

数据共享环境下统一信息标准的建设与应用

李学俭

(广东女子职业技术学院 网络中心, 广东 广州 511450)

摘要:数据共享可以解决应用系统互联互通的问题,共享数据中心的建设是实施数据共享的有效方案。共享数据中心是数据交换的中继链,也是综合信息的集市,实施的关键在于共享主题库的分类组织与上下行数据的准确转换,因此数据共享环境下统一信息标准的建设尤其重要。分析了实现信息标准的意义与作用,基于高等教育管理信息标准并结合实际探讨了信息标准体系的建设过程,将其应用于共享主题数据库的建设和数据中心的数据桥接与交换当中,具有很强的实用价值。

关键词:共享数据中心;数据分类;转换;信息标准

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2011)05-0242-04

Construction and Application of Uniform Information Standards Based on Data Sharing

LI Xue-jian

(Network Center, Guangdong Women's Polytechnic College, Guangzhou 511450, China)

Abstract: The problems of systems interoperability can be solved by data sharing, Shared Data Center (SDC) is an effective way of data sharing implementation. SDC is the relay chain of data exchange and also a comprehensive information pool, so the classification and accurate data conversion up and down is the key theme of SDC. The realization of the significance and role of information standards is analysed, the actual process of building standards system is discussed based on higher education management information. The application fields based on information standards are listed such as building the center databases pool of SDC and bridging and switching data, which makes SDC have a strong practical value.

Key words: SDC; data classification; conversion; information standards

0 引言

文献[1~5]从数据层面的角度,探讨采用数据共享方法解决应用系统互联互通的问题,然而文章仅从数据交换角度讨论数据交换的实现方案,忽略了建立完整且可执行的统一信息标准体系这一基本前提。事实上,在校内存在两大类的数据共享交换需求,一类是跨部门应用的公共环境数据,如学生基本数据、教工人事数据等,另一类是学校领导层关心的汇总统计数据。从校园信息建设的整体出发,共享数据中心的建设是系统地解决数据共享的有效途径。

共享数据中心平台,一方面担任数据中继的角色,将源应用系统的数据分类存放到对应的主题数据库并基于主题库将数据路由到定制数据的目标系统;另一方面作为校园的数据集市,为数据挖掘工作提供了现

实的数据基础,为智慧型校园的实施奠定了基础。

共享数据中心实施的重心在于共享主题库的分类组织与上下行数据的相互准确转换,因此在数据共享环境下统一信息标准的建设尤其重要。文中基于《CELTS-33_高等教育管理信息标准》并结合实际项目,从信息共享的角度研究与探讨高校统一信息标准的制定与实现方案,并就统一信息标准的拓展应用进行了论述。

1 统一信息标准的意义

统一信息标准是参照教育部就高等教育信息标准、国家通用标准及行业通用标准,综合校园内各类应用系统的数据(含业务数据及统计数据),建设一套权威的可行的信息标准集。实现校园统一信息标准,具有如下重要意义与作用:

(1)信息标准化虽不是一个具体的应用系统,但属于信息化校园的重要基础工作,是打通各个校园内各个信息点的桥梁与枢纽^[6];

收稿日期:2010-10-23;修回日期:2011-01-27

基金项目:广东省科技计划攻关项目(2007B010200036)

作者简介:李学俭(1978-),男,江西九江人,硕士,研究方向为校园数字化和信息管理。

(2)在国家相关信息标准的基础之上建立的信息标准库及基础数据库,并将应用系统建立在对基础数据库的引用之上,是有效解决信息共享的方法;

(3)确立信息标准,是信息一致性与完整性的有力措施,可以实现对现有系统的归并,实现旧数据的归纳与重用,容易找出信息化建设的薄弱点,指导信息化的深入开展,提升信息化建设水平;

(4)高校信息化的广泛开展,必然会引起应用系统之间的交叉引用,统一信息标准将解决在实际业务之中由于各应用系统信息存储及表示格式不一致而带来的种种问题^[7]。

2 统一信息标准制定与实施

信息标准的制定并不是一蹴而就的事情,相反由于它涉及到的应用系统及人员众多,甚至与各部门自身的利益相关,因而标准化的工作任务艰巨。作者所在单位在进行数据整合项目过程中,为统一信息标准的制定工作,形成了由院领导牵头、各业务部门配合及信息中心组织与实施的三级层级结构,加大了标准化工作的便利性。

在信息标准的制定与实施过程中,遵照如下几条规则:

- 1)涉及到国家或教育部相关标准时,严格遵照已颁发的标准,如数据项名称及取值,严格按照相关标准执行;
- 2)满足学校的管理需求,包括日常办公及信息管理的需求;
- 3)当数据管理存在多种信息标准时,根据面向对象的频率选择对应的信息标准,如资产信息在当前高校管理当中,存在国标《固定资产分类与代码》、教育部《高等学校固定资产分类与编码》以及科技部的《科学仪器设备分类编码表》等三套编码体系^[8],由于高校每年需要向教育厅提交高基表,向上汇总的主要对象是教育部,因而采用教育部的标准;
- 4)在涉及到的数据没有官方执行标准时,依照“谁主用,谁产生,谁维护”的原则,将数据源所在应用系统的数据规范做为该数据体系的标准,形成学校自定标准。

2.1 信息标准的层次

为了实现信息有效交换与存储,需要将数据进行分类。分类时,参照高等教育管理信息标准^[9,10],将数据集分为数据子集、数据类、数据子类以及数据项。数据子集为各类数据的归并档案,是数据类的集合。数据类是描述同一对象(业务环节)的相关数据项的集合,数据子类则是根据业务环节的需要,将数据类按对

象的独立性与相关性进行进一步划分。数据项则是具体的信息描述,是具有独立含义的最小标识单位,由编号、数据项名、中文简称、类型、长度、约束、值空间、解释/举例和引用编号等构成。信息的分类存放方法与主流的面向对象思想十分吻合,较好地反映出“信息—信息”以及“信息—信息集”之间的关系,图1是数据子集、数据类、数据子类以及数据项之间的关系图。

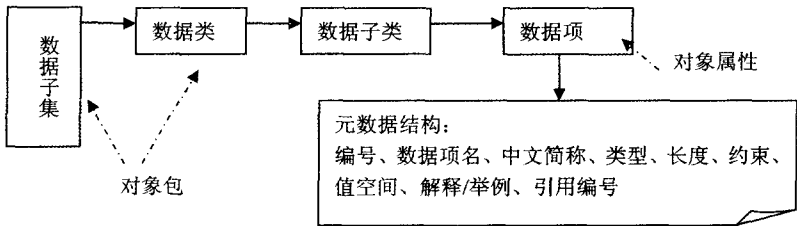


图1 信息标准的分类方法与面向对象的层级结构

按照以上信息结构可结合学校的主要业务活动,构建适合学校校情的数据集,信息标准工作小组可以从“人—财—物—管”等四个角度进行深度挖掘。涉及到“人”的角度,可以分别创建学生数据集、教工数据集;涉及到“财”的角度,可以创建财务管理数据集;涉及到“物”的角度,可以创建资产管理数据集;涉及到“管”的角度,可以创建教学管理、科研管理、办公管理等数据集。

2.2 代码集的建立

代码是制定与执行信息标准的重要基础,是信息标准中数据项的取值空间,严格定义了数据的格式及取值范围。高教管理信息标准中定义了四个代码子集,分别是国家标准代码子集(GB)、教育部标准代码子集(JB)、高教行业标准代码子集(HB)以及学校标准代码子集(XB),这样分类清晰,给各学校留有自定义编码的空间。

在实施代码库时,可以参照标准中的代码库,根据学校自身情况进行取舍,构建出适合学校应用的代码库,如人的性别代码、民族代码、政治面貌代码、学位代码等均有相关的国家标准代码,可以直接取用。而对于标准代码库中就同类数据的取值有冲突的情况,在选择代码库时,可参照第2节中规则3),根据数据应用的频率来定,如资产类数据代码一般选择高教行业标准代码子集。对于没有相关标准代码做参考时,可以根据正在使用的应用系统的编码体系制定出符合实际情况的代码集,如高职院校的部门代码可以根据学校的架构,采用各部门约定的规范进行编制^[11],如经济系代码为“01”,外语系为“02”,然后以此构建出学校内的专业代码,如“0101”表示物流专业,“0102”表示电子商务专业,由此可进一步构建班级及学生学号的编码规则(见表1)。

表 1 部门代码—专业代码—学号代码之间的关联

系部代码	系部名称	专业代码	专业名称	学号举例
01	经济系	0101	物流专业	200901010101
		0102	电子商务	200901020101
...				

2.3 信息标准的管理平台

信息标准的制定不是静态的工作过程,而是动态的管理过程,需要在实践中不断应用、不断修订、不断完善,即需要引入信息标准的动态管理机制。文中采用定制的管理平台^[12]实现信息标准的部署、实施与管理。管理平台包括三个主要功能模块,其一是数据子集的管理,包括数据子集的管理(新增、编辑、更新)、数据类的管理(新增、修改、更新)以及数据项的管理(新增、修改、删除);另一部分为代码集的管理,实现新增代码子集,新增代码;第三部分是标准比照模块,能够将现有业务系统的数据字典与标准库进行对照,构建出业务系统的标准评判体系,进而指导业务系统建设的规范化。

3 统一信息标准的应用

统一信息标准的建立,是信息交换的枢纽,对于全校的应用系统数据共享具有十分重要的作用。

3.1 建立全局共享数据库

基于统一信息标准,构建全校的共享数据库。文中在实现共享数据库时,依照标准创建了六大主题库,包括学生库、教工库、教学管理库、科研库、财务库、资产库,各个库的定义严格参照信息标准体系,即数据表的定义参照信息标准中的数据子类,表中的字段属性参照信息标准中的数据项,这是通过标准可实现的部分。然而标准付诸于实践的过程中,还有一部分不确定因素需要结合实际的业务才能确定,即标准的应用不能生搬硬套,需要灵活应变:

1) 共享库中数据表的主键约束需要根据实际情况来考量,如学生基本信息表中学号为主键,而在学生成绩表中,学号并不能做为唯一主键;

2) 数据项之间的引用与关联需要通过合适的外键约束予以实现,如科研主题中与教工号有关的属性

应与教工基本信息表中的教工号做外关联,即通过设置外键保证数据的完整性;

3) 数据项的值域与代码库的关联,只能通过第三方工具进行转换才能实现。

图 2 与图 3 是共享库与标准对应的设计界面,图 2 为共享数据库与后台实现对应的物理数据表,图 3 为学生库中学生基本信息的表结构,其中民族代码、性别代码等的取值均参照代码集中相关的标准,因而引用了对应的参照表。

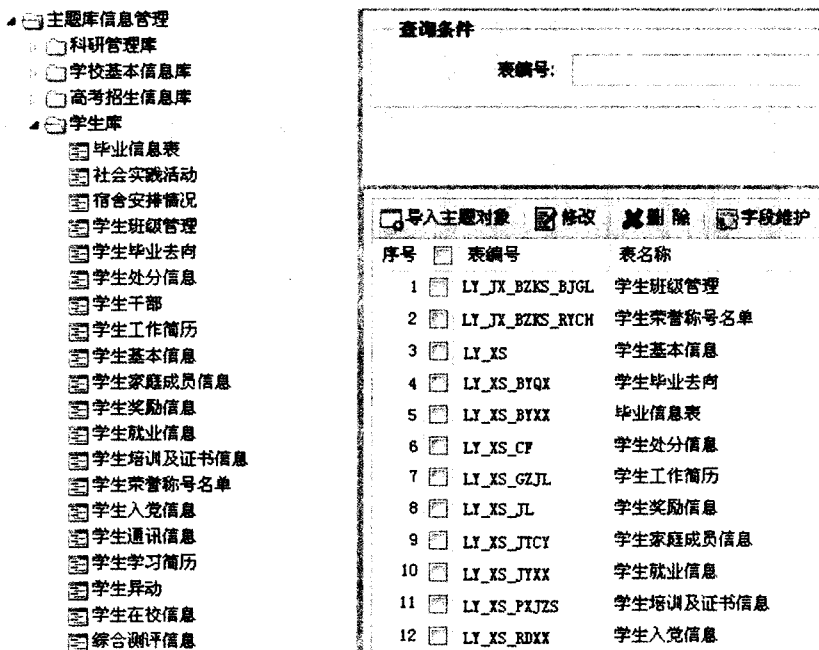


图 2 共享库与物理表的对应实现对照

中文简称	字段名称	类型	长度	是否主键	参照表名称	参照表中文名
学号	XH	C	20	1		
姓名	XM	C	60			
姓名拼音	XMPY	C	60			
曾用名	CYM	C	60			
民族代码	MZDM	C	2		LY_ZXBZ_MZ	民族代码
性别代码	XBDM	C	1		LY_ZXBZ_XB	人的性别代码
血型代码	XXDM	C	1		LY_ZXBZ_XX	血型代码
政治面貌代码	ZZMDDM	C	2		LY_ZXBZ_ZZMDDM	政治面貌代码
证件类型代码	ZJLXDM	C	2			
身份证件号	ZJH	C	60			
出生日期	CSRQ	D	7			
籍贯	JG	C	60			

图 3 学生主题库中学生基本信息与取值代码对照

3.2 基于信息标准的数据交换

如前所述,统一信息标准建立后,可以在共享数据库的基础之上基于信息标准实现应用系统之间的数据交换,优化跨部门的办公流程。数据交换时,需确保同一类型的数据只有一个源头,如按照现行的高校管理体制,学生的基本情况信息由学工系统产生,因此在全校信息交换流中,凡与学生基本情况相关的数据均采用学工系统提供的数据,这样以来保证了数据的权威

性,这是数据交换工作要解决的一个问题。除此之外,还必须保持各引用数据与数据源的一致性,即保持数据产生方与数据引用方的同步更新,这是数据交换工作要解决的另一个问题。

数据同步交换的基本思路是,先将数据产生方的数据抽取并按标准进行转换清洗之后存放至对应的主题库,引用方引用主题库中的数据并转换成对应的目标格式。如心理测评系统中需引用学工系统的学生基本信息(见表2),需将共享库中的学生基本信息(见表3)推向心理测评系统并进行格式转换,可以借助第三方的中间件工具予以实现,如可采用金蝶中间件公司的 AESB(金蝶企业服务总线)^[13]中间件解决方案,实现将数据从源端推向目标端并实现字段匹配与值转换。

表2 心理测试系统需要的学生基本信息

字段名	Code	Name	Sex	Nation	Native	Pe1	Pe2	...
中文名	学号	姓名	性别	民族	籍贯	院系	专业	...
举例	200902033104	张三	女	汉族	广东广州	外语系	商务日语	

表3 学生基本情况标准化格式

字段名	Xh	Xm	Xbdm	Mzdm	Jg	Yxdm	Zydm	...
中文名	学号	姓名	性别代码	民族代码	籍贯	院系代码	专业代码	...
举例	200902033104	张三	2	01	广东广州	02	0203	

4 结束语

高校信息化建设是一个不断发展的过程,在这个过程中,信息资源的有效整合是一个必然的过程,通过整合可以实现现有应用系统之间的数据交换与共享,而信息的标准化是数据交换的关键^[14]。然而由于各高校的具体情况各有不同,信息标准化在各校实践过程中存在许多不同的地方,需要结合实际情况进行标准的研制与应用,而这将必然优化高校信息化工作的成果,文中正是从这个立场研究了信息标准的制定与

应用过程,旨在探讨将标准付诸于实践的过程要领。

参考文献:

[1] 许鑫,苏新宁,吴乃冈. 高校共享数据中心平台的设计与实现[J]. 现代图书情报技术,2005(6):48-53.

[2] 张钦德. 一个面向 EAI 的数据交换平台的设计与实现[D]. 上海:上海交通大学,2007.

[3] 梅立军,付小龙,刘启新,等. 基于 SOA 的数据交换平台研究与实现[J]. 计算机工程与设计,2006,27(19):3601-3603.

[4] 谢晓祥. 基于 SOA 架构的学校应用整合[D]. 上海:华东师范大学,2007.

[5] 李学俭,何文华. 基于 SOA 架构的高校数据资源整合研究[J]. 计算机技术与发展,2010,20(1):78-81.

[6] 张计龙,陈翼,宓咏. 复旦大学管理信息标准研制实践[J]. 中国教育网络,2007(7):60-62.

[7] 李幼军,张广庆,刘炳兴. 高校信息化建设中信息标准的确立及应用[J]. 计算机时代,2008(10):66-68.

[8] 李润娥,张治坤. 高校信息标准系列文章之七—《资产设备数据子集》的研制[J]. 中国教育网络,2007(5):60-60.

[9] 沈锡臣,蒋东兴. 高校信息标准呼之欲出—《CELT-33 高等学校管理信息标准》[J]. 中国教育网络,2007(3):59-61.

[10] 全国信息技术标准化委员会. CELTS-33_高等教育管理信息标准[EB/OL]. 2008[2008-09-23]. <http://www.celtsc.edu.cn/>.

[11] 蒋磊宏. 高校信息标准系列文章之三—统一管理校内组织机构—标准校内单位基本数据类的产生[J]. 中国教育网络,2007(4):58-59.

[12] 广东女子学院共享数据中心平台. 信息标准管理[CP]. 2010[2010-04]. <http://sjzx.gdfs.edu.cn:8080/gdfs>.

[13] 金蝶中间件门户. APUSIC ESB 面向服务的企业总线架构[EB/OL]. 2008[2008-06]. <http://www.apusic.com/products/esb.faces>.

[14] 蒋东兴,佟秋利,蒋磊宏,等. 高等学校管理信息标准体系研究[J]. 中山大学学报(自然科学版),2009,48(3):56-62.

(上接第 237 页)

的研究[D]. 武汉:武汉理工大学,2008.

[10] 张艳花,钟一环,徐淑华. 电工电子实验教学中心的信息化管理[J]. 实验技术与管理,2009,4(26):106-107.

(上接第 241 页)

[7] 吕红,王凤琴. 计算机硬件技术基础实验教程[M]. 北京:兵器工业出版社,2007.

[8] 施德路,孟晨,于洪涛,等. 任务驱动式教学在计算机应用基础课程中的实验研究[J]. 教育探索,2009(7):26-27.

[9] 王岳斌,王惠宇,甘靖,等. 计算机技术基础“233”实验教

[11] 李万宝. ASP.NET 2.0 技术详解与应用实例[M]. 北京:兵器工业出版社,2007.

[12] 明日科技,张跃廷,王小科,等. ASP.NET 数据库系统开发案例精选[M]. 北京:人民邮电出版社,2007.

学体系研究[J]. 湖南理工学院学报,2009(9):82-84.

[10] 邹逢兴,陈立刚. 计算机硬件技术基础[M]. 北京:高等教育出版社,2005.

[11] Mazidi M A, Mazidi J G. The 80x86 IBM pc and compatible computers: Volumes I & II. Assembly language Design and Interface[M]. 3rd ed. 北京:清华大学出版社,2002.