交通运输信息数据标准化及标准符合性检测技术

JIAOTONG YUNSHU XINXI SHUJU BIAOZHUNHUA
JI BIAOZHUN FUHEXING JIANCE JISHU

张绍阳　等◎著

人民交通出版社
China Communications Press

交通运输信息数据标准化及标准符合性检测技术

JIAOTONG YUNSHU XINXI SHUJU BIAOZHUNHUA
JI BIAOZHUN FUHEXING JIANCE JISHU

张绍阳　等◎著

人民交通出版社
China Communications Press

内 容 提 要

本书在对当前交通运输数据标准研究和编制现状介绍的基础上，对交通运输数据标准化的基本思想、数据标准规划和标准建设中存在的不足进行了分析，建立了数据标准的关系，最后对数据标准符合性检测理论、方法和系统开发等工作进行了全面介绍。

本书是交通运输信息数据标准符合性检测技术人员的培训教材，也可作为交通运输信息化设计和开发人员、交通运输信息数据标准编制人员以及相关信息化管理部门的参考资料。

图书在版编目(CIP)数据

交通运输信息数据标准化及标准符合性检测技术 /
张绍阳等著. —北京：人民交通出版社，2014.5
ISBN 978-7-114-11339-0

Ⅰ.①交… Ⅱ.①张… Ⅲ.①交通运输－数据检测－
标准化 Ⅳ.①U-65

中国版本图书馆 CIP 数据核字(2014)第 064595 号

书　　名：交通运输信息数据标准化及标准符合性检测技术
著 作 者：张绍阳　等
责任编辑：丁　遥
出版发行：人民交通出版社
地　　址：(100011)北京市朝阳区安定门外外馆斜街 3 号
网　　址：http://www.ccpress.com.cn
销售电话：(010)59757973
总 经 销：人民交通出版社发行部
经　　销：各地新华书店
印　　刷：北京市密东印刷有限公司
开　　本：720×960　1/16
印　　张：13.25
字　　数：230 千
版　　次：2014 年 5 月　第 1 版
印　　次：2014 年 5 月　第 1 次印刷
书　　号：ISBN 978-7-114-11339-0
定　　价：48.00 元
(有印刷、装订质量问题的图书，由本社负责调换)

前　言

QIANYAN

“十一五”时期，交通运输行业全面推进信息化建设，以示范、试点工程建设为依托，在交通运输动态信息采集与监控、交通信息资源整合开发与利用、交通运行综合分析辅助决策和交通信息服务四个方面取得了明显的成效。近年来，交通运输部确立了信息化在发展现代交通运输业中的战略地位和引领作用。作为发展现代交通运输业的根本途径、加快转变交通运输发展方式的重要支撑以及服务人民群众安全便捷出行的关键载体，交通运输信息化工作的重要性和关键性日益凸显。

“十二五”期间，交通运输信息化的主要工作思路是要实现从效率到效能、从分散到集约、从封闭到开放的三个转变，即从注重提高政府管理水平和工作效率转变为注重提升科学决策水平和公共服务能力及政务效能；从以往各自分散、独立的建设转变为集约化建设，大力推动整合共享项目建设，建立统一的通信信息基础网络、数据中心和应用支撑平台，倡导整合建设模式，避免重复建设，提高投资效益；从各自为政、相互封闭，转向注重顶层设计、加强总体规划、落实协同机制、实现资源共享的开放模式。其中，协同和共享是手段，集约和开放是方法，最终目标是提高交通运输行业的服务能力和决策水平。

要实现交通运输信息资源的集约和开放，交通运输信息数据的标准化是其中的关键。我国交通运输管理涉及公路、水路、铁路、航空、邮政等业务领域，业务面广、数据量大，并且我国区域经济发展水平不同，使得各省、各交通运输业务领域信息化发展也不均衡。随着交通运输信息化工作的深入开展，交通运输信息数据中存在的问题逐步暴露，阻碍了交通运输信息数据的共享交换及整合应用。为此，交通运输部发布了《交通信息基础数据元》(JT/T 697)等系列数据标准，编制了交通运输信息化标准体系表。经过多年的建设和宣贯，数据标准

不断充实，数据质量得到了一定的改善。但是，由于最初数据标准的编制大多是基于特定的业务系统等原因，数据标准总体规划不足、内在联系不足、标准贯彻缺乏监管等深层次问题逐渐显现出来。

为此，交通运输部启动了《交通运输信息数据与标准规范符合性检测关键技术及规范研究》（项目号：2011-364-812-50）、《交通信息化基础性标准研究（一期）》（项目号：2012-364-223-500）等项目，对交通运输信息基础性标准和标准符合性关键技术进行研究。本书作者是这两个项目的主要研究成员，书中大部分内容是这两个项目部分工作的总结。本书在介绍当前交通运输信息数据标准研究和编制现状的基础上，对交通运输信息数据标准化的基本思想、数据标准规划和标准建设中存在的不足进行了分析，建立了数据标准的关系，最后对数据标准符合性检测理论、方法和系统开发等工作进行了全面介绍。

全书由多位作者合作完成。其中，长安大学张绍阳主要编写了第1、3、4、6、7、8、10、14等章；曲卫东编写了第11、12章；曹金山参与了项目的大部分工作，并参与编写了第11、14章；安毅生编写了第9章；中国人民公安大学解源源编写了第2、5、13章；交通运输部科技司邹力、高翔参与编写了第6、10、11、13章；多位研究生参与了各章节的编写工作，其中中国电子集团第20研究所武伟参与编写了第13、14章，西安亿阳信通软件科技发展有限公司李男男、高航、王峰等参与编写了第6、7、12章，在读研究生李欣参与编写了第11章，张恒、关胜超参与编写了第8章，葛丽娟参与编写了第4、9章，刘静参与编写了第12章。本书中的数据采集部分由中国交通信息中心有限公司唐菁、刘昕、齐硕、辛文等人完成。全书由张绍阳统稿，交通运输部科技司邹力、高翔主审。

由于时间较紧且作者水平有限，书中难免存在错漏之处，请读者批评指正。

目 录

MULU

第一篇 交通运输信息数据标准化概述

第二篇　交通运输信息数据标准研究

第三篇　交通运输信息数据标准符合性检测方法和技术研究

第一篇

交通运输信息数据标准化概述

在交通运输行业信息化主管部门、标准编制人员以及信息化从业人员等的努力下，交通运输信息数据标准化已经取得了大量成果。

本篇对交通运输信息数据标准化的内涵、框架和发展方向等进行了系统的研究和总结。其中，第1章在对数据的基本概念、作用等分析的基础上，对数据标准化的内涵和意义、交通运输数据标准化方法等进行了论述；第2章对交通运输信息数据标准的编制和规划进行了总结；第3章从交通运输信息数据流程出发，提出了当前交通运输信息化发展阶段的数据标准化要求，形成了交通运输信息数据标准体系的整体初步框架；第4章对未来交通运输信息数据标准化的研究方向进行了讨论。

第1章 绪 论

1.1 数据及其作用

1.1.1 数据的定义

本体论认为,信息是事物运动的状态和状态变化方式的自我表述或者自我显示。人类认识世界和改造世界,首先必须实现对客观事物信息的获取和把握。电子计算机技术帮助人类实现了海量信息的获取和管理,大大扩展了人类在时间和空间两个维度对客观事物的感知和控制能力。但是,在电子计算机中,信息是无法直接进行存储的,信息被转换为"数据"的形式进行存储。因此,电子计算机中的数据被定义为承载客观事物信息、以电子信号形式进行存储和交换,并记录在磁、光或机械介质上的数字符号[1]。

1.1.2 数据是信息化的核心资源

通常所讲的信息化,就是利用信息技术手段,改变传统信息的获取、处理、传递、存储、利用的方式方法,从而实现对传统业务的改造和效率提升,甚至引领传统业务模式的变革。

在信息化过程中,会产生多种资源积累,包括软件资源、硬件资源和数据资源。软件资源实现了业务工作的信息化,提高了效率和效能;硬件资源是一种有形的信息化资产,是软件资源的载体;数据资源是信息化工作的核心资源,比其他两种资源更有生命力和价值。随着业务流程、目标和方法等的改变,以及信息技术的发展,软件需要不断地进行升级,因此,软件资源很快就会达到其寿命周期。随着集成电路技术的快速发展,硬件资源的淘汰速度也越来越快。1965 年,著名的摩尔定律就指出:当价格不变时,集成电路上可容纳的晶体管数目,约每隔 18 个月便会增加一倍,性能也将提升一倍。换言之,每一美元所能买到的电脑性能,将每隔 18 个月提升一倍。摩尔定律所阐述的趋势一直延续至今。可见,信息化过程中的软硬件资源随着时间的推移会不断淘汰和更新。

但是，数据资源作为信息的表现形式，永不过时，而且非常宝贵。广义上讲，数据资源所承载的信息是人类的财富。人类对文字出现以前的信息，只能从化石、遗迹片段等进行推断猜想，信息量非常有限。有了文字记载，人类可以更多地了解历史信息。现代计算机技术对人类生产、生活的各类信息进行了更加充分的记录，帮助子孙后代更完整地认识历史、以史为鉴、更好地生活。狭义上讲，数据资源是下一代信息系统的建设基础，能为信息系统提供基于时间维度的分析能力，是决策支持的重要信息来源。因此，数据资源是信息化工作的核心资源。我国已将信息数据所承载的信息资源提升到与能源、材料等同等重要的战略资源高度。

交通运输信息数据资源是指在交通建设、生产和管理过程中产生，并通过信息化手段形成的电子数据的集合，是交通运输信息的数字化表现。交通运输信息数据具有海量、结构复杂、增长迅速等特点。

发挥数据资源作用的关键是要对其进行充分利用。在现有计算机科学与技术的框架下，数据尚未能实现自由流转和任意应用，需要对其进行标准化。数据标准化对于实现数据资源的高效利用具有重要意义。

1.2 数据粒度

1.2.1 数据的产生及数据粒度

信息系统中的数据来源大致可以分为以下两种情况。第一种情况：数据产生自客观事物信息的数据化。客观事物信息的数据化过程就是将现实世界中的信息转化为计算机中存储数据的过程，通俗来讲就是由计算机应用系统自身产生的数据，这个过程一般都是通过应用系统专用的软硬件完成。为了能够对数据进行长期利用，一般要将数据在永久存储器上进行存储，防止断电、设备故障等原因造成数据的丢失。第二种情况：从其他系统交换而来。信息的互联互通，是信息化发挥综合效益的重要途径，因此，从其他系统交换而来是信息系统数据的另一个重要来源。

信息有粒度，数据也是有粒度的。从数据的产生过程可知，客观事物信息数据化时，即产生了单个的独立数据；同一类独立数据的集合，形成数据组织的最原始单位；对多个数据的有序组织，就形成了数据集。因此，从数据组织角度，可将数据分为三个粒度：独立数据、数据项、数据集。

1）独立数据

独立数据是指单个的、具体的客观事物的属性值。例如“张三”，该数据代表

某个人的姓名。独立数据具有数量庞大、分散的特点，不易管理。归类是人类学习和认知的一个基本方法。在数据管理中，一般也对数据进行归类。同一类独立数据的归类，就是数据项。

2）数据项

数据项是客观事物某个属性的标识及其内容的总称，也称为数据元素，可以理解为同一类数据的集合。数据项的定义即为该类数据的定义。例如，“姓名”代表了一个数据项，“张三”是该数据项的一个特定值，该值必须服从“姓名”数据项的定义。数据项将数据按类别进行有效组织，起到了提纲挈领的作用。数据项的定义包括类型、格式等，是计算机中对数据进行组织的最小单位。数据项的定义对独立数据形成了约束。在关系型数据库中，数据项与“字段”的定义相对应。

3）数据集

数据集是指有限数据项及其内容的集合。在交换中，一般是以数据集的方式进行交换。数据集是交换数据的集合，可大可小。数据集的属性包括其组织方式、内容、表示方式等方面。

这种分类方式和具体的数据管理系统的实现技术无关，符合人类认识事物从特殊到一般的过程。数据集的标准化和数据项的标准化具有不同的要求，本书将数据标准化分为数据项的标准化和数据集的标准化进行讨论，重点研究了数据项的标准化方法。

1.2.2 计算机中的数据层次及其粒度

在计算机中不同的技术层面，同一数据面向的主体不同。为了便于各主体对数据的理解，数据在各层面的表现形式不尽相同，如表1-1所示。

计算机中的数据层次 表1-1

数据层次	数据内容	面向主体	表现形式
应用层	完整的信息	用户	多个表示层数据共同构成一条完整的信息，例如“明明是男孩，今年15岁”
表示层	自然语言形式的片段信息，是客观事物属性的可理解形式	软件系统、开发人员	age = 15，sex = “男性”
逻辑层	二进制0/1形式	CPU、总线、存储器	00001111，00000001
物理层	光、电磁、机械信号	光、电磁识别装置等	光信号、电磁信号等

计算机系统帮助人类实现信息的存储和组织，经过复杂的转换过程，最终表现为人类可识别的符号。其中，物理层存储的光信号、电磁信号等通过光、电磁等识别装置被转换为二进制形式，二进制形式的数据可被计算机系统识别。二进制数据在编码方式、数据管理系统的约束下，转换为软件系统及开发人员所理解的表示形式，这类数据是一种自然语言形式的片段信息。多个片段信息的组合，便形成完整的应用信息。表 1-1 中的“表现形式”列展示了信息的转换和形成过程。

物理层数据的表示方法属于电磁学等领域的研究内容，技术标准也在不断发展，自计算机出现以来先后出现了纸带、磁盘、磁带、光盘等多类存储介质，每种介质的存储容量和标准也在不断提高。

逻辑层数据的表示方法是电子计算机的基础技术，经过多年的发展已经形成了较为完整的体系。Unicode 码的建立实现了跨平台、跨语言的信息编码方法，声音、图像、视频等信息的编码也逐步完善并形成标准体系。

表示层数据面向软件系统和开发人员，由软件系统在程序逻辑中理解和应用，具有面向业务应用的特点。例如，业务系统可能会根据年龄数据的大小进行干部选拔，也可能根据年龄进行离退休手续办理，或者进行公司员工年龄结构优化等，其数值是一种开发人员和应用系统都可以理解的表示方式。

应用层的数据表示方法是信息的应用模式，根据应用场合和需求，即信息获取者的要求而产生。例如，在表示层的几个独立信息，“年龄”、“姓名”、“性别”等，有的人关心某人的年龄信息，有的人关心其性别信息。

在以上几种表示中，物理层和逻辑层两类数据表示都属于计算机科学与技术、电磁学等相关学科技术领域研究范畴，和业务系统没有关联，经过多年的发展，已经形成了完善的标准。应用层数据表示根据应用场合的不同而不同，属于信息学的研究范畴。只有表示层的数据将业务信息和计算机数据进行关联，数据为软件系统所使用，和业务系统紧密相关。

与数据粒度的概念相对应，表示层数据属于数据项的范畴。表示层的数据是对现实世界信息的一种可理解的表示，是描述客观事物特征或性质的独立的信息。由于其面向开发者，因此，是进行数据标准化的关键和主要层次。

1.2.3 主流数据存储中的数据概念及其粒度

根据描述客观事物对象特性的不同，为了程序开发的便利性，一般将表示层数据分为数字、文字、图像、声音、视频等多种类型。

在计算机中，数据库是数据存储的一种主要组织方式。其中，关系型数据库是一种普遍应用的主流数据库。在关系型数据库中，所有数据都存储在关系（通

常所说的二维表)中,关系符合一定的范式要求。客观事物的同一特性的数据存储在二维表的一个列中,该列称为一个字段(或属性),字段的概念与数据项的概念相对应。每一行对应着一个客观事物对象,称为记录(或元组)。因此,在以关系型为主的数据存储中,数据项的标准化就是对关系表的列(即字段)定义的统一规定。

列的逻辑层编码方式可将其中存储的二进制数据转换为约定的、可识别的形式,这种编码方式便确定了一个字段的类型。数据库根据数据所表示的客观事物特性的不同,并考虑存储的方便性、占用空间的大小,对字段类型进行了详细的分类,大类上可分为数值型数据字段和非数值型数据字段。对于数值型数据字段,也有多种类型,例如 SQL Server 数据库的数值型数据的存储类型有 int、smallint、numeric 等,它们最主要的区别就是其所能表示的数据范围和精度不同;对于非数值型数据,数据库采用不同的类型进行存储,如 char——存储字符型数据、bit——存储真假二值型数据、Image——存储图像数据等。数据类型之间最主要的区别就是其运算方式不同,例如"长度"是一种数值,其可能的运算方式包括大小的比较、算术运算等,因此,适宜于用数值型方式存储;而"姓名"则一般不会使用大小比较、算术运算等操作,仅是一种可识别的符号,因此,适宜于使用非数值型(字符串)方式存储。

在数据库物理实现中,字段的主要描述属性包括类型和格式,这也是字段之间进行区别的主要属性。图 1-1 是 SQL Server 进行字段定义时的界面。

cname	nvarchar(600)	☐
ename	nvarchar(400)	☑
spell	nvarchar(600)	☑
field	nvarchar(400)	☑
num	nvarchar(400)	☑
version	nvarchar(400)	☑
department	nvarchar(400)	☑
circumstance	nvarchar(400)	☑
synonym	nvarchar(400)	☑

图 1-1 典型的字段定义内容

图 1-1 中,第一列是字段名称,是字段的唯一标识,用于进行字段之间的区别;第二列是字段类型 + 格式,是字段的主要定义内容;第三列是是否允许为空,是对字段中数据内容的填写进行规定。可见,第二列构成了一个字段定义的主要内容,也是字段所存储数据类别的主要表征。

1.2.4 数据交换中的数据概念及其粒度

信息的价值是在交换中体现的，因此，从其他系统交换数据是信息系统数据的一个重要来源。如前所述，关系型数据库是目前应用系统中主流的数据存储方式。下面首先对关系型数据库之间的数据交换过程进行分析。

在关系型数据库的交换过程中，通常采用系统内部的专用交换接口，或者是异构系统交换接口。在专用交换接口实现时，交换数据集可能是一个 RecordSet 对象、DataSet 对象、JTable 对象、一组自定义顺序的数据或数据结构等。其格式是交换双方约定好的，虽然效率较高，不需要转换，但无法在异构系统之间进行交换且无法实现非预定交换。通用的、异构系统之间进行数据交换时，主流采用基于 XML 的数据格式进行交换，这在业界已形成共识。

无论是哪种模式，其目的都是将 A 数据库中的数据传递到 B 数据库中，其技术关键是将所要交换的数据及其识别信息进行完整描述。这样才能实现非预定交换，实现交换的自由性。

在交换过程中，单个数据是最基本的交换内容。交换数据集就是要进行单个数据交换的组织。从通用角度出发，本书提出一种包括三个层次的交换数据集组织方式。第一层次，采用以行为主的组织方式。由于单个数据自身信息的不完整性，例如，数据库 A 中的字段 A112 中存储了人员的身高数据，需要传递到目标数据库中，但大多情况下仅传递身高数据到目标数据库中是无意义的，目标数据库不知道身高数据是谁的数据。因此，一般情况下，数据及其识别信息（数据表的主键）需要一起传递，如果将数据和其识别信息分离，在入库时将增加复杂性。因此，交换数据集的数据结构以记录（一组相互关联的数据）为组织更为合理，便于数据的识别。第二层次，以相互独立的交换单元容纳多组关系数据。一个表的多条记录能够构成一个内部循环的交换单元，由于各个表的结构不同，放在一起将增加数据解析的复杂性，因此，将多个单元相互独立，在一个交换数据集中就可容纳多个表的数据。第三层次，在交换数据集和交换单元中都增加描述信息，使数据集能够自我描述。这样，在一个交换数据集中便可自由、完整地表示出包含多个数据关系（表）、多个属性（字段）中任意数据（记录）的一次交换。关系型数据库的交换数据集的数据结构如图 1-2 所示。

特定系统内部的专用交换数据接口虽然使用其规定的格式组织交换数据，但在概念上也具有和图 1-2 类似的结构。例如 Microsoft 的 Dataset 对象，里面包含多个表及关系，存储为 XML 格式时，与图 1-2 中的交换数据集具有类似的结构。