Research and Application of Data Standard Construction

Haitian Jia[1, a] and Chun Jia[2, b]

[1] Suzhou Institute of Trade & Commerce, Suzhou 215009, China;
[2] Yellow River Conservancy Technical Institute, Kaifeng 475004,China
Email: [a]11804709@qq.com; [b]smy1161666@163.com

Abstract. The construction of intelligent campus data center is guided by the "13th Five-Year Plan" (2016-2020) of the College of Education, and insists on promoting the deep integration of information technology and education and teaching. This paper introduces the main theory and implementation method of data standard construction in the process of university data center building. Our school carries on the research and the exploration in the data standard construction, Each independent business system data is collected by data combing and data check, the data standard check and the data supplementary approval, the data cleaning and the conformity exchange, Establish shared data center, data management and quality assessment, design of service interface, Construction and display of data value model; This paper establishes standardized data by Oracle ODI and application system, and by the data exchange tool to build the data center, the standardization of the data platform construction management and maintenance.

1. Introduction

With the development of intelligent campus construction, a lot of business has been completed or is being built. Most systems are built on the basis of some business requirements, without considering the repetition of functions and data with other systems. The contradiction between data consistency and availability is obvious. Main performance aspects: (1) lack of standard specification for data requirements, resulting in multiple storage objects, different storage structures, serious impact on data sharing; (2) different data standards, resulting in statistical standard can't be matched; (3) business standard is not unified, Causes communication difficulty, cause the ambiguity.

The construction of the school data standard system is imminent , the construction of data standard can provide a constructive basis for the future big data application. JY/T 1006-2012 《Educational Management Information in Colleges and Universities》 has been promulgated for many years, and the data standard system of our university has been built in the last two years. The specific performance includes metadata (standard data) management, code standard management and master data management. Metadata management mainly refers to the standards of the Ministry of Education and the industry standards of the education industry, which mainly involve 11 data domains; the code standards mainly refer to the basic units of data in metadata management, such as gender, religion, etc. It is defined in the code standard. Master data management, mainly to achieve the definition of school standards, school level data standards define the main source and metadata ,code standards ,school custom standards. The development of data standards can be supported in business, technical and management aspects. Business can enhance the normative and enhance the data support for business analysis, through the data standards, the data information can be unified and consistent, so that the data can flow more easily between the business departments; first of all, technical aspects, the same structure of data, It is easier to share and exchange, second, the same data standards, reduce a lot of conversion,

cleaning, greatly improve the efficiency of data processing; In terms of management, data standards can provide complete, timely, accurate and high-quality data, provide support for decision support management.

According to two cases in the actual work, Through the original Human Resource system to complete the teaching staff basic information table data standardization, completes the data standardization of the student education system, and completes the construction of data standards in the actual work.

2. The data analysis

2.1 Personnel data and student data collection.

The school human resource system is an independent business system, there is an independent server system, the use of the Sqlserver database, through the original developer communication, the original system to open the user to the view interface.

School student management platform construction is a business system, there is an independent server system, the use of the Oracle database, through the original developer communication, the original system to open the user to the view interface.

2.2 metadata analysis.

Metadata is the data describing specific objects of information resources, and it can identify and manage the object, and realize the effective discovery and acquisition of information resources. Metadata definition comes from data country or industry standard. Schools need to update and maintain their own conditions.

The human resource standard data element T_JZG data consist of the individual basic data items specified by the staff.

The data element T_BZKS data of school students consist of basic information, student status information and source information of this college student.

The classification information of in the metadata standard includes: the general information of the school, the basic information of the personnel, the personnel management, the management of scientific research, the management of educational work, the management of educational affairs, the management of assets, the management of the party, the management of the organization of the party, the management of the United Front work, the management of the organization of the students, the management of foreign affairs, the administration of the office, the management of sports and health, the management of the archives, and the archives management. Alumni management and the unified identity management.

2.2 Target code analysis.

The target code is a detailed description of the part of the metadata. The target code and the data element definition are similar to the data units that are described by a series of attributes such as definition, identification, representation and permissible values, and It is the smallest data unit that can't be split in a particular semantic environment.

The human resource data standard data element relies on codes for political affiliation table and the Department corresponding table (the Department corresponding table is set as the school standard according to the situation). Because the original human resource system provides the limitation of the data, more other related target code can't be matched, and The new system can be defined and restricted in the future.

The student data standard data element depends on the national code table, the Department correspondence table, the class information table, the professional information table and so on. Due to the limitation of the original educational system, more other related target codes can't be matched, and the new system can be defined and restricted in the future.

2.3 Master data analysis

After the completion of the metadata analysis and the target code analysis, the school's main data standard can be built. This work is an extremely standard work and a long-term work. In principle, in order to make the data used and exchanged within and outside the school consistent and accurate, a normative document formulated by consensus and approved by the relevant competent authority for joint use and reuse is adopted. The main data specification is not only a set of specifications, but a system of management standards, control processes and technical tools. It is a process of gradually realizing information standardization by this system. Master data specification standardization is a set of complete data specification control processes and technical tools to ensure all kinds of important information, such as students, staff, institutions, teaching, and so on, The use and exchange within and outside the school are consistent and accurate. In addition, the main data specification standard is not only a matter of a department of technology or business, it is a unified specification for important business topics on the data level, but also the implementation of the business specification on the data level. The implementation of data standards depends on the consensus among business departments and the coordination between business and technology.

The master data specification consists of metadata specification、target code specification、school definition of data specification. According to different data domains, master data specification can be divided into basic class master data, analysis class master data and proprietary class master data. It is similar to the standard classification of metadata. The final goal of master data construction will directly determine the success or failure of data standard construction.

3. Data standardization implementation program

3.1 Implementation of key processes

ODI (Oracle Data Integrator) is a data integration tool provided by Oracle, which can efficiently extract, transform and load batch data. ODI can realize the integration of most mainstream relational databases (Oracle, DB2, SQL Server, MySQL, Sysbase).

ODI provides graphical client and agent runtime programs. Client software is mainly used for the design of the entire data integration service, including the creation of connection architecture for data sources, the creation of models and reverse table structures, the creation of interfaces, generation schemes and plans. The Agent running program is a service initiated by the command line on the ODI server, which executes periodically the execution plan under agent.

The construction of data standard in our university is based on Oracle ODI. The data standardization implementation mainly includes the following steps:

Topology manager,

(1)Create a data server and a physical architecture

(2)Creating a logical architecture

(3)Create an agent

Design and operate Designer and Operator:

Designer defines rules for data transformation and data consistency and data filtering conditions. Operator is mainly used for monitoring production data processing.

(1)Create a model

(2)Create a project

(3)Create an interface

(4)create package

(5)Generating plan, plan

The Agent, the Agent of ODI, is a JAVA service that can be used as a TCP/IP listening port, and the agent service includes some preset time schemes, and when the agent is running, it will be executed automatically according to the time and cycle set by the scheme.

Detailed operation is not described in detail.

3.2 human resource master data standardization implementation program

human resource master data standardization process mainly includes:

(1)Connect the original human resource system data V_JZGJBXX to the data center T_JZG_JBXX by ODI.

(2) The cleaning data is compared with the target data table, and the T_JZG standard table is written by ODI.
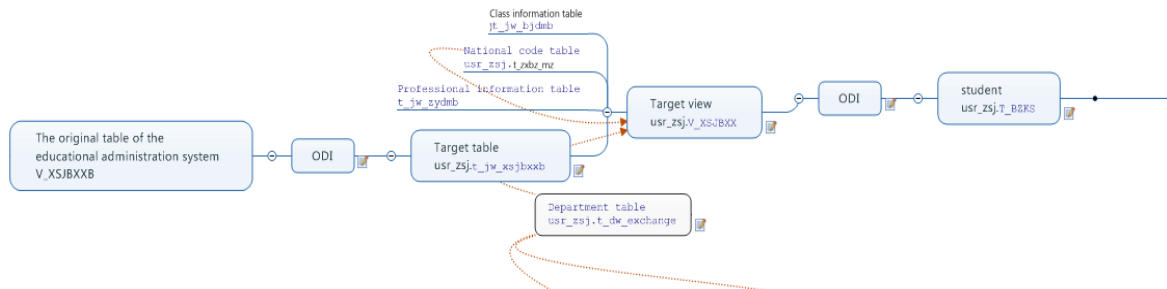
3.3 Student master data standardization implementation program

The standardization process of student master data mainly includes:

(1)by ODI, transform the original educational system data V_XSJBXXB to the data center T_JW_XSJBXXB;

(2) The cleaning data is compared with the target data table, and the V_XSJBXX standard table is written by ODI.

(3)Write view V_XSJBXX to T_BZKS standard table by ODI.
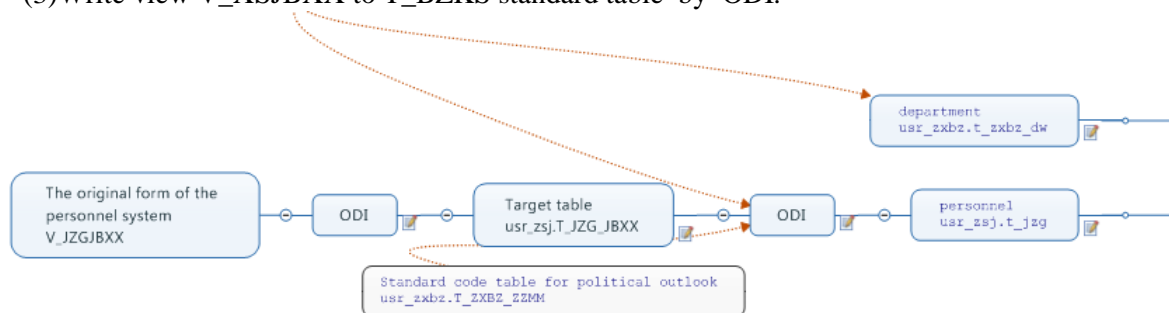
3.4 The definition of core data by data standards

First of all, standards are not models, but standards are core elements that can be achieve.

Secondly, considering the core data standard topic selection, we should consider it in many dimensions. Analyze the three aspects with business impact, system management and enforceability from data standards.

(1)In view of the degree of business impact, we can organize a variety of research activities such as centralized explanation, interview and questionnaire, and the number of problems involved, the number of problems affecting business, and the importance of problems affecting business.

(2)The application system association degree can be analyzed by the frequency of each department's attention, the use of each system and system module, and by combing the function of the application system, refining the related entities, and making the data theme of the related entities, and forming the distribution of the theme in the system.

(3)It can be implemented through the product manual and the business department system files to obtain the topic definition and classification, and the information item situation; analyze the data difference; obtain the degree of inconsistent degree of data definition and the difficulty of the integration of business rules.

According to the analysis, the number of business systems in each topic is different, the degree of concern is different, the degree of implementation is different (difference quantity, technology, etc.), and the final analysis chart of the theme selection is formed.

3.5 Data standards include both technical and business attributes
(1) the data standard is mainly aimed at business. The semantic of many schools depends very much on the manual combing of the business personnel. It is very difficult and inefficient. It is likely to cause the problem that the business semantics can't be found and managed in time because the carding personnel are not combing in time.
In the future, schools will face digital transformation. From unstructured documents, most of the business semantics are extracted and managed together to become the future development trend. This ability can be realized by natural language analysis technology. The school can analyze the most of the same business through the description of the same business in a comprehensive number of materials. The new and most widely recognized business definition, identified by the business staff, identifies the business semantics, which greatly reduces the workload of the business staff and improves the enthusiasm of the business staff to sort out the business semantics.
(2) in the school data management, any data standard, if there is no corresponding technical ways, will be difficult to achieve, so when the school establishes the data standard, it needs to join the English name of the information item to correspond to the field in the actual database table. Adding English names to information items in data standards can bring two benefits to school data governance:
When making model design, standards can be integrated directly with model design tools, and standards can be directly quoted when designing models.
For the existing system, the standard can be directly associated with the related fields of the application system by English names, and automatically discover and incompatible fields and notify the corresponding system directly by metadata.
(3) technology and business information are required in the standard, and effective association is needed to make it effective. For the school data management, the premise of technology to understand the business is that there is a correspondence between technology and business. This correspondence can't be done by a large number of manual combing. Otherwise, the burden of business department is very heavy and the enthusiasm is not high. It is necessary to use the technical means to accumulate the industry practice of the data management tool provider, to form an automatic Association Library of business and technology, and to automatically complete the correspondence between business and technology. It will greatly reduce the workload of the business personnel, improve the accuracy of the connection between technology and business, and eliminate the gap between business and technology.
3.6 **Data standards need to be updated continuously**
For the school data management, there are many data standards established, often only a set of books, not according to the development of the school business in time to make updates, a long time to be a set, in fact, the data standards need to be constantly revised with the changes in the school business, such as the new business in the school. It is necessary to increase the corresponding standards and to discard the standard of no value in time. Only in this way can we ensure that data standards have always been able to adapt the needs of business development and promote standard achievement.

4. Conclusion
Some school data standards are completed, only in the booklets and books, the lack of realized ways, can't be effectively implemented; in addition, the data standard itself is lack of management, can't effectively adapt to the development of new business.
The idea of school data management is focused on data standard formulation, and the requirements of data management in various fields are integrated into the system, and from the source of demand writing and demand analysis. In strict accordance with the requirements of data standards, the whole software life cycle is connected with the metadata management of data quality management. In this

process, the data standards are constantly verified and revised, so that the data standards can always adapt the development needs of new businesses.

5. Acknowledgement

6. References

[1]. WANG Jingchun. Digital Campus Data Integration Architecture Analysis [J]. Journal of Changchun University of Science and Technology(Natural Science Edition). 2015(03) p. 148-151

[2]. Xuan Zu Guang, Wang Xiaoxia, Xu Xiaohui, Zhang Hai. Application of ODI in Massive Online Learning System Management [J]. Journal of Zhejiang Wanli University. 2015(02) p. 79-81

[3]. FENG She-miao. Study on Location Model of Aviation Logistics Transfer Center [J]. Railway Transport and Economy. 2015(01) p. 45-47

[4]. Tammaro A M, Ross S, Casarosa V. Research Data Curator: The Competencies Gap［J］. BOBCATSSS 2014 Proceedings, 2014, 1 (1): p.95 －100

[5]. Kouper I. CLIR/DLF Digital Curation Postdoctoral Fellowship — The Hybrid Role of Data Curator ［J］. Bulletin of the American Society for Information Science and Technology, 2013, 39 (2): p. 46 －47.