

王小宁, 夏巧利. 陕西省风和气温自记纸数字化建设成果 [J]. 陕西气象, 2015 (2): 35-37.

文章编号: 1006-4354 (2015) 02-0035-03

陕西省风和气温自记纸数字化建设成果

王小宁, 夏巧利

(陕西省气象信息中心, 西安 710014)

摘要: 介绍陕西省 EL 型电接风向风速自记纸和气温自记纸数字化预处理、扫描、质量检查、数据集制作等数字化建设过程及数字化成果数据集, 方便用户有效利用数字化成果。

关键词: 风向风速; 气温; 自记纸; 数字化成果

中图分类号: P416

文献标识码: B

自 2009 年以来, 陕西省开展了六期气象历史资料数字化建设工作, 对部分气象资料已经进行了图像扫描和录入等数字化处理^[1], 这些数字化成果在气象业务、科研和其它社会部门中得到广泛应用, 发挥了重要作用。随着气象业务和社会的发展, 对时间序列长、空间和时间分辨率更高的气象资料的需求, 特别是精细化预报服务和气象能源开发, 对逐小时风、温度等资料的需求尤为突出^[2-3]。

1 建设内容

按照中国气象局《气候变化应对决策支撑系统工程》、《山洪地质灾害防治气象保障工程》建设要求, 陕西省需要做好本省国家一般站 1951 年至自动站正式使用前一年 EL 型电接风自记图像扫描和质量检查工作; 完成本省国家基本

(准) 站和国家一般站 1951 年至自动站正式使用前一年的气温自记图像扫描和质量检查工作。

2 技术方案

2.1 总体思路

在中国气象局预报与网络司的统一组织和领导下开展项目工作, 由陕西省气象局组织, 陕西省气象信息中心承担, 完成陕西省地面气象站 EL 型电接风自记纸、气温自记纸的调研和数字化, 按照中国气象局项目组制定的统一技术方案和标准上报数字化成果。历史气象资料数字化工作采取外包数字化专业公司的形式组织。采用陕西省气象局和中国气象局两级质量审核, 陕西省气象信息中心负责对委托数字化专业公司的管理和监督并负责本省数字化成果的质量审核。受委托公司的实施方案和样品经陕西省信息中心审查

收稿日期: 2014-12-26

作者简介: 王小宁 (1965—), 女, 汉族, 陕西岐山人, 高级工程师, 从事气象资料应用与开发工作。

基金项目: 中国气象局重点建设项目《气候变化应对决策支撑系统工程》、《山洪地质灾害防治气象保障工程》

测数据与人工观测数据不存在明显的数学函数关系, 基本上没有相互订正的意义。

参考文献:

- [1] 中国气象局, 地面气象观测规范 [M]. 北京: 气象出版社, 2003: 17-20.
- [2] 世界气象组织. 气象仪器观测方法和指南 [M].

北京: 气象出版社, 1998: 186-201.

- [3] 张乐琪, 刘韬, 张桂梅. 中国北方大气气溶胶微物理特性研究 [J]. 陕西气象, 2014 (3): 12-16.
- [4] 沈宏彬, 宋静. 成都双流机场能见度气候特征及气象相关性分析 [J]. 成都信息工程学院学报, 2013, 28 (6): 672-675.

合格后方允许批量加工。

2.2 图像扫描及处理

2.2.1 资料借阅 所有记录纸扫描必须在陕西省气象局气象档案馆内进行,原始记录纸原则不得拿出档案馆。资料扫描严格按照档案管理规定办理借用手续,公司方要根据工作进度和保管条件定期分批借用,完成扫描后将资料恢复原装订,按照原顺序及时归还。陕西省气象档案管理人员及时检查,确保所借记录纸准确无误。

2.2.2 资料预处理 在扫描前,首先要对扫描的 EL 型风自记纸、气温自记纸进行预处理。按照“风向风速自记纸预处理技术规定”、“气温自记纸扫描技术规定”对需要扫描的风自记纸、气温自记纸进行处理,即按册对自记纸进行逐页清点,将装订成册的自记纸拆分为单页,检查自记纸是否标有省市名称、台站名称、年、月、日、换上换下时间是否完整,是否正确,有无缺测、备注等等。

2.2.3 扫描设备及参数 由于风、气温自记纸质较好,扫描设备选用柯达 i2600 扫描仪,扫描速度较快,影像清晰。采用“彩色扫描”方式,扫描分辨率不低于 200 dot/英寸(1 英寸=2.54 cm),扫描文件存储为 JPG 格式。

2.2.4 图像处理 为确保扫描图像清晰、可用,对扫描图片进行图像处理,避免重页、错页、漏页、扫描黑屏、页面不全等问题;应去除扫描过程中造成的图像杂点和脏点;进行图像纠偏处理,图像与水平位置夹角小于 1 度,以达到视觉上基本不感觉偏斜为准。

2.3 质量检查

2.3.1 图像文件质量检查 包括程序检查和人工检查两部分,程序检查主要检查扫描图像文件的完整性,检查是否存在漏扫、命名错误等问题;人工检查重点是检查图像清晰度、图像是否歪斜、颠倒、扭曲变形、图像黑边与污点处理、文件命名与内容是否一致等问题。

2.3.2 数据集质量检查报告 基于质量检查结

果,编写数据集质量检查报告,将图像文件质量检查错情及处理、数据集相关文档编制情况进行说明^[4]。

3 数字化成果——数据集

数字化的最终成果是数据集,按照中国气象局气象数据集制作技术规定和气象数据集组织命名规定,数据集包含五方面内容:数据集实体文件(datasets)、说明文档(description)、元数据文档(metadata)、附加文档(documents)、readme.txt 文件。一个完整的数据集由四级目录组成,具体命名规则如下(以气温为例):

SURF_CLI_xx_TEM_MIN_JPG (一级目录,数据集标识符),

datasets (二级目录,存放扫描图像文件),

IIiii (三级目录,台站标识符),

yyyy (四级目录,年标识符,存放各年自记纸图像文件,文件格式为 TIIiiyyyym₁m₁d₁d₁d₂d₂.JPG、TIIiiyyyym₁m₁d₁d₁d₂d₂A、JPG 等);

metadata (二级目录,存放元数据文件,文件名为 SURF_CLI_xx_TEM_MIN_JPG_META.doc);

description (二级目录,存放说明文档文件,文件名为 SURF_CLI_xx_TEM_MIN_JPG_DOCU.doc);

documents (二级目录,存放质检报告和台站信息文件等,台站信息文件名为 SURF_CLI_xx_TEM_MIN_JPG_STATION_IIiii.doc);

数据集实体文件是指扫描的图像文件,以区站号、年为单位,存放于 datasets 文件夹下。例如气温自记纸有 101 个站,datasets 文件夹下有以区站号命名的 101 个文件夹,每个区站号文件夹下,是以年份命名的文件夹,年份文件夹下是图像文件。数据集附加文档中包含扫描自记纸的台站信息文件(例如风向风速 SURF_CLI_SN_WIN_TMIN_JPG_STATION_

IIiii.doc), 内容有序号、区站号、纬度、经度、台站名称、自记纸年月、自记纸页数、备注文件数、自记纸换纸时间、备注等内容。数据集说明文档(例如风向风速 SURF _ CLI _ SN _ WIN _ TMIN _ JPG _ DOCU _ C.doc) 及元数据文档(SURF _ CLI _ SN _ WIN _ TMIN _ JPG _ META _ C.doc) 是对数据集实体文件的详细说明和解释, 便于用户对数据的使用。Readme.txt 文件, 是对数据集的名称、内容、数据大小、实体文件数量、数据集结构等的说明^[5]。文件命名中各项含义见表 1 和表 2, 数据集清单见表 3。

表 1 数据集命名的标识含义

标识	含义	备注
SURF _ CLI xx	数据类型 省区标识符	表示地面气候资料 陕西标识符为“SN”
WIN、TEM	要素代码	WIN 表示风, TEM 为气温
TMIN、MIN	观测时段代码	表示分钟级
JPG	数据形式	数据集为 JPG 文件

表 2 图像文件命名的标识含义

标识	含义	备注
F、T	要素指示码	F 为风向风速自记, T 为气温自记
IIiii	区站号	
yyyy	自记纸起始年份	
m ₁ m ₁ d ₁ d ₁	自记纸起始月份、日期	位数不足, 高位补“0”;
m ₂ m ₂ d ₂ d ₂	自记纸结束月份、日期	位数不足, 高位补“0”;
A	一日 2 张自记纸标识	第一张
B	一日 2 张自记纸标识	第二张(如有多张的, 从字母 D 往后顺延)
C	缺测标识	该文件为缺测空文件
BZ	备注标识	自记纸背面有备注时扫描的图像文件
JPG	文件后缀名	图片类型

表 3 陕西省风向风速、气温自记纸扫描图像文件数据集清单

数据集名称	台站数	资料年代	实体文件数/个	数据量/GB
陕西省 EL 型电接风向风速计自记纸扫描图像文件数据集	101	1956—2008 年	1 075 492	1 276
气温自记纸扫描图像文件数据集	101	1951—2007 年	1 503 077	1 110

注: 实体文件数(个)和数据量是指数据集实体目录(datasets)下的文件总数和大小。

4 结语

气象资料数字化建设是对以纸质报表和图纸形式保存的历史资料的有效拯救和保护, 数字化产生的信息化气象资料便于搜索、查找, 提高使用效率, 也可以建立气象要素长序列数据集, 提升陕西气象资料共享服务能力。

参考文献:

- [1] 张聪娥. 陕西省历史高空气象观测记录表数字化成果 [J]. 陕西气象, 2014 (2): 36-38.
- [2] 刘冬梅, 王钰, 黄少平. 江西历史地面气象报表

- 数字化常见问题分析 [J]. 江西科学, 2010, 28 (2): 269.
- [3] 刘熔熔, 封秀燕, 马仙妹, 等. 浙江省历史地面气象报表数字化处理和应用 [J]. 浙江气象, 2011, 32 (2): 34.
- [4] 文芳一, 谭斐, 王超球, 等. 广西区高空观测记录表数字化建设工作探析 [J]. 气象研究与应用, 2012, 33 (4): 78-79.
- [5] 王小宁, 阴秀菊, 汤浪. 陕西省气象科学共享数据集的制作及处理 [J]. 陕西气象, 2009 (4): 34-35.