

基于气象行业共享数据标准化处理的设计与实现

高婷¹ 陈晴¹ 孙晓燕¹ 徐海明¹ 杨明¹ 陆明²

(1 浙江省气象信息网络中心, 杭州 310017; 2 浙江省气象局, 杭州 310002)

摘要: 在数据融合与共享的大环境下, 气象部门与水利、交通、电力等气象相关行业依托共享的数据日益增多。行业气象观测站选址特殊, 站点密集, 对气象监测与预报起到了很好的补充与扩展作用。但由于目前并无标准规范化的行业间共享数据处理流程, 阻碍了各类数据在气象部门的应用。因此, 设计与实现气象相关行业间共享数据的标准化处理流程, 建立行业间共享数据的标准化数据集, 具有重要的意义。

关键词: 行业间, 共享数据, 标准化, 数据集

DOI: 10.3969/j.issn.2095-1973.2020.05.005

Design and Implementation of Standard Preprocessing Flow for The Sharing Data Between Industries Related to Meteorology

Gao Ting¹, Chen Qing¹, Sun Xiaoyan¹, Xu Haiming¹, Yang Ming¹, Lu Ming²

(1 Zhejiang Meteorological Information Network Center, Hangzhou 310017 2 Zhejiang Meteorological Bureau Office, Hangzhou 310002)

Abstract: With the environment of sharing and merging data, the meteorological department gets more data which comes from other industries, such as hydrology bureau, electricity bureau, traffic bureau. The location of the shared observation stations is special. Moreover, the numbers of stations are large. They play a positive role in meteorological monitor and forecast. Due to the lack of the standard preprocessing flow, the meteorological department cannot use the shared data. By designing and implementing of standard preprocessing flow for the sharing data between industries, we can establish the standardized data set coming from other departments. This work has important value. It can accelerate the application of shared data and provide the base indemnification for weather services and disaster prevention.

Keywords: other industries, shared data, standardization, data set

0 引言

近年来, 中国气象局提出“加快集约整合、推动气象大数据资源共建共享”的目标。各地气象部门积极推进本部门与交通、水利等气象相关行业的数据共享, 以实现跨行业的数据价值挖掘。目前浙江省气象局已获取来自环保、水利、交通、电力、海洋等气象相关行业的共享数据。行业的特种观测数据不仅可以补充气象站点数量的不足, 还可以扩展观测内容与种类, 对气象监测起到有益的补充与扩展作用, 充分利用多种数据源, 可以为气象预警预报与防灾减灾提供更高的保障。

由于行业气象数据标准各不相同, 导致各类数据在共享方式、数据格式、数据质量上均有较大的差异, 且与气象部门对观测数据的要求也有很大偏差, 难以直接应用在气象业务中。随着气象现代化信息化的不断发展, 国家及各级政府都在强调标准体系的建立及完善^[1], 气象标准化是气象现代化的表征之一^[2]。因此, 设计气象相关行业间共享数据的标准化处理流程, 实现从数据接收到数据存储共享的全流程规范化管理, 制作出行业间共享观测数据标准化数据集, 具有重要的意义, 可为加快推进多源信息在气象中的应用, 实现各类数据的融合提供基础与保障。对气象数据起到了很好的补充与扩展作用, 在气象服务中不可或缺。

1 数据介绍

目前浙江省气象局已获取海洋、交通等共计6个部门的共享数据, 具体资料情况汇总如表1。

收稿日期: 2019年7月22日; 修回日期: 2019年11月15日
第一作者: 高婷(1988—), Email: gt621_nuist@163.com
通信作者: 陈晴(1984—), Email: zjhjxlchenqing@163.com
资助信息: 浙江省气象局重点项目(2018ZD08)

表1 行业间共享数据汇总表
Table1 The summary of sharing data between industries related to meteorology

部门	站点个数	观测频率	共享形式	观测要素
水文局	约4340	1次/h 1次/10 min	数据库型数据	小时雨量、日雨量、水位、流量
电力局	约1200	不定时	数据库型数据	风向风速、降水量、气温、相对湿度、本站气压
交通局	约52	1次/h	文件型数据	气温、相对湿度、水汽压、降水量、风速风向、路面温度、能见度、路面冰雪层厚度、天气现象
环保局	142	1次/h、1次/d	文件型数据	AQI、PM _{2.5} 、PM ₁₀ 、CO、NO ₂ 、O ₃ 、SO ₂ 要素的小时数据和日统计数据
环境监测中心	137	1次/h	Websevice接口数据	负氧离子、臭氧、PM _{2.5} 的小时数据
海洋局	25	1次/h	文件型数据	气温、气压、湿度、降水、风向、风速、能见度

从表1可见，行业间数据种类丰富，观测要素中既有气象中常用的温压湿风等数据，又有水位、流量、AQI、NO₂等行业特色数据，大大扩展了气象观测数据的数据种类。分析表1，还可以看出不同来源的行业间数据在共享方式、观测频率、数据计量单位等方面均有差异与不同。

利用站点的经纬度信息，可获取气象相关行业间共享数据在空间的分布特征，其中交通站、电力站、水文站的空间分布如图1所示。

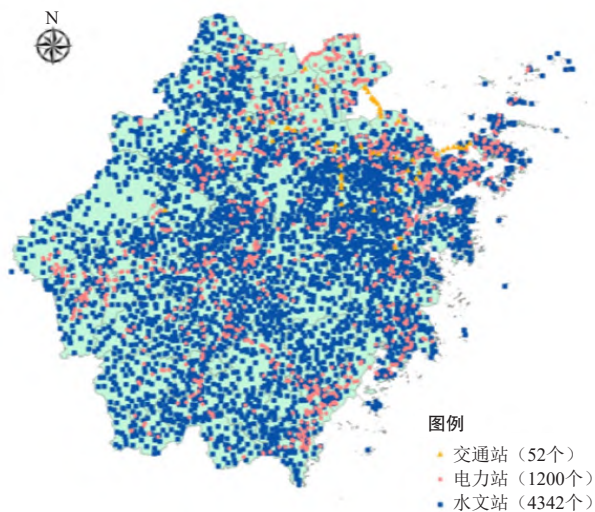


图1 目前浙江省行业间已共享数据站点空间分布
Fig. 1 Distribution of sharing weather stations between industries in Zhejiang Province

从图1中可以看出，行业间自动站在全省数量众多、覆盖密集，且所在位置与气象站互不重叠，充分利用行业间的数据，可以填补气象监测的盲点与空白。

2 设计原则与方法

行业间共享数据的处理流程设计应充分考虑以下三个方面：一是尽量保持原始观测数据的完整性，二是确保数据的可靠性，三是满足数据查询的便捷性。同时，应考虑在保留各类行业间特色数据的基础上，尽量将其中常用的气象观测要素与本部门的观测数据

进行统一。基于以上原则，初步设计了行业间共享数据标准化处理流程（图2）。

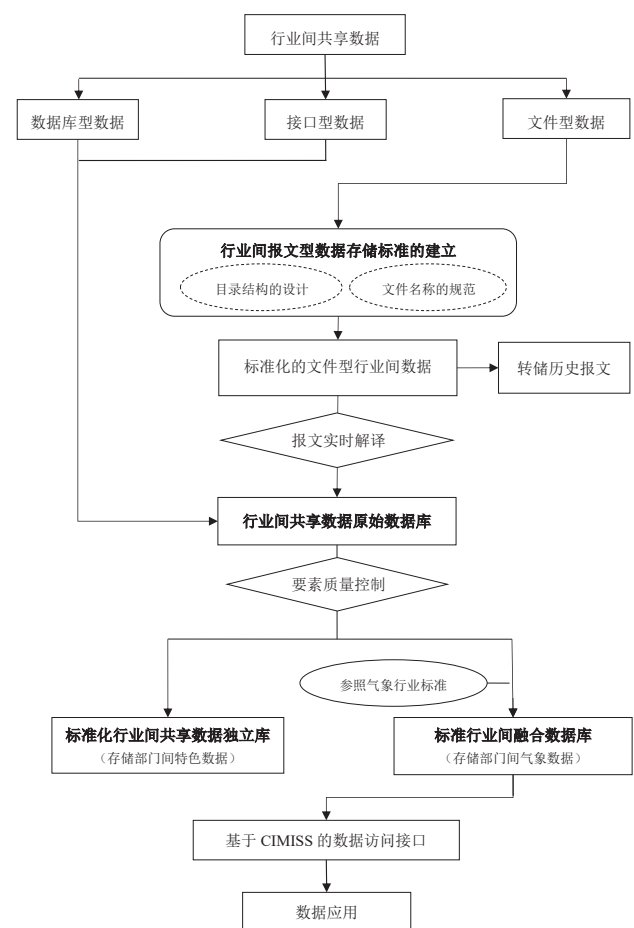


图2 行业间共享数据标准化处理流程设计方案
Fig. 2 The design flow of the standard preprocess for sharing data between

针对气象相关行业间共享数据标准化处理流程，主要内容有以下几个方面。

2.1 建立行业间报文型数据的存储标准

2.1.1 文件目录结构规范

为了实现行业间观测数据与气象数据的统一，考虑依据数据分类，以树形结构为存储目录结构。参

考中国气象局制定的《气象要素分类与编码》(QXT 102—2009),将数据分为16大类,其中第16大类,为“其他数据”,其解释为其他行业观测获得或收集加工的各种资料及相应数据产品。故可将行业间数据划分到此类中,作为目录的第一层次;其次,根据行业间数据的生产单位,进行目录第二层次的分类;若在以上两层次的基础上,数据还未能划分,则可根据数据的格式、时空属性等不同特点进行第三到第四层次的划分。

依据以上数据分类原则,结合各类行业间数据的实际特点,设计了浙江省行业间共享报文数据目录结构(表2)。

表2 行业间报文数据目录结构
Table 2 The directory structure for sharing data

数据名称	中文目录结构	实际目录结构
交通数据	其他数据\交通\	Other\traffic\
环保数据	其他数据\环保\小时数据 其他数据\环保\日数据	Other\environment\hour Other\environment\day
海洋数据	其他数据\海洋\	Other\ocean

2.1.2 文件命名的规范

参考《国家气象中心文件命名规范》《全国公共气象服务产品库文件命名规范》《气象数据传输文件命名》(QX/T 129—2011),并结合行业间数据的实际特点,将行业间观测文件命名格式规范如下:

`pflag_productidentifier_department_oflag_originator_yyyyMMddhhmmss_Ftype_Timer.type`

1) `pflag`字段:用来指示对后面`productidentifier`字段如何进行解码。目前`pflag`的取值均为Z,表示文件为国内交换的资料。

2) `productidentifier`字段:用来标识文件中数据的实际类型,当`pflag`为Z时,编码如表3。

表3 Productidentifier编码含义
Table 3 The schematic table for the meaning of productidentifier

类型名称	productidentifier	说明
地面气象资料	SURF	包括业务化运行的人工和地面观测台站等获得的资料
高空气象资料	UPAR	包括业务化运行的高空观测手段获得的高空气象探测资料
海洋气象资料	OCEN	包括海洋船舶等方式获得的海洋资料
气象服务产品	SEVP	直接应用于决策服务、公众服务的各类产品
其他资料	OTHE	指无法归并到上述资料内的气象资料和产品
文件交换系统相关信息	NOTES	业务通知、日志信息、系统之间交换信息等

3) `department`:用来指示数据生产的机构。

4) `oflag`:用来指示对后面的`originator`字段如何

进行编码。目前,`oflag`字段可以有以下允许的值:C、I、S,分别表示后面`originator`字段按四位编码(即CCCC码)进行解码、按目前台站区站号5位码或站名进行编码、按站号+子站号进行编码。

5) `originator`:不定长,用来代表文件的生成地点,一般可用台站号或台站名代替。

6) `yyyyMMddhhmmss`:文件生成时间“年月日时分秒”。

7) `Ftype`:代表文件的属性,常见取值有B(代表业务通知)、O(代表观测资料)、P(代表加工产品、反演资料、预报产品、服务产品)、C(代表测站站点信息)、R(代表各种统计信息、质量管理信息)、L(代表系统运行log文件)。由于行业间报文数据均为观测数据,故`Ftype`取固定值O。

8) `Timer`:指示是否为定时观测,定时观测用FTM、非定时观测用ITM表示。

9) `Type`字段:是一个可变长度的字段,用以标识文件的一般格式类型。

依据以上命名规范,将现有的行业间报文型数据命名如表4所列。

表4 行业间报文数据文件命名规范
Table 4 Standard name of sharing data between industries

数据名称	文件名示例
交通数据	Z_OTHE_traffic_I_JA001_20170724160000_O_FTM.txt (JA001为站号)
环保数据	Z_OTHE_environment_I_330200_20170724160000_O_FTM.txt (330200为站号)
海洋数据	Z_OTHE_ocean_I_鳌江_20170724160000_O_FTM.txt (鳌江为站名)

2.2 行业共享数据数据库的设计

由于报文型数据不方便用户读取,考虑以数据库形式存储共享行业间数据。依据各类行业间数据观测要素的不同,分别建立各行业间数据的独立数据库。数据库设计包括数据库名、表名、字段名的设计,字段类型、数据单位的规范,具体设计以《气象数据库存储管理命名》(QX/T233—2014)为标准。对于其中的气象要素,需保证其记录单位与本部门的一致。对于行业间的特色观测要素,结合行业应用为设计实施准则。

依据以上,将行业间数据库命名规范为:数据库分级代码_数据管理中心代码_数据分类代码_数据所属大类_数据生产单位标识代码。如行业间共享海洋数据库命名为:P_BEHZ_D_OTHE_oceanDB。

数据表的命名规范为:气象数据存储管理代码[_气象数据存储管理代码]_数据库对象代码[_数据库对象代码][_数据库对象扩展属性代码]。如海洋

小时数据表命名为: OTHE_OCEAN_BEHZ_MUL_HOR_TAB。

数据表中的相关字段采用V、Q、C和D等字母作为前缀,用以表示其数据类型(表5)。

表5 字段类型前缀说明

Table 5 The description for the prefix of field type

字段类型前缀	字段数据类型含义
V	LONG (整数型数据); NUMBER (小数型数据)
Q	相应V类型字段的质量控制段
C	VARCHAR (变长字符型数据); CHAR (定长字符型数据)
D	DATE (日期型数据)

对于其中的气象要素相关字段命名,气象要素代码由XXYYY[_ZZZZ]组成。XX为要素类型码,YYY为要素码,[_ZZZZ]是要素特殊码,为自由码。气象要素以外的其他要素字段采用帕斯卡(Pascal)命名规则执行。

2.3 基于行业间共享数据的接口设计与开发

在建立统一规范的数据环境的基础上,充分考虑

在多种环境下的适用性,基于全国综合气象信息共享平台(CIMISS)系统开发部门间共享数据的访问接口,实现部门间共享数据的集约、高效共享。

开发的数据访问接口提供全面的数据获取功能,包括:行业间站点资料的按站检索、按时间检索、按要素检索以及数据的按时间、按区域、按站统计等功能,并提供多种选择的返回格式。

3 结论

参照行业共享数据的标准化业务流程,将行业数据进行相应的规范化处理,实现了行业共享数据的标准化管理,并建设了存储行业共享数据的存储数据库,通过接口调用,实现行业间共享数据的实时访问。如图3为“浙江气象决策服务云平台”中接入的水文、交通、电力站数据的实时数据。从图3中可以看出,通过数据接口获取行业间的气象数据,实现了各类部门间资料与本部门数据的一张图显示,体现了多源信息在气象中的融合,保障了行业间数据的应用。



图3 浙江气象决策服务云平台中行业间数据的实时显示

Fig. 3 Display for the sharing data between industries in meteorological decision service platform for Zhejiang

本文表明,行业共享数据进行标准化的业务流程管理,符合中国气象局提出的“加快集约整合、推动气象大数据资源共建共享”的目标,有利于各类业务平台方便获取行业间资料,为行业间数据的应用提供了基础性保障。由于目前中国气象局还未颁发针对行业间共享数据的处理办法,此文中提出的标准化处

理流程可为各省、市气象部门提供参考,具有实用意义。

参考文献

- [1] 马锋波,周勇,刘东君.关于气象信息化标准体系框架编制的探讨.标准科学,2015(增刊):36-40.
- [2] 纪翠玲.气象标准制修订过程管理探究.厦门:第一届提升标准质量交流会,2014.