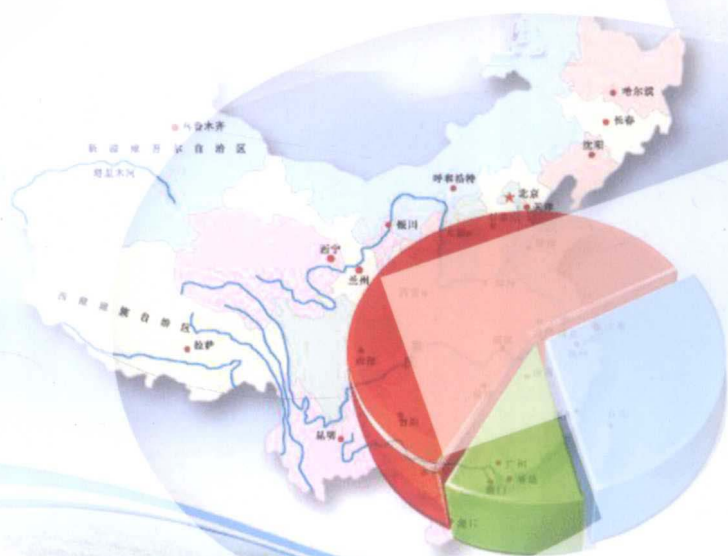




地理信息系统理论与应用丛书

农村社会经济统计信息 空间管理与分析

● 李小娟 孙永华 王彦兵 连健 杨灿坤 著



科学出版社

地理信息系统理论与应用丛书

农村社会经济统计信息 空间管理与分析

李小娟 孙永华 王彦兵 连 健 杨灿坤 著

科学出版社

北 京

内 容 简 介

本书论述 GIS 在农村社会经济统计信息空间管理与分析中的应用。全书共分 8 章, 主要内容包括: 统计信息分类标准与规范、统计时空数据模型、统计时空数据库设计与实现、农村社会经济统计空间数据挖掘、统计数据可视化方法、统计数据分析模型库构建、统计信息网络发布技术和农村社会经济统计数据分析应用。

本书可以作为地理信息系统、统计学等相关专业研究生的参考书; 同时可供地理学、计算机科学、统计学等领域从事空间信息处理、管理与分析的研究人员和工程技术人员参考。

图书在版编目 (CIP) 数据

农村社会经济统计信息空间管理与分析/李小娟等著. —北京: 科学出版社, 2011

(地理信息系统理论与应用丛书)

ISBN 978-7-03-031615-8

I. ①农… II. ①李… III. ①农村-经济统计-经济信息-研究 IV. ①F302.5

中国版本图书馆 CIP 数据核字 (2011) 第 115072 号

责任编辑: 韩 鹏 马云川 赵 冰/责任校对: 刘亚琦

责任印制: 钱玉芬/封面设计: 王 浩

科学出版社出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

骏杰印刷厂印刷

科学出版社发行 各地新华书店经销

*

2011 年 6 月第 一 版 开本: 787×1092

2011 年 6 月第一次印刷 印张: 12 1/4

印数: 1—2 500

字数: 277 000

定价: 59.00 元

(如有印装质量问题, 我社负责调换)

前 言

随着我国社会经济的快速发展，以周期性普查为基础、经常性抽样调查为主体，以必要的统计报表、重点调查和综合分析为补充的统计调查业务也得到快速地发展，统计数据和统计信息量以几何级数迅猛增长。与此同时，国家统计局部门、政府决策部门和社会公众对统计工作也提出了更高的要求，需要统计信息和空间信息资源的全面整合，深入的统计分析和更及时、更完善的统计服务。

针对统计工作对空间信息技术的迫切需求，“十一五”期间，国家科学技术部在国家高技术研究发展计划（“863”计划）地球观测与导航技术领域设立重点项目“国家统计遥感业务系统关键技术研究与应用”，并委托国家统计局作为主持部门组织实施。其中该项目第四专题开展了基于地理信息系统（GIS）技术、数据库技术和网络技术的“农村社会经济统计信息空间管理与分析运行系统研究”，旨在为政府决策和科学研究提供统计数据与空间数据一体化的信息资源，为全面深入的统计调查、统计分析和及时、完善的统计服务提供先进的技术支撑平台。

课题主要研究内容包括：①农村社会经济空间信息系统建设与应用过程中的主要技术标准和规范研究；②农村社会经济统计数据空间化处理与分析关键技术研究，包括统计数据空间化处理、统计空间信息分析、统计数据多维可视化、统计信息综合及海量统计遥感信息发布等关键技术；③在用户需求调查、业务化运行模式研究基础上，分别研制适于国家级、省级、地/市级、县/市级的农村社会经济统计信息空间管理与分析业务化运行系统，实现统计数据和空间数据一体化的存储、管理、分析和发布；④开展示范应用研究，按照统计遥感项目的总体设计，在国家粮食主产区选择农村社会经济统计工作基础较好的示范区，开展业务化系统的示范应用研究，为在我国统计部门推广应用本系统奠定基础。

在该项目的研究成果基础上我们编写了本书，旨在探索我国农村社会经济统计空间信息系统建设与应用过程中的一般性规律和认识。为便于读者深入浅出地了解 and 掌握本书的主要精髓，我们按照农村社会经济统计空间信息系统建设的体系结构、设计思想、技术实现和应用示范四个环节，由表及里、由浅入深，从理论到实践的顺序编排本书各章节内容，具体包括：第1章：统计信息分类标准与规范；第2章：统计时空数据模型；第3章：统计时空数据库设计与实现；第4章：农村社会经济统计空间数据挖掘；第5章：统计数据可视化方法；第6章：统计数据分析模型库构建；第7章：统计信息网络发布技术；第8章：农村社会经济统计数据应用。

本书主要由李小娟、孙永华、王彦兵、连健、杨灿坤编写。参加编写的还有周春平、滕俊华、程勇、刘晓萌、刘强、张嘉、赵惠敏、张海波、杜伟等。在编写过程中，我们得到了国家统计局农村社会调查司张淑英、赵建华、余新华、万国强、施开分、魏

锋华等的大力帮助和支持，对他们的帮助和支持深表谢意和敬意；参阅、引用了其他书籍和论文的部分内容和思想，在此对相关作者表示衷心感谢。

受时间仓促和研究水平所限，书中难免存在不足之处，恳请各位同仁和读者批评指正。

李小娟

2010年8月于北京

目 录

前言

第 1 章 统计信息分类标准与规范	1
1.1 概述	1
1.1.1 统计信息的分类	1
1.1.2 国内外信息分类编码体系	3
1.2 统计信息分类体系的构建	5
1.2.1 分类体系的构建原则	5
1.2.2 分类体系的构建思路	7
1.3 统计信息编码体系的构建	9
1.3.1 编码方案	9
1.3.2 编码原则	9
1.3.3 统计信息编码体系结构	10
1.4 小结	12
第 2 章 统计时空数据模型	13
2.1 概述	13
2.1.1 时空数据模型的概念	14
2.1.2 时空数据模型的功能	14
2.1.3 国内外研究现状	15
2.1.4 农村社会经济统计数据特征分析	17
2.2 现有时空数据模型	19
2.2.1 序列快照模型	19
2.2.2 基态修正模型	20
2.2.3 时空立方体模型	21
2.2.4 时空复合模型	21
2.2.5 基于事件时空数据模型	22
2.2.6 现有时空数据模型存在的问题	23
2.3 基于事件和面向对象的行政区划时空数据模型	24
2.3.1 基于面向对象模型实现空间实体的表达	24
2.3.2 行政区划数据的特点和变化规律	28
第 3 章 统计时空数据库设计与实现	31
3.1 概述	31
3.2 农村社会经济时空数据库设计	33

3.2.1	总体设计	33
3.2.2	数据库表格结构	36
3.3	农村社会经济时空数据库实现	38
3.3.1	社会经济统计数据入库实现	38
3.3.2	行政区划单元变更维护的实现	40
3.3.3	统计数据时空查询实现	41
第4章	农村社会经济统计空间数据挖掘	44
4.1	数据挖掘与知识发现	44
4.1.1	数据挖掘产生的背景	44
4.1.2	数据挖掘的研究现状	45
4.1.3	数据挖掘的知识表达	47
4.2	空间数据挖掘的基础理论	49
4.2.1	空间数据挖掘概述	49
4.2.2	空间数据挖掘的体系结构与基本过程	51
4.2.3	空间数据挖掘可发现的知识类型	53
4.2.4	空间数据挖掘的方法	54
4.3	农村社会经济统计空间数据挖掘方法体系	55
4.3.1	农村社会经济统计空间数据挖掘的必要性	55
4.3.2	农村社会经济统计空间数据挖掘的总体框架	56
4.3.3	农村社会经济统计数据空间挖掘可发现的知识类型	57
4.3.4	农村社会经济统计空间数据挖掘的流程	59
4.4	小结	60
第5章	统计数据可视化方法	61
5.1	概述	61
5.1.1	统计数据可视化理论研究现状	61
5.1.2	统计数据可视化技术应用现状	61
5.1.3	统计数据可视化主要内容	62
5.2	统计数据可视化理论	63
5.2.1	农村社会经济统计数据的时空特征	63
5.2.2	农村社会经济统计数据的组织	64
5.2.3	传统统计数据可视化方法	65
5.2.4	统计数据可视化	65
5.3	统计数据可视化技术与方法	66
5.3.1	统计数据的空间化	66
5.3.2	统计数据专题地图表达	67
5.3.3	统计数据数字表面模型的构建	71
5.3.4	统计数据与 Google Earth 的可视化集成	75
5.4	统计数据多维可视化子系统的设计与实现	78

5.4.1 农村统计数据多维空间可视化子系统设计	79
5.4.2 农村统计数据多维空间可视化子系统实现	79
第6章 统计分析模型库构建	91
6.1 概述	91
6.1.1 模型库系统理论	91
6.1.2 国内外研究现状	92
6.2 模型库系统基本概念	96
6.2.1 模型特点分析和分类整理	96
6.2.2 模型库	99
6.2.3 模型库管理系统	101
6.2.4 模型库系统	102
6.3 模型库构建方法	103
6.3.1 COM 组件技术	103
6.3.2 系统决策支持问题的分解	105
6.3.3 原子模型的表示与组合	106
6.3.4 组合模式的模型库构建	106
6.4 模型库系统设计与实现	108
6.4.1 模型库系统的总体设计	108
6.4.2 模型库管理系统总体框架	117
6.4.3 模型库与系统集成	122
6.5 应用案例	124
6.6 小结	127
第7章 统计信息网络发布技术	128
7.1 概述	128
7.1.1 统计信息网络发布技术研究现状	128
7.1.2 统计信息网络发布技术发展趋势	129
7.2 统计信息网络发布基础理论	130
7.2.1 WebGIS 的基本原理	130
7.2.2 WebGIS 的设计与实现	130
7.2.3 农村社会经济统计信息发布平台选型	131
7.3 农村社会经济统计信息网络发布系统	132
7.3.1 技术线路与系统架构	132
7.3.2 系统关键技术研究	135
7.3.3 数据组织和数据库设计	138
7.3.4 系统主要功能模块的设计与实现	140
第8章 农村社会经济统计数据应用	151
8.1 统计数据相关分析方法	151
8.1.1 探索性空间数据分析方法	151

8.1.2 多元统计分析方法	155
8.2 贫困区经济发展时空分析	156
8.2.1 贫困时空数据处理及特征提取	157
8.2.2 分析方法与流程	160
8.2.3 区域经济发展综合定量分析	161
8.3 农村社会经济统计数据空间挖掘分析与应用	167
8.3.1 乡镇农业经济发展空间结构模式	167
8.3.2 区域经济发展时空演进结构模式	174
8.4 小结	180
参考文献	181

第 1 章 统计信息分类标准与规范

农村社会经济统计为揭示农村社会经济发展的特点、问题及变化规律，为认识和研究农村经济和社会现象、指导科学决策提供了可靠的依据，从而促进农村社会经济又好又快发展。为适应农村社会经济统计信息处理的需要，促进农村社会经济统计信息规范化管理和信息系统建设，必须确立一个科学而全面的分类系统，并在此分类系统的基础上进行编码，建立我国农村社会经济统计信息的分类编码标准和规范，以指导我国农村社会经济统计空间信息系统建设。

农村社会经济统计信息分类编码是根据我国农村社会经济统计信息管理和应用的实践，综合考虑农村经济特点、农业生产特点及对统计的影响等三方面的要求，选择合适的分类策略、编码方法，建立一套面向信息化领域的农村社会经济统计信息的分类与编码体系以及规范的统计信息分类编码标准及统计信息分类数据库，进而推广为国家及省市级《农村社会经济统计信息分类与编码》的标准。

本章主要包括三个部分：①概述国内外信息分类方法及其研究现状；②农村社会经济统计信息分类体系构建的原则与方法分析；③基于农村社会经济统计信息分类体系，对农村社会经济统计信息进行编码，提出统计信息编码体系，编写农村社会经济统计信息分类编码表。

1.1 概 述

1.1.1 统计信息的分类

1. 信息分类编码的目的

信息分类编码是信息存储、处理、交换及共享的基础（何建邦等，2003）。随着统计信息各种专题数据库应用系统先后建成，有关信息采集、处理与共享技术标准的制定相对滞后。对信息进行分类能够使信息系统条理清楚、易于管理，并能够方便地从中提取不同类型的信息加以分析利用，而采用信息分类编码是解决信息分类问题的有效方法。自从开展信息分类编码标准化工作以来，一直没有适合自身业务特点的农村社会经济统计信息分类编码标准，因而迫切需要进行农村社会经济统计信息的分类与编码，以满足各有关方面对农村社会经济统计信息共享的需求。

农村社会经济统计就是摸清我国农村经济和社会发展的实际情况，为政府制定农村政策、编制国民经济和社会发展规划提供依据，并检查政策和计划执行情况，为我国社会主义现代化建设事业服务。

《农村社会经济统计信息分类与编码》标准提供完整规范的统计信息分类编码体系，

从宏观、系统的角度对农村社会经济统计信息进行全面、细致的分类与编码。它以国家标准的权威性加以规范,促进分类与编码体系在统计基础数据库系统中的应用,为统计基础数据库的建设提供基础的分类保证。

2. 统计信息分类编码的意义

21 世纪是社会信息化、经济一体化的时代,只有充分而快速地获取、处理、传播和统筹利用各种农业信息,进行科学决策,才能确保农业系统沿着良性循环方向高效运转(王亚东等,2002)。当前,“三农”问题始终关系着我国现代化建设的全局,是我国全面建设小康社会首先需要解决的问题,而农业信息化是解决“三农”问题的重要途径之一,这已经成为全社会的共识。农业信息化成为农业现代化发展的必然途径,是农业发展的必然方向(孙九林,2002)。在农业信息化建设中,首先遇到的问题就是编码问题。信息的编码只有遵循有关标准,才能在信息系统中进行有效的数据管理、访问查询和加工处理。农业信息分类编码标准化是农业信息化建设中重要的技术基础,是实现农业信息化的基本前提。

随着农村改革的逐步深入和农村商品经济的更大发展,农村统计已成为农村经济管理中不可缺少的重要手段,是了解和掌握农村改革和农村形势发展变化的实际情况和存在问题的重要途径。开展农村统计是农村经济改革及整个国民经济改革的需要,是管理和指导农村工作的需要,也是实现农业现代化、提高农民生活水平的需要。

农村社会经济统计信息分类编码标准化是统计信息标准化工作的一个专业领域,是统计信息化过程中的一项基础工作。建立统计信息分类编码标准体系及其编码系统将有效增强信息资源的共享,提高信息资源的利用率。对于统计信息管理者而言,在统一的分类与代码标识下进行相似类型的系统开发时,不需要重复对统计信息进行分类与标识,可直接将其纳入统计信息基础数据库,实现了信息最大限度的共享和重用,对农村社会经济统计基础数据库的建设和各类应用系统的建立至关重要。

在“农村社会经济统计信息空间管理与分析系统”中,农村社会经济统计信息分类编码标准占有基础性的地位。标准将统计指标分类直接录入基础数据库,更好地服务于信息检索、统计分析,从而指导我国农村社会经济统计空间信息系统建设,全面建立 2000 年以来历次调整的国家行政区划基础空间数据库,支持统计数据的空间化处理,为决策支持提供依据。

3. 统计信息分类编码存在问题

到目前为止,我国已制定了上百个信息分类编码国家标准,同时,各部门根据本行业的需要,也制定了很多信息分类编码标准的行业标准。目前,我国信息分类编码标准体系建设还不完善,人们对信息分类编码的作用和复杂性认识不够,农业信息分类编码标准化工作薄弱。例如,农业综合统计信息共有 1000 多个调查指标,关于农业生产一项就有 78 个指标,这些指标目前规范化的分类和编码不完善。随着经济全球化进程的加快、全球信息技术及我国信息化的迅猛发展,加强我国信息分类编码标准化工作已时不我待,我国农业信息分类编码标准体系构建研究势在必行。

1.1.2 国内外信息分类编码体系

1. 国际信息分类编码体系

国外早在 20 世纪 60 年代就开始了信息分类编码的研究工作, 针对不同的应用需求, 迄今已开发上百种信息分类编码系统。例如, 西德 Pizt O 教授研制的用于机械产品数字化的混合式刚性码结构系统 (OPITZ 系统); 日本提出的 KK-3 分类编码系统 (Zeng, 1992); 美国波音公司的 BUOCS 系统 (丁宗红和陈军, 1994), 主要应用于企业设计、制造和管理; 荷兰应用科学金属研究所 20 世纪 70 年代初推出 MUTICLASS 编码系统 (Ding, 1994), 等等。

在信息资源分类中, 比较著名的如《美国国会图书馆图书分类法》(LCC)、《杜威十进制分类法》(DDC)、《通用十进制图书分类法》(UDC) 等。这些分类法采用的分类体系较为宏观, 被应用于各种资源检索系统中。

美国国会图书馆图书分类法是现今不少大学图书馆采用的图书分类法, 它用 26 个英文字母做分类编码, 可达成比 0~9 的数字编码更仔细的分类。

杜威十进制图书分类法是针对图书资源来设置分类的, 在分类方法上采用学科分类法, 遵循传统的学科分类依据; 分类码用 3 位数字表示, 除了 3 位数分类外, 一般会有 2 位数字的附加码, 以代表不同的地区、时间、材料或其他特性的论述, 分类码与附加码之间则以小数点 “.” 隔开。

通用十进制图书分类法是由两位比利时书目专家 (Otlet P、la Fontaine H) 19 世纪末在杜威十进制图书分类法的基础上继续研发的分类方法。此种分类方法由数字和特殊符号组成, 对涉及各种知识体系及学科的书籍予以分类。鉴于数字和符号无语言障碍的特性, 该分类法在世界各地的图书馆中被广泛应用。

另外, 一些专业性分类法也较为突出, 如瑞典大学科技图书馆联合会开发维护的《工程信息分类法》、美国数学协会“数学主题分类资源”所依据的《数学主题分类法》, 但这些分类方法行业专属性的色彩明显。

2. 国内信息分类编码体系

在国内, 20 世纪 70 年代末才开始着手研究信息分类编码工作, 在一些国外分类编码系统基础上, 开发各类分类编码系统, 具有代表性的有 JCBM-1 系统、JLBM-1 系统、CS-1 系统和 NGC-CIMS 系统等 (李祖庚, 1987a, 1987b; 曾平春和杨光薰, 1989; 皮德常等, 2002)。

JCBM-1 系统是我国第一个信息编码系统, 该系统是在 OPITZ 系统的基础上制定的, 用于机床行业实施成组技术。JLBM-1 系统是原机械工业部研制的, 该系统结合了 OPITZ 系统、KK-3 系统和 JCBM-1 系统, 克服了 OPITZ 系统分类小而全和 KK-3 系统分类环节过多的缺点。CS-1 系统是台湾研制的一种机床行业编码体系。NGC-CIMS 系统是南京高速齿轮箱厂研制的, 编码对象既有数值类型的, 又有非数值类型的图形数据。

在图书馆领域,国内最具有权威价值的分类方法是《中国图书馆分类法》(简称《中图法》)。在分类体系上,《中图法》从学科分类和知识分类的角度来揭示文献内容的区别与联系,按学科和专业集中文献;在编码体系上,《中图法》采用拉丁字母与阿拉伯数字相结合的混合制标记符号,以拉丁字母标记基本大类,再展开一位字母标记二级类目。

刘延章等对《中图法》进行改造,从而对网络农业信息进行分类编码研究(刘延章和余义芳,2004;王美琴和侯汉清,2005;王崇良,2004)。陈树年(2000)、陈树年和李青华(2005)在保留《中图法》科学、完整、严密的知识分类体系特性的基础上,提出了改造方案,研究了初步实现《中图法》版搜索引擎的方法和途径。

除《中图法》这种适用范围广泛并且影响深入的分类方法外,对于全新分类体系的探索,国内相关学者也进行着不懈的努力,并且取得了一定范围内的成果,提出了几种基于某种理论思想的具体分类体系。

华中农业大学高等教育研究所的杨志惠等(2007)对一般信息资源的分类方法进行了分析比较,提出了基于数字化教学资源不同用户需求的“数字化教学资源分类体系”。王键和甘国辉(2004)提出多维农业信息分类体系的构想。他们在分析农业系统构成和运行特点及农业信息在农业生产、农业行业/产品、市场、社会经济、科学技术、自然资源等方面的内涵和特性的基础上,提出包括农业生产、市场、自然资源、社会经济、科技、时空和农业信息属性在内、多维交叉的分类立方体数据模型。牛振国等(2003)在分析现有农业信息资源分类的基础上,提出面向不同用户需求和网络农业信息资源管理的分类方案:在农业行业分类的基础上,依据农业信息内容的属性特征,可将农业信息分为农业空间信息、农业科技信息、农村社会经济信息、农业相关机构信息、农业自然资源信息、农业生产资料信息和农产品市场信息7类;依据用户对信息需求的差异,又可将各类信息分为基本信息和全集信息,并初步建立了农业信息资源分类编码体系。郭书普(2003)在此基础上,提出一种较为简洁的农业信息分类的方案,将农业信息行业简化为4个门类,分别为种植业、养殖业、林业及其他行业,下设10个大类,59个中类,每个中类下可以再细分。刘世洪等(2006)分析了体系框架二维结构的优缺点,根据印度魏尔曼最早提出的标准体系表三维结构的思想、结合标准化三维空间的概念,在广泛比较的基础上,采用了农业领域(对象)、信息化内容(内容)和标准层次(级别)的三维结构,在每一维结构中又增加小门类,延伸了结构的空间,大大地扩展了标准的存储容量,为标准体系的未来发展准备了广阔的空间,结构上体现了框架的先进性和科学性。钱巍和盛苏伦(2006)基于信息资源的基本适用原则,提出农业信息资源开发和利用标准体系的研究内容,如农业信息资源分类编码、用户视图、农业信息资源描述与表示以及农业信息资源的存储方式几个基础标准。张雪和祝玉华(2004)在国家粮食局的支持下,率先将粮食信息分类与编码提升为行业标准。张茂震等(2005)从森林资源信息的类型特征和管理需要出发,提出基于特征的森林资源信息分类的基本方法和分类编码的实用方案。

3. 信息分类编码标准现状

信息分类编码标准化,就是将信息按照科学的原则进行分类并加以编码,经有关方面协商达成一致,由标准化主管机关批准发布,作为各单位共同遵守的准则,并作为有关的信息系统进行信息交换的公共语言使用。

在国外,世界各工业发达国家都在信息分类编码标准化方面投入了大量的人力、物力。日本早在1920年第一次人口普查时就注意信息的分类编码问题,历经多年的完善与修订,日本的信息分类编码标准已较为完善。苏联在20世纪60年代初由全苏技术信息分类编码科学研究院组织48个部门的各行各业专家学者1000余人,花费10年左右时间,于1971年出版了第一版《全苏工农业产品高位分类集》。同期美国各大公司也推行产品分类编码制度,美国凯斯公司实行零件分类编码制度(CACS)。近年,随着计算机技术的快速发展,农业统计信息分类编码标准化更受到越来越多国家的重视。美国农业信息机构已达700余个,以政府为主体,国家农业统计局、经济研究局、世界农业展望委员会、农业市场服务局及外国农业局等信息机构构建了国家、地区、州三级农业信息分类编码标准。在信息的收集规范上,美国农业部对每一种信息收集工作,统一编写出版工作手册,规定标准化内容。例如,基础农业种植生产信息包括作物面积、土地使用、牧畜存栏数、猪牛羊的产仔和死亡数、农产品库存以及农户数目等信息。1994年年底,日本形成政府主导的农业信息规范化建设,其所承载的信息主要有农业生产信息和农业市场信息,通过分类编码,建成农业信息服务全国联机网络,从而保证较高的农业信息利用率及共享性。

在国内,随着信息技术的发展,各行各业也制定了一些标准,有《国土基础信息数据分类与代码》(GB/T 13923—1992)、《全国主要产品分类与代码》(GB/T 7635.2—2002)、《林业资源分类与代码》(GB/T 14721.1—1993)、《粮食信息分类与编码》(LS/T 1702—2004)、《电子政务典型信息分类与编码标准》等标准,此外水利、图书、药品、档案、疾病等行业也制定了一些标准。农业各行业标准及信息技术分类标准的研究较为成熟。自1964年我国制定第一个农业种子国家标准以来,我国已累计制定并发布了农业国家标准1365项、农业行业标准3396项、农业地方标准8194项,涵盖种植业、养殖业、饲料、农机、再生能源、气象、水利和生态环境等领域(刘海凤和佟桂芝,2004)。截至2001年年底,我国公布并实施的信息技术类国家标准也已有765项,涉及计算机系统互联、信息分类及编码、数据通信、多媒体表示、电子数据交换、交易信息处理规则等信息共享方面(赵苹,2002)。

1.2 统计信息分类体系的构建

1.2.1 分类体系的构建原则

农村社会经济统计信息涉及的内容非常复杂,指标众多、结构复杂,要建立面向农村社会经济统计信息的分类体系,必须遵循科学性、实用性、系统性和可扩展性

的原则。

1. 科学性原则

农村社会经济统计信息分类与编码是农村社会经济统计信息的一个重要标准,必须坚持科学客观的编制原则。

在进行分类设置时,必须遵循农村社会经济的客观实际,进行广泛的调研,综合考虑分类对象、分类形式、应用环境等多种影响因素,选择统计信息相对稳定的本质属性或特征作为分类的基础和依据,使整个分类体系能覆盖农村社会经济领域,最大程度地符合当前我国信息化农村社会经济领域的需要。设计类目时,保证分类正确反映类目间的概念逻辑关系,并可保持相对稳定,从而使得不管载体如何变化,分类体系都能够充分准确定位统计信息,同时揭示各种信息之间的内在联系和相互关系。

2. 实用性原则

农村社会经济统计信息分类体系用于规范农村社会经济统计信息分类与编码方式,提高信息规划水平,需要得到普遍的推广,因此所确定的分类体系必须具有实用性高的特点,才能便于实践推广。

实用性原则就是要求分类体系的设置讲究精致全面、不能烦冗复杂,要精而有用。分类体系要简单、实用,类目名称尽量采用自然语言,明确其含义;类目设置应突出重点,方便浏览;类目级次不应太多,类目如果划分得太细,虽然知识的系统性强了,但用户的使用也比较困难,基本上以3、4级为宜。

3. 系统性原则

系统性原则要求将统计指标按一定的排列顺序加以系统化,从而形成合理的分类体系。

农村社会经济统计信息分类体系是一种多维的分类体系,揭示多维统计信息的联系,同时要保持系统化的整体感。设计分类体系时,建立若干摒弃从属关系却具有内在联系的子系统;而子系统的分类类目的划分是连续的、逐渐深入的、具有从属关系的,从而避免了跳跃式的展开。

4. 可扩展性原则

在实际应用中,遵照分类体系提供的扩展原则,利用规定的扩展方法,用户可根据实际情况对各分类进行类目扩充。

农村社会经济统计信息分类体系面向的是农村社会经济领域中的各种统计信息,因此必须考虑此种环境下信息更新的快速性、动态性等特点。诸如此类的可预见因素要求本标准所确定的分类体系也必须是一个动态的可扩展系统,能够提供灵活、有效的扩展机制,从而在资源发展到需要对分类体系进行扩展的情况下,用户可在符合标准的原则下,对分类体系进行合理的扩展。

1.2.2 分类体系的构建思路

信息分类是根据信息内容的属性和特征,将信息按一定的原则和方法进行区分和归类,并建立起一定的分类体系和排列顺序(古法辉等,2008)。信息分类的基本方法有线分类法、面分类法和混合分类法。线分类法也称为层级分类法,是以所选定的若干个属性或特征为分类的划分基础,将初始的分类对象逐次地分成相应的若干个层级的类目,并排成一个有层次的、逐级展开的分类体系(张茂震等,2005)。它的表现形式为大类、中类、小类等,将分类对象一层一层地具体划分,逐级展开,各个类之间构成并列或隶属关系。我国《国民经济行业分类与代码》(GB/T4754—1994)就采用了这种方法。面分类法是将所选定的分类对象的若干个属性或特征视为若干个“面”,每个“面”又可分成许多彼此独立的若干个类目。

实践证明线分类法和面分类法各有其优缺点。线分类方法具有线性单向式结构,体现的是纵向的概念逻辑关联,因此它的优点是层次好、简单实用,能较好地反映类目之间的逻辑关系。但由于其要求类目间具有严格的上下级隶属关系,不利于在复杂系统中应用,当分类层次较多时,会影响数据处理速度。面分类法能反映复合对象中各个主题因素及其联系,可识别性强、容纳性强,但对复杂对象中某一具体方面的层次关系反映不够。

农村社会经济统计信息分类体系根据农村社会经济统计信息的特点,确定层级分类的策略,对信息进行区分和归类。因此,采用线分类法将统计信息分门别类地、系统地组织起来以揭示层级关系。但鉴于线分类法与面分类法各自具有不同的优缺点,同时考虑统计信息多属性特征,必须综合利用这两种分类方法,进行有效的组合,才能达到多角度、多层次分类的目的。采用面分类法从多个方面反映信息属性特征,体现横向的概念逻辑关联,同时利用线分类法兼顾分类细化的纵向下移,体现分类的清晰度、精准度。

根据信息分类原则,结合《全国主要产品分类与代码》(GB/T7635—2002)、《国民经济行业分类》(GB/T 4754—2002) A门类以及农业生产结构特点,农村社会经济统计信息可分成4个等级:8个大类,39个中类,383个小类,900个细类(表1.1至表1.3)。在数据库结构设计中,每一层都留有足够的空间用于类别的扩充。

大类:是依照农村社会经济统计信息门类而建立的类别,依次分为农产品分类、农业生产、农业经济产值与经济效益、农业市场信息及乡镇社会经济发展情况信息等。

中类:是大类中按照生物学科和农业学科基于生产实践中种养的植物、动物划分成若干分支的类群,如种植业、林业、畜牧业、渔业等。

小类:是中类按照生物学科和农业学科分类法细分成若干分支的类群,如种植业下的粮食作物、豆类、薯类、油料等。

细类:是小类中细分的基本单位,也是各级单位的起点,即物种(简称“种”),如具有北方特色的谷物主要有玉米、谷子、高粱等。

表 1.1 分类的框架结构

层级	一级	二级	三级	四级
名称	大类	中类	小类	细类

表 1.2 一级类目

一级类	代码
农产品分类	A000000
农业生产	B000000
农村经济产值与经济效益	C000000
农业市场信息	D000000
乡镇社会经济发展情况信息	E000000
农业生产资料信息	F000000
农业科技信息	G000000
农产品批发	H000000
总计	8 类

表 1.3 详细类目

代码				类别名称
大类	中类	小类	细类	
A				农产品分类
	01			种植业
		01		粮食作物
			01	水稻
			02	玉米
			03	谷子
			04	高粱
			05	其他谷物
		02		豆类
			01	大豆
			02	绿豆
			03	豌豆
			04	红小豆
			05	蚕豆
⋮	⋮	⋮	⋮	⋮

分类完成后,按照定义好的分类原则进行分类论证,这个过程需要相关业务、技术、管理人员共同参与。论证方法可采用反推法。例如,分类方法是否存在重码现象、是否具有一定的灵活性、是否适合各业务运营和管理需要等。

最后,根据分类论证,需修改分类方案,对不全的信息进行补充,最终完成信息分类过程。