

建立水保基层单位规范化的数据文档文件和查询系统

——以南安市水土保持试验站为例

陈永宝¹, 黄传伟², 颜沧波³, 郭志民¹, 潘文聪¹, 陈志伟¹, 王维明⁴
(1. 南安市水土保持试验站, 福建 南安 362300; 2. 南安市水土保持办公室, 福建 南安 362300;
3. 泉州市水土保持办公室, 泉州 362000; 4. 福建省水土保持站, 福州 350003)

摘 要: 建立统一的数据文档文件是实现数据资源共享的基础工程之一。以南安市水土保持站为例, 详细地介绍了数据文档文件格式、数据编码, 不同条件下的查询系统和数据的安全性。
关键词: 数据文档文件; 数据编码; 数据查询系统; 数据安全性; 水土流失
中图分类号: S 157; T P311. 1 文献标识码: A 文章编号: 1005-3409(2003) 04-0320-03

Setting up a Standard Metadata Format and Query System

——A Case Study for Nanan Soil and Water Conservation Experiment Station

CHEN Yong-bao¹, HUANG Chuan-wei², YAN Cang-bo³,
GUO Zhi-min, PAN Wen-cong¹, CHEN Zhi-wei¹, WANG Wei-ming⁴
(1. *Soil and Water Conservation Experiment Station of Nanan*; 2. *Bureau of Soil
and Water Conservation of Nanan 362300, Fujian*; 3. *Bureau of Soil and Water Conservation
of Quanzhou, Fujian 362000, China*; 4. *Soil and Water Conservation Experiment
Station of Fujian Province, Fuzhou 350003, China*)

Abstract: It is a basic project of agricultural information to set up a uniform form of the metadata. The emphasis was put on the discussion of the metadata format, data coding, data security, and the query modes of different conditions, as well as on a idea, the design feature of agriculture information was from bottom organization to top organization.
Key words: metadata; data coding; query system; data security; soil and water loss

1 引 言

全国大部分市县都有水土保持站或水文站, 成立时间虽有先后, 但都积累了丰富的有关水土流失的科学试验数据。目前各站现有的数据资料, 尚未有统一的格式, 缺少详细的说明, 无法交流, 不能实现数据资源共享, 造成极大的浪费。南安市水土保持试验站成立于 1984 年 5 月, 迄今为止, 已积累了 40 余万个观测数据和科学试验数据。每年站上都组织人力对观测试验数据以及散落在书本、杂志、个人手中的数据进行搜集归类, 并编号装订成册存放于资料室。但存在如下问题: (1) 随着时间的推移, 有些数据记录纸已发黄、发脆, 极易损坏; 尤其是使用圆珠笔的记录, 油迹扩散, 有些字迹不易辨认; (2) 有些数据只有当事人才知道干什么用, 时间长

了, 有的连当事人也忘记了; (3) 数据没有统一的格式, 缺少详细的说明。 (4) 数据不能共享。一些专家学者已发出呼吁, 例如: 李智广、曾大林^[1]指出: “必须对每一个数据项定义标准的限定信息, 建立完整的、统一的文档。……应尽可能快地进行历史资料的整编和数据库的建立。” 吴礼福^[2]认为: “空间基础数据缺乏标准和规范, 一致性很难保证, 几乎没有共享。收集的空间数据, 仅仅只能为某一具体项目服务, 别的应用项目又必须重新收集空间数据。大量的重复劳动产生一堆零碎的没有规范性的数据。……”
综上所述, 南安市水土保持站的做法是: (1) 建立统一的数据文档格式; (2) 按照数据文档格式整理现存的数据资源; (3) 录入计算机; (4) 数据文件的编码和分类; (5) 建立基于数据库和网络的数据文件查询系统, 实现资源共享。

¹ 收稿日期: 2003-05-05
基金项目: 国家自然科学基金“监测水土流失的定量新方法”资助项目(40171060); 福建省自然科学基金资助项目(D0210020)。
作者简介: 陈永宝(1957 -), 男, 福建南安人, 南安市水土保持试验站高工, 从事水土保持工作。

2 数据文档文件

数据文件是建立多学科综合数据库的基础。数据文件包括两部分: 一是数据的文字说明部分(documentation), 即数据文档文件(metadata), 用于描述数据的背景、目的、测试方法、时间、地点、负责人等信息; 二是观测记录的二维数值表, 即数据文件的数据(数值)部分(data)。目前流行的商用数据库仅包括数据文件数据(数值)部分, 而缺少或淡化了数据文件的文档部分。美国著名的生态数据管理专家 Dr. Walt Conley 和他的学生美国 LTERN (Long Time Ecology Research Net) 数据管理专家 James W. Brunt 共同创立了生态系统数据管理中的数据文档文件格式(Walt Conley and James W. Brunt, 1992), 又称之为站际间文档交换文件(Inter-site Archive & Exchange File), 简称 Inter-site。Inter-site 是存储科学数据的一种格式和规范, 其存档文件结构通俗直观, 编辑方便灵活, 并以最常用的 ASCII 码文本格式存储, 能方便地从一种平台向另一种平台移植, 不受任何计算机操作系统和软件环境的限制。该数据文档文件得到不同学科的专家的共识和推荐(John B. Goorentz, 1992)。

国际长期生态研究网络 ILTER(International Long-term Ecological Research Network) 和中国生态研究网络 CERN (Chinese Ecological Research Network) 都采用了该格式^[3,4]。为了实现与国内外的大型网站进行数据交流, 南安市水土保持站也采用上述数据文档文件格式, 它的具体内容如下:

2.1 数据文档文件格式

2.1.1 文档日志

- 文档日志用于数据文件的建立、维护等有关信息。
- (1) 文档编写者: 数据采集者及文档编写人员。
 - (2) 文档录入者: 文档录入人员。
 - (3) 校对者: 文档校对人员。
 - (4) 数据负责人: 对本数据集的使用有决定作用的有关人员。
 - (5) 建档时间: 指建立文档的时间, 如果文档变更或修改, 必须加上更新时间。

2.1.2 文档说明

- (1) 数据集名称: 应根据数据集的内容、用途命名。
 - (2) 数据集编号: 用数字组合表示, 编号必须惟一。
 - (3) 数据来源: 指明是观测分析数据还是引用摘录数据。
 - (4) 摘要: 根据如下不同数据类型编写相应内容。
 - (5) 字段(变量)说明: 对字段(变量)的文字说明。
- 指标说明必须详细完整, 通用缩写名称用括弧标出, 名称单位要标准化, 如果和目前标准化不一致的请给出转换公式。

2.1.3 数据类型

此域用于记录数据类别。

内容: 观测数据、实验分析数据、统计数据、文字数据、图形图像数据。

2.2 数据文件

数据格式要与字段(数据变量)的说明一致。有些变量必须通过计算才能得到的, 请在数据观测项目说明中给出计算公式。

数据文件的格式建议采用电子表格式(xls)。字符型缺失值以 NA 表示, 数值型缺失值用满格 9 表示, 例如字段长度为 6, 小数点位数为 2, 则缺省数据以“999.99”表示。

2.3 存储格式

- 文档文件存储格式: .txt 纯文本文件格式
- 数据文件存储格式: .xls 格式
- 图形、图象数据存储格式: dxf, tif, bmp, jpg 等

3 数据文档文件的分类和编码

在数据管理中, 如果数据量很大, 经常采用编码方案。编码的一般原则是: (1) 编码的惟一性, 元素与其相应的编码一一对应; (2) 简单适用性, 编码尽量简单, 含义清晰; (3) 合理性, 代码层次清楚, 要与分类体系相适应; (4) 可扩充性, 编码应考虑到系统今后的发展; (5) 规范性, 尽量采用国家现有标准和其它行业标准。

数据编码的思路很多, 本文主要参考邮政、邮电系统、图书分类系统、身份证等的编码思想。例如邮政编码: 省、市、县……电话编码: 省、市、支局……图书分类: 从大类到小类。中国人信封地址的排列也是从大到小: 省、市、区、街道、门牌号。身份证号码共有 15 位数字, 前 6 位是持证人所属行政区域代码, 分别代表省、市、区(县); 第 7 位至 12 位是持证人的出生年月日; 第 13 位至 15 位是分配顺序代码, 第 15 位还表示性别。我们的编码思想也是如此。

3.1 数据分类

南安市水土保持站现存资料 40 余万个, 其来源可分为 4 类: 课题数据、常规观测数据、图形图像数据、资源环境及社会经济数据。

3.2 数据编码

- 根据全国行政区域编码, 福建(35)、泉州(05)、南安市(83)。每个数据文件名长度为 8 个字符, 第一第二个字符为“83”, 为南安市的代码, 第 3 个字符为数据分类符, 其余 5 个字符为数据文件的编号。如果区域在福建省内则数据文件名长度为 10 个字符, 在南安市“83”前加上泉州的代码“05”, 如在全国范围内, 在数据文件名前加上附件的代码“35”。例如,
- 83100001: 表示南安水土保持站课题数据文件, 编号为 1。
 - 83200121: 表示南安水土保持站观测数据文件, 编号为 121。
 - 83300105: 表示南安水土保持站图形图像数据文件, 编号为 105。
 - 83400019: 表示南安水土保持站资源环境及社会经济数据文件, 编号为 19。

3.3 数据文件总表格式

编号	数据文件名	文档文件名	数值文件名
83100001	1 号试验地土壤侵蚀量	83100001.doc	83100001.xls
.....
83200025	泉州市工程侵蚀调查数据	83200025.doc	83200025.xls
.....

4 数据查询系统

用统一的数据文档文件格式整理数据,按照数据文件的分类和编码规则编号,将文档和数值文件录入计算机。如果将其打印出来装订成册,这本身就是一部很好的专业数据查询工具书。根据上述 3.3 数据总表格式可知,条目编码、文档文件编码、数值文件编码是惟一对应的。人们可以根据分类和编码查询所需的数据文件条目,根据它的条目可以查到它的文档详细说明和相应的数值文件。这是一项基础性的工作,实现了字典式的查询和共享,但是速度慢,不方便。我们需要一个基于数据库和计算机网络的查询系统。

我们采用 Office 套装软件,它包括: Word, Excel, PowerPoint, ProntPage, Access。文档文件是文字说明,用 Word 输入和编排。数据文件是数据,使用 Excel 软件。用 ProntPage 编制网页,用 Access 建立数据库。根据需要用 PowerPoint 建立各种专题报告。Office 的各个软件都能直接上网, ProntPage 很容易把它们在網上组织起来,实现内部外部的查询和共享。文档文件是公开和共享的,促进基层水保部门与其它有关部门交流和合作。数据文件有些是公开的,有些是保密的,有些是需要花钱购买的。这就是数据的安全性。由于基层单位人力资源的差异和不同的配置就有不同查询系统方案,下面详细论述。

4.1 单机条件的查询系统

有些水土保持单位计算机少,或者没有连成局域网(内部网),计算机应用水平也不高,可以建立单机查询系统。

4.1.1 Prontpage 查询系统

有些单位不会使用数据库 Access 软件,怎么办?使用 Prontpage 这个软件,编制一个最简单的查询系统。Prontpage 这个软件容易学,可以从 Word 的格式直接转换过去。把本单位的数据归成几类,形成一级一级的菜单,超链接后,即是一个简洁明快的查询系统。缺点:不能按关键字或条件查询,不能直接在查询系经充上修改和增减数据。

参考文献:

[1] 李智广,曾大林.对水土保持监测网络数据管理的思考[J].资源生态环境网络研究动态,2001,12(1):13- 17.
[2] 吴礼福.基于 GIS 的防汛指挥系统[J].3S 世界,2001(1):7.
[3] 曹月华,赵士洞.世界环境与生态系统监测和研究网络[M].北京:科学出版社,1997.
[4] 中国生态系统研究网络科学委员会秘书处.中国生态系统研究网络数据目录[M].北京:气象出版社,1998.

数据的安全性:可以编制两套查询系统,一套供外部查询检索,一套是供内部使用。外部查询系统仅仅是文档文件查询,或且包括可以公开的数据文件。内部查询系统是在外部查询系统的基础上,通过文档文件连到相对应的数据文件上。

4.1.2 Access 查询系统

Access 是一个数据库软件,其数据文件可以直接从 Excel 转过来。它有窗体模块,能方便地建立漂亮的菜单界面。它有宏功能模块,可以设置各种方式的查询,例如,时间、关键字段等查询,还可以设立条件查询。它的最优秀的功能,根据数据产生各种图表,一目了然,供使用者决策。它可以修改和增减数据。

数据的安全性:可以编制两套查询系统,一套供外部查询检索,一套是供内部使用。外部查询系统仅仅是文档文件查询,或且包括可以公开的数据文件。内部查询系统除文档外,保括所有的数据文件。

4.2 网络查询系统

南安市水土保持试验站计算机人才量比较雄厚,有自己的网站,http:// www. nany. gov. cn。站内每台计算机都已连网。数据库软件使用 Access,文档文件和数据文件由 Access 管理。网页使用 ASP 软件。界面和菜单使用 ASP 编写,而不使用 Access 窗体模块。ASP 能方便地编写与 Access 的接口,完成各种咨询要求。

数据的安全性不需要用两套查询系统。用 ASP 编程界面的用户登录名和口令,凡有权共享数据的用户,经登录后,即可共享数据资源;反之,仅能浏览数据文档文件。

5 结 语

南安市水土保持站的做法值得其它基层水保部门参考:
(1)软硬件全部在微机上实现。硬件:PC 微机,即使目前一台高档微机的价格,基层水保单位也能接受。软件:使用目前普及和通用型的 Office 套装软件包(Word, Excel, Access)。财力、人力在最基层水保单位的允许范围内。在有条件的单位可采用 ASP 编程。

(2) 建立统一的数据文档文件格式,组织人力收集整编数据并录入计算机,保证了数据的永久性。

(3) 根据各基层水保单位的具体情况建立相应的查询系统。文档文件是公开和共享的,数据文件有些是保密的,设立数据的安全级别。