

***Executive Summary***

***TBR ERP Data Integrity Task Force***

***April 6, 2005***

## ***Introduction***

The TBR Data Integrity Task Force was formed to review standards and guidelines and to help make decisions regarding maintenance of data in the administrative systems. This task force has been charged to develop model policies, enact model procedures, and recommend priorities for utilization of resources used to support institutional data management systems. The membership is a cross-section of technical resources from two-year and four-year institutions and the SMO.

## ***Recommendations***

The Task Force recommends that the TBR as a governing body and each two-year and four-year institution:

- 1. Establish a data standards committee whose purpose is to establish policies to assure the accuracy and consistency of data within all systems.**

For the TBR the recommended minimal membership is:

- TBR Chief Information Officer (Ex Officio)
- TBR Associate Vice Chancellor for Research and Assessment
- Three members of the existing Data Integrity Task Force to provide continuity and technical representation
- Nine members representing a cross-section of the two and four year institutions, the various functional areas, and the geographic sections of the state. These members should be selected from the Data Standards committees established at the individual institutions.

For the individual institutions, the recommended minimal membership should include the areas listed below. (Individual positions and designation of the chairperson will vary depending on local organizational structure.) Areas not included in the ERP system such as the Library and One Card system should be considered for this committee to ensure effective flow of information outside the ERP system.

- Institutional Research
- Internal Audit
- Alumni/Development
- Business Office (Budget and Bursar)
- Human Resources
- Financial Aid
- Admissions
- Records
- Information Technology

Sample documents regarding data quality policy, data correction, and data custodians are included as Appendix A.

**2. Identify important data collections and systems (shadow systems).**

A shadow system is a collection of data outside the ERP system that is used for a business function. These systems receive data from the legacy system either through a direct feed, or manually. Examples of such systems are Access databases containing data extracted from one of the current systems which may then have additional information added for manipulation and report generation. There may also be spreadsheets with current enrollment data used for projections for class section creation for future terms. The data in these systems could be used for only a short period or could be maintained over time. Other examples are feeds to library patron databases or “one card” systems. A document with more detail regarding important data collections and systems is included as Appendix B.

**3. Identify exceptional data elements.**

Exceptional data elements are those data which should be identified and examined independently of and prior to the migration of data to the ERP system. Specific examples of such elements are ones which exist in more than one of the current systems but will be merged into a single element in Banner. Other examples are data elements which have been used by an institution for local purposes or elements which are being used for a purpose other than the one originally intended. A more detailed document regarding exceptional data elements is included as Appendix C.

**4. Begin scrubbing data in legacy systems.**

Through the years, inconsistent data input has caused errors in the legacy systems. Regardless of which cohort an institution is in, all colleges and universities should begin the process of scrubbing (cleansing) the data in their legacy systems. Functional users and IT staff should work together to identify and correct problems. Examples of cleansing data include checking for and eliminating duplicate IDs in each of the current systems; execution of programs which purge data files such as billing detail or applicants who never enrolled; review of duplicate or obsolete vendor information; and review of date information for possible correction.

In addition, the Data Integrity Task Force recommends the Data Standards Committee formed by the TBR address the following:

- Coordinate all data standards and issues from other TBR and institutional process teams
- Adopt a formal data warehouse strategy to insure accurate and timely reporting across the TBR
- Adopt and publish data standards and valid values for all data elements necessary for TBR reporting to ensure consistent reporting
- Establish an automated methodology to streamline the flow of data from the campuses to TBR to reduce manual intervention as much as possible

## ***Appendix A***

**<Institution Name>  
Policy on Data Quality and Data Correction**

**Background and Purpose:** The effectiveness of information systems for **<Institution Name>** is directly related to the quality of data within those systems. The new systems planned for **<Institution Name>** will require the migration of data from legacy systems and it is the intent of this policy to assure the quality and correctness of our data. The current environment of isolated systems has led to questionable quality of data which has, in turn, led to invalid data and confusion in the utilization of information at **<Institution Name>**

Therefore:

**Policy:** To assure the accuracy of data within all systems at **<Institution Name>** it is our policy to:

1. Make all data corrections to the data within the appropriate data source system.
2. Assign a data custodian to be responsible for each data element.
3. Identify, in the case of multiple systems using the same data, a system of “official record” - that is, a system in which data will originate and will “feed” other systems requiring the data element.
4. Establish a process insuring the accuracy and quality of data in all systems.

The Data Administrator, working with the Data Custodians, is assigned the responsibility of assuring that this policy is followed and for providing a process for data correction.

**<Institution Name>  
Description of Duties for Data Custodians**

In recognizing the strategic value of data, <Institution Name> has established the responsibilities of Data Custodians as guardians of these corporate assets. Assignment as a Data Custodian is an important recognition that an individual has achieved the experience and business acumen needed to assure <Institution Names'> community that our data are well-managed and able to support all of our information needs.

The duties of a Data Custodian follow:

1. Maintain membership in and attend meetings of the <Institution Name>Data Custodians.
2. Work with the <college/university> Data Administrator to establish standards for data management such as a *data element naming standard* and a *data element definition standard*. Maintain the official *data element abbreviations list*.
3. Provide business definitions for <college/university> data.
4. Be certain that the data elements assigned to your care are properly understood and used appropriately by the <college/university> community by assuring that data definitions:
  - a. are based on actual usage
  - b. follow <college/university> standards
5. Periodically audit the elements for accuracy and integrity and take corrective actions whenever necessary while adhering to <college/university> policies.
6. Review data capture and update processes to assure that:
  - a. Data capture is done once at the business event of its origin.
  - b. Data collection is automated whenever possible.
  - c. Business processes are designed to ensure data quality.
  - d. Data update procedures are consistent across units authorized to update data.

7. Review all data models and recommend changes to the <college/university> Data Administrator.
8. Act as liaison to the <college/university>community in publicizing and seeking community input on any anticipated changes to data elements.
9. Be responsible for data access requests and for assuring that <college/university>data policies are followed.
10. Plan future data needs.



## ***Appendix B***

## **Exceptional Data Elements**

SunGard SCT is providing a set of tools to help move data from Plus to Banner. This will handle the vast majority of the data. However, there is a group of data that must be considered in this conversion that is not or may not be handled by these tools. These exceptional data elements fall in the five classifications and must be considered in the conversion effort.

### **1. Merging Data Elements**

First and most obvious are the data elements that are contained in more than one Plus system that will be merged into one Banner data element. This includes such elements as name, address, sex, marital status and many more. Most of these elements will be considered in General Person planning. Each Plus system may have different formats and values for these elements. A common format and set of valid values must be developed with a conversion table for each system to the new values.

### **2. Institutional Added Elements**

The second type is the data elements that the institution has added to the Plus system. There may be corresponding data elements in Banner to receive this data. Each of these elements must be identified and matched to the equivalent Banner element if one exists.

(Example 1 below is a sample spreadsheet for documenting this type of element.)

### **3. Institutional Abused, Commandeered, or Hijacked Elements**

The third type is the hardest to identify. This is the data element that is being used for some purpose other than the purpose specified in the design of the system. In some cases only the office using that portion of the data may know about this usage. Probably every institution is using the SIS field Mother Tongue for something other than the original language spoken by the student. Some of these will be relatively easy to identify because the DBD values have been changed. In other cases no DBD values are assigned or the values happen to match the usage.

Another type of abused data element is those elements that are not entered or maintained by the custodian office but are needed by another office at the institution. An example could be ACT scores in SIS. Admissions does not need ACT score for transfer students and does not collect or enter it but Institutional Research needs it. Some of these may be required data elements in converting to Banner.

(Example 2 below is a sample spreadsheet for documenting this type of element.)

### **4. User Defined Elements**

The fourth type is the user defined elements in Plus such as the “report flags” and other elements that are defined within the Plus system to give the institution flexibility on data usage. These elements must be defined along with the usage and corresponding element in Banner.

## **5. Subsidiary System Data Elements**

The fifth is closely aligned with the second but is data that is stored in files outside the Plus umbrella that could be added to Banner. There are areas where new functionality in Banner will replace subsidiary systems that institutions have developed. This data must be identified and matched with the appropriate Banner data element.

With all of these types of exceptional data elements there is the possibility that the usage has changed over time and the element may have a different set of values for different periods of time. This is a relatively common occurrence in the Financial Aid System. This situation will be difficult to handle in the data conversion.

(Example 3 below is a sample spreadsheet for documenting this type of element.

## Example 1

Example of Institutional Added Elements

ITD AISS  
FRS Conversion Data Mapping – MTSU Fields

DBD element	COBOL name	Description	Type	Length	Pic	New DBD element name
FG040	FG-DESC-CODE	Part of User Filler in DBD	alpha-num	2	X(02)	FG-DESC-CODE
FG294	FG-USER-RESERVED-4	Part of User Filler in DBD	alpha-num	35	X(35)	MTSU AGENCY
FGU06	FG-USER-RESERVED-4	Part of User Filler in DBD	alpha-num	3	X(03)	FED-AGENCY-CODE
FGU07	FG-USER-RESERVED-4	Part of User Filler in DBD	alpha-num	1	X(01)	RESEARCH-CODE
FGU08	FG-USER-RESERVED-4	Part of User Filler in DBD	alpha-num	1	X(01)	FEDERAL-CODE
FS040	FS-DESC-CODE	Part of User Filler in DBD	alpha-num	2	X(02)	

## Example 2

### Example of Institutional Abused, Commandeered or Hijacked Elements

#### ITD – AISS FRS Conversion Data Mapping – Cannibalized Fields

DBD element	SCT COBOL name	SCT Description	Type	Pic	MTSU DBD element	MTSU COBOL name	MTSU description	Type	Pic
FG250	FG-DSGN	DESIGNATION	alpha-num	X(1)	FG250	FG-DSGN	COST REIMB	alpha-num	x(01)
FG258	FG-SPONSOR	SPONSOR	alpha-num	X(04)	FGU05	FG-SPONSOR	CONTRACT NUMBER	alpha-num	X(20)
FG260	FG-SPONSOR-AWARD-NO	SPONSOR AWARD #	alpha-num	X(04)		FG-SPONSOR-AWARD-NO			
FG262	FG-FEDERAL-STATE-ID	FEDERAL/STATE ID#	alpha-num	X(12)		FG-FEDERAL-STATE-ID			
FG278	FG-TECH-RPT-DATE-8	TECH REPORT DATE	alpha-num	X(08)	FG278	FG-TECH-RPT-DATE-8	QTR 2 BILL DATE	numeric	9(08)
FG280	FG-FISCAL-RPT-DATE-8	FISCAL REPORT DATE	alpha-num	X(08)	FG280	FG-FISCAL-RPT-DATE-8	QTR 1 BILL DATE	numeric	9(08)
FG282	FG-INVENT-DATE-8	INVENT REPORT DATE	alpha-num	X(08)	FG282	FG-INVENT-DATE-8	QTR 3 BILL DATE	numeric	9(08)
FG284	FG-RENEWAL-DATE-8	RENEWAL REPORT DATE	alpha-num	X(08)	FG284	FG-RENEWAL-DATE-8	QTR 4 BILL DATE	numeric	9(08)

### Example 3

Example of subsidiary system or data files

ITD – AISS  
FRS Conversion Data Mapping – MTSU Files

File name	Description	Fields (COBOL name)	Type	Pic	File Total length
GSE-TABLE-FILE					
	GSE-REC				
	KEY	GSE-REC-KEY		PIC X(10)	
	HI SUBCODE	GSE-HI-SUBCODE		PIC X(04)	
	GSE VALID SWITCH	GSE-VALID-SW		PIC X(01)	
	FILLER	FILLER		PIC X(02)	
					17
IND-TABLE					
	MS-PAGE				
	ACCOUNT NAME	MS-ACCT-NAME		PIC X(30)	
	BR SUBSIDIARY	MS-BR-SUBSIDIARY		PIC X	
	FRS SUBSIDIARY	MS-FRS-SUBSIDIARY		PIC X	
	FEED OTHER	MS-FEED-OTHER		PIC X	
	FILLER	FILLER		PIC X(51)	
					84

## ***Appendix C***

## **Important Data Collections and Systems (Shadow Systems)**

This recommendation is to help establish procedures for identifying, cataloging and coping with important data collections and systems that are sometimes referred to as shadow systems.

The original definition of a shadow system is: “a system that replicates information and processes of a centralized system outside the control of that central system.” Over time the definition has evolved to include other types of auxiliary data stores and systems. For simplicity this document will refer to all of these important data collections and systems as shadow systems.

A shadow system is a collection of data outside the ERP system that is used for some business function. It can be a spreadsheet, Access database or a Word document or a complete application system. The key factors in defining shadow systems are:

1. Information that is contained in the ERP system is stored outside the ERP system database.
2. The data is used in some business function.
3. The data is retained beyond initial extraction.
4. The system depends on an identifying index that is generated in the ERP system such as the personal identification number.

Shadow systems can be classified into three groups:

### **a) Quasi-shadow**

This would include downloaded information from the ERP system to a spreadsheet, database, word document or any other storage media where the data is manipulated, printed and deleted. This is not a true shadow system but should be cataloged for migration purposes because:

- a. The process may be replaced or not needed with the new system.
- b. Data formats, location and extraction methods will change with the new system.

Example: Downloading data from the central system to a spreadsheet, adding additional information and manipulating the data for a report, and then discarding the spreadsheet when the report is printed.

### **b) Real shadow system.**

This includes down loading information from the ERP to a spreadsheet, database or document, manipulating the data, adding additional information, using it in a business process, and then retaining the data over time. Issues involved:

- a. With the data being retained it can become out of sync with the ERP data and no longer accurate.

- b. Syncing with the ERP data.
- c. In the conversion to the new ERP system data formats, extraction methods and data will change.

Example: Downloading student grades at the end of each semester to monitor the progress through a program of study.

**c) Auxiliary shadow system.**

This includes complete applications that interface with the ERP system through extracted data or common data and indexes that are not part of the actual ERP system. This would include such systems as Parking pass and ticket system, institution “one card” systems, library patron, etc. These systems may be third party with little flexibility in changing data formats to match the ERP system.

Example: Campus ID card system, Library Automation, NCAA compliance, parking registration and traffic ticket, WebCT, etc.

It is very important that all of these important data collections and systems be cataloged and analyzed as part of the ERP implementation process. Some of these may be mission critical and their failure could cause serious damage to the mission of the institution or an office. **With the implementation of SunGard SCT Banner most if not all interfaces to shadow systems will become non operational.** The use of data warehouses for the extracted or interface data for both SunGard SCT Plus and Banner may help this situation.

**The recommended information to be collected includes:**

1. Department name
2. Contact person
3. Departmental databases
4. Hardware and software used
5. System administrator
6. Primary purpose of the system
7. List and description of data in the system duplicated from the ERP system and additional data elements
8. How long has the system been in use?
9. Who uses the system?
10. What reports are generated by the system and who receives the reports?
11. Who modifies or maintains the system when needed?