1985 年以来，其生产的数据库总量一直保持为其它各国总和的一倍，约占世界数据库总量的近百分之七十。以下数据库产量超过一百个的还有：英国、加拿大、德国、法国和澳大利亚。

数据库按类型分布的情况列于下表：

数据库类型	1985	1988	1989
文字型	1926	3147	3409
书目	1094	1162	1223
专利-商标		55	58
全文	535	1258	1412
指南	287	613	707
字典	10	32	9
数值型	1084	1278	1250
图象		16	34
声音		1	2
电子邮政		101	136

从上表中的数据可以看出，各类数据库中，文字型中的全文型数据库、图象数据库和电子邮政增长最为明显。

数据库按专业类型分布的情况列于下表：

专业类型	数据库数量	所占百分比
商业	1687	33
科学/技术/工程	996	19
健康和生命科学	576	11
法律	447	9
社会科学	393	8
交叉学科	335	7
综合	327	6
新闻	186	4
人类学	184	3

从上表中可以看出，有关商业的数据库占有最大的比重，其次为科学技术。

数据库按语言分布的情况是：英语的数据库占全部数据库的 84.79%，明显地占着主导地位；以下所占比例在 1% 以上的依次为：法语、德语、西班牙语和日语。

世界最大的联机数据库服务系统 DIALOG 系统, 其中所收录的二次文献已覆盖五十多种语言的六万五千种期刊; 而据称, 该系统所用的数据库约占世界各系统所用数据库总数的一半; 据此估算, 大致可以认为国外(特别是美国)的主要信息资源已基本机读化, 并以机读形式提供服务。这反映国外信息资源的开发已达到相当高的程度。

下表列出了 1974-1988 美国主要联机数据库服务系统文字型数据库使用量的发展情况:

年份	1974	1976	1978	1980	1982	1984	1986	1988
使用量 (百万次)	0.75	1.2	2.7	5.0	7.5	13	18.5	28.3

按此估算, 联机的各类数据库的服务总量应可达每年亿次以上; 而 CD-ROM 光盘数据库的发行量在九十年代初已超过千万张盘片, 即使按每张盘片每年只使用十次计算, 总的利用量也将超过每年亿次。因而, 可以认为, 国外仅以机读形式提供世界性公开服务的信息资源的利用量已达每年亿次的量级。也就是说, 经济发达国家公用机读信息资源的年利用量已与人口量处于同一量级。

1.3 信息化指标浅析

1.3.1 国外信息化指标简介

对于信息资源及其开发利用情况的衡量国外尚无比较明确和通用的指标。而多用信息化指标作为综合反映包括信息资源及其开发利用情况在内的综合指标, 作为反映和衡量一个国家、地区、行业或部门信息化程度的尺度。

国外现用的一些信息化指标大致有以下三种类型:

(1) 按从业劳动力或产值的比重表示信息化程度

这是目前世界上比较流行的一类指标。此类指标中常用的有:

* 信息劳动者占总就业人口的比重

$$\text{信息劳动者比重} = \frac{\text{信息劳动者总量}}{\text{总就业人口量}} \times 100\%$$

* 信息产业产值占 GNP 或 GDP 的比重

$$\text{信息产业比重} = \frac{\text{信息产业产值}}{\text{GNP或GDP}} \times 100\%$$

与此类似, 日本在其信息化指标体系中, 分别将计算机、信息处理服务、软件服务、信息提供服务、电信等在各行业的销售额与 GNP 的比值作为信息化的副指标(主指标下面另行介绍)以分别反映工厂自动化、办公室自动化、商店自动化和家庭自动化的进展情况。

* 信息经济增加的价值占 GNP 或 GDP 的比重

$$\text{信息经济比重} = \frac{\text{信息经济增加价值}}{\text{GNP或GDP}} \times 100\%$$

此指标通常还分别按第一信息部门(包括所有向市场提供信息产品和信息服务的部门)和第二信息部门(包括为内部信息生产和信息消费活动服务的政府部门和非信息企业)进行统计。

(2) 按产业从业人员人均信息处理设备、软件、通信能力的占有率及其变化情况作为信息化的指标

这是日本提出的三维信息化指标体系中的主指标 I3, 其指标是:

$$I-H(i) \text{ 硬件装备率} = \frac{\text{产业计算机设置金额}}{\text{该产业从业人员数}} \quad (\text{金额/人})$$

$$I-S(i) \text{ 软件装备率} = \frac{\text{产业五年软件费用累计}}{\text{该产业从业人员数}} \quad (\text{金额/人})$$

$$I-T(i) \text{ 通信能力装备率} = \frac{\text{产业通信线路容量}}{\text{该产业从业人员数}} \quad (\text{bps/人})$$

$$E-H(i) \text{ 硬件装备率比} = \frac{\text{对应年硬件装备率}}{\text{基准年硬件装备率}} \times 100$$

$$E-S(i) \text{ 软件装备率比} = \frac{\text{对应年软件装备率}}{\text{基准年软件装备率}} \times 100$$

$$E-T(i) \text{ 通信能力装备率比} = \frac{\text{对应年通信能力装备率}}{\text{基准年通信能力装备率}} \times 100$$

(3) 信息利用潜力指数 (index of Information Utilization Potential)

这是联合国教科文组织研究提出供国家信息政策和规划制订部门用的一种复合指标。IUP的基础是由230个变量组成的变量集，其中27%的变量反映国家的基本情况；20%的变量反映信息需求和信息使用；其余53%的变量反映信息资源和信息活动。所有这些变量分别聚集入21个结构化的集合，分别计算出相应的各个集合的指数；然后，计算出国家背景情况、信息需求和利用、信息资源和活动等三个子集的合成指数的平均值；再根据这三个合成指数的平均值即可计算得到综合的信息利用潜力指数(IUP)。这一指标尚处于研究阶段，仅在几个发展中国家作过试验。

1.3.2 指标体系简析

从根本上说，反映信息化程度的首先和最直接的应该是信息资源量及其实际开发、利用程度。但从以上介绍的国外一些反映信息化程度的指标来看，实际采用的基本均系间接反映指标。其中部分指标已实际成为国际通用的可比性指标，我国信息经济界也多采用这些指标。这可能反映了两个基本情况：一是这些间接指标在一定条件下，具有实际可用性；二是采用以信息资源及其开发、利用程度为基础的直接指标体系存在实际困难。

但是不难看出，目前国外使用的这些指标对我国并不完全适用；换句话说，这些在某些发达的资本主义国家实际可用的指标如放在我国使用可能就难以反映信息化的实际程度，甚至可能还会产生错误的导向作用。例如，信息产业的劳动力和产值所占的比重反映的是信息产业的发展程度，但我国在目前和在相当长的一个时期内，由于劳动就业、机构体制和价格关系等难以一下理顺，特别是信息的价值尚未被很好认识、信息市场尚未健康发育的情况下，这一指标可能就难以准确反映我国的实际信息化程度。再如，各行业的信息技术装备率等指标只是从设备一个侧面反映信息化的程度，而根据我国的实际情况，由于信息设备基本尚系由国家投资，信息设备的购置与实际的信息需求和信息资源的开发利用程度实际远非是一致的，计算机等设备不是按需购置，买了不用等等情况在我国目前可以说是比比皆是，因而，在我国这一指标只能反映国家对信息化的投资强度，实际上难以真实反映各行业和各地区的信息化程度。

这里需要强调的是，指标体系在计划经济占有重要地位的我国具有很强的导向作用。例如，我国经济发展中长期存在过热和重数量、轻质量的问题，不能不说与经济指标中存在的问题直接有关。而在规划和发展我国的信息产业中如果采用劳动力或产值比重，特别是如果采用信息设备的装备率作为衡量信息化程度的指标体系，则将产生明显的错误导向作用，助

长一些错误倾向的发展，从而势将严重影响我国信息产业的健康发展。

邓小平同志关于“开发信息资源，服务四化建设”的指示，从根本上反映了开发信息资源是发展现代化信息系统和现代化信息服务产业的基础，而服务四化建设则是我国发展这一现代化产业的目标这两个基本点。根据这一精神，探索建立以信息资源及其开发、利用程度为基础的直接反映信息化程度的指标体系，以此作为衡量和指导发展我国现代化信息系统和信息服务产业的基础，对我国信息服务产业的合理而健康地发展进而对我国经济的发展具有重要的和深远的意义。

但是，根据已掌握的国内外情况，全国性的信息资源及其开发利用程度的调查研究在世界上尚未见先例。一些文献中曾提到过世界或某国信息资源总量的估计数字，但均未提及取得这些数字的方法。因而，信息资源的调查研究从理论到方法基本均需要从新探索；而调查如需全面进行，其难度与工作量估计将与人口调查相当，甚至更大。从总体来看，全国信息资源及其开发、利用情况的调查研究是一项需要费大力量长期进行的、巨大的系统工程。这也可能是国外回避这一条路的重要原因。

因而，摆在我们面前的似乎只有两条路：一条是沿用与我国实际情况不很适用的国外指标体系；另一条则是去闯、去探索一条建立符合我国实际情况和实际需要而又实际可行的指标体系的路子。如前所述，从根本上反映信息化程度的是信息资源量及其开发利用程度。因而，不论是探索在我国建立和采用以信息资源量及其开发利用程度为基础的直接指标体系，或是探索与此指标相关的衍生指标体系，都需要从我国信息资源及其开发利用程度的研究入手；同时以此为基础，把我国信息资源及其开发利用程度的实际情况研究清楚，我国这一现代化产业的规划和发展才能做到基础情况明，才能真正做到从我国信息资源及其开发利用的实际情况出发，按照信息资源开发利用的实际需要和已经达到的程度正确规划和指导信息系统和信息服务产业的发展。因而，可以不过份地说，信息资源及其开发利用的研究是正确规划我国信息产业并使之健康发展的必不可少的基础。

2 我国信息资源及其开发利用的现状

2.1 信息资源及其开发利用调查研究的方法

开展本课题研究的基础，同时也是本课题研究的主要内容就是要从宏观上初步掌握我国信息资源量及其开发利用程度的大致情况。在这基础上才能分析实际存在的问题和从实际出发提出相应的建议。

但是，在本课题所做的文献工作中，尚未见到国内外进行信息资源量及其开发利用程度调查和研究的先例，没有可供借鉴的方法和经验。而信息资源如同空气和水资源一样，渗透、充斥于社会的所有领域。而且每日每时还在不断地产生和交流；并在传播中使自身得到延伸扩展。同时又产生相互交叉重复，情况极其错综复杂。如果将所有信息产生的主体——人脑中的信息和信息交流的主要形式——人际交往都考虑在内，信息资源在日前条件下可以说是无法调查统计的。因而，一是必须如前节所述，根据目的合理界定其范围；二是需要探索实际可行的调查方法。这是本课题开展研究，初步掌握我国目前信息资源及其开发利用概况所必须首先解决的前提。

2.1.1 调查一般方法的讨论

一般来说，调查的方法有：全面普查，抽样调查和专家调查。现根据本课题的实际需要

和可能，简单分析如下：

如果条件可能，当然全面进行普查是掌握情况最直接和最可靠的方法。就信息资源及其开发利用程度的调查而言，由于信息人人皆有，处处皆在，进行普查就更有必要。因而，问题的焦点就在于是否存在实际可能性。研究中有同志提出以人为统计对象，统计每人产生的各类信息量（包括将集体产生的信息分解到个人的部分），然后进行汇总统计的建议。由于人是产生和接收信息的主体，从理论上说，以个人为统计对象是合理的。但实施起来有很大的技术难度，而且整个工作量将明显大于人口普查，因而实际上将难以在短期内实现。

抽样调查是在不可能或不需要进行全面普查时所经常采用的方法。抽样调查的结果能否基本反映全面的情况，首先决定于抽样点的数量，特别是在所调查的对象在各点的分布没有什么规律，或是分布规律还未被认识的条件下，必要的抽样数量是保证抽样调查成功所必不可少的条件；但是如果能够研究分析得到所调查对象的分布规律，那么就可以相应减少抽样调查的抽样点数。

专家调查法，国外称为德尔菲法，是一种利用专家的分析判断能力，帮助进行宏观定性预测的方法。在无法进行直接定量调查的情况下，用此法帮助进行间接的、定性的分析也是可取的。

考虑到这个课题的研究时间很短，根本不可能考虑进行全面或局部的普查，因而确定采用以抽样调查为主，对一些难以进行调查和调查所得数据难以进行分析之处，再辅之以必要的专家调查。这样，这一课题的成败在很大程度上就决定于抽样调查的质量，而这又在很大程度上决定于抽样调查的方法。

2.1.2 抽样调查具体方法的研究

对信息产品量的抽样估算：

各类信息产品在质量和数量上是构成信息资源的主要组成部分，并是信息资源中业已开发、利用最广和可以控制的部分，比较容易进行统计。根据我国的实际情况，只要对各类单位自生产的机读型信息产品（含数据库、软件等）、印刷型信息产品（包括各类出版物和文件、资料等，统计中需与数据库去重）以及档案（需与印刷型产品去重）进行合理抽样统计，参照现有有关各类单位数的统计数字，就有可能估算出信息产品的总量。

对生产活动产生的信息量的抽样估算：

生产活动（含知识产品的生产活动）是社会各类活动的基础。生产活动所产生的信息也是信息资源中数量最大、最基础的主要组成部分。同时，也是产生各类管理信息的直接基础。因而，也是这次抽样调查的重点。

从理论上讲，由于生产活动和在生产活动中产生的信息量存在明确的因果关系，因而反映生产活动的各类指标与生产活动中产生的信息量之间应存在某种函数关系。国外也有人对此一关系作过探索，例如有人估算：生产量每增加一倍，信息量将增加三倍等等。探索这一关系是探索信息资源发展规律的必要基础，对合理规划信息系统随生产发展而发展具有重要而深远的意义。但探索这一关系的基础是必须掌握生产活动中产生的信息量。

因而，掌握生产活动产生的信息量及其规律是掌握信息资源及其发展规律的基础；是抽样调查中所要突破的焦点。经研究分析，各行业（或各种类型）的生产活动基本都存在一种“基本活动单元”，这些基本活动单元产生的信息量在一个行业或同一类型的单位中差异不大。例如：

铁路和公路运输中的车次与空运和水运中的航次。一个车(航)次产生的信息量比较稳定,随运输距离和客、货运量变化的影响很小。

工业中的一个班组一个班次的生产活动所产生的信息量在同一类型的工业中也相对比较稳定,产量、产值与产品型号的不同对产生的信息量影响不大。

科研中的一个课题(从开题到结题)。虽然所产生的信息产品(如成果报告等)的信息量会因课题的不同存在较大的差别,但作为一个课题所产生的科研管理信息量则基本不随课题的大小和性质而变。

教育中的一个班次(从入学到毕业)。与科研相类似,虽然教育中所用的和所产生的信息产品(如教材、讲稿和其他教学材料)会因专业和课程性质的不同而不同,但所产生的教学管理信息量在一类学校中是相同的。

同样,商业中的一票货(从进货到售出)、金融业中的一次存取或借贷、其他服务业中的一次服务等所产生的信息量也基本不随货的品种、存取的金额或服务的费用等等而变。

根据上面的简单分析可以看出,只要分析找出这类产生的信息量相对恒定的基本活动单元,按照基本活动单元分类选择抽样点,抽样调查得出这一基本活动单元在单位活动周期内产生的平均信息量和全行业(或同类型单位)的基本活动单元总数,即可估算出全行业(或同类型单位)生产活动产生的基本信息量。同时,还可从中探索生产活动产生信息量的规律。

对管理用信息量及其开发利用程度的抽样估算:

前一阶段我国信息系统的建立和信息资源的开发利用的重点主要侧重于管理信息方面。“七·五”重点建设的十二个大系统中的绝大部分是以管理(指挥、控制)为目的的。而在今后,根据我国的实际情况,管理也将仍然是信息资源开发中的重点。因而大致掌握管理信息资源量及其开发利用的现状,是总结以往,规划今后的必要基础。所以也是本次抽样调查的重点。抽样调查目前各级、各类管理部门实际拥有和使用的信息量及其实际的开发利用程度,就可能从而估算出全国管理所需的信息量及其开发利用程度。根据这次研究的实际需要和可能,将抽样的重点放在部委和基层这两头。

对信息流量的抽样估算:

信息流量是从一个侧面反映信息资源量和信息利用量的重要数据。通过对邮政、电信、会议以及报刊、电影、电视、广播等大众传播媒介进行合理抽样调查就有可能大致估算全国的信息流量。

此次调查研究即采用以上方法进行抽样试点,并在实践中多次修改和调整了抽样调查和估算的方法。

2.1.3 抽样调查点的选择

根据上面讨论的原则,为了便于按基本活动单元进行抽样估算,这次抽样调查以行业为主分类选点,辅之以按城市的抽样。在行业中则以能够进行估算为基本要求,分别确定抽样点的数量和分布。在实践中经调整后,实际选定和完成的抽样点是:

行业:

工业:机械电子部,化工部;

农业:农业部;

交通:交通部;

商业、财政、金融:商业部;