

MT. SAN ANTONIO
COLLEGE
1100 N. GRAND AVE.
WALNUT, CA 91789
(909) 594-5611
RESEARCH@MTSAC.EDU

Research Tips

FEBRUARY 2009

RESEARCH & INSTITUTIONAL EFFECTIVENESS OFFICE

- Barbara McNeice-Stallard,
MSc, Director
- Priyadarshini Chaplot,
M.B.A., Educational
Research Assessment
Analyst
- Lisa DiDonato, M.A.,
Project Manager, Basic
Skills Projects
- Odette Richardson, M.A.,
Research Assistant
- Cathy Stute, M.A.,
Graduate Research
Assistant, Basic Skills
Projects
- Chiou Yueh "Maria" Tsai,
Ph.D., Research Analyst
- Jennifer Tucker, Ph.D.,
Educational Research
Assessment Analyst

We're on the Web!

[http://www.mtsac.edu/
administration/research/](http://www.mtsac.edu/administration/research/)



Keeping a Clean Local Database

By

Chiou Yueh "Maria" Tsai, Lisa DiDonato, and Odette Richardson

As Mt. San Antonio College moves closer to full implementation of the BANNER information management system, and because of database capabilities within BANNER, creation of a standalone local database is discouraged. (See details in the NOTE section at the end of this article.) However, if you absolutely must create one, then this article gives you a great starting point.

All too often database elements in some local databases are set up in a way that is not conducive to efficient use of the data. Thus, "cleaning" (detecting errors in the data and correcting them) must occur before meaningful analysis can be conducted. You may have heard the adage, "Good data in means good data out." This is most certainly the case with any database management. Three typical errors we have found in some local databases are: not collecting enough data and/or the right data; selecting an inappropriate method of housing the data; and using multiple ways to code the same data. Consequently, here are the primary areas to consider in creating your local or shadow database.

► What is the purpose of a database?

The data collected in a database should be essential and unique to your program so that it can be used to track a special population or to meet external accrediting requirements, accountability reports, or program review assessment needs.

Tip 1: Before implementing a database, make yourself a list of what you want to be able to know from the data you are collecting. Meet with others involved in the process and ask them, "What do we want to know, show, or prove by collecting this data?"

Tip 2: Review the questions or list regarding what you want to find out and decide what data items would best answer your questions.

Tip 3: There is no need to keep duplicate data that can be extracted from the Mt. SAC Data Warehouse (e.g., birth date); however, there must be a *common* identifier with a matched data type and format in order to link your local database with Mt. SAC's centralized database. Only use the Mt. SAC student ID number as the *common* identifier.

► How should the database be designed?

A well designed database helps eliminate problems and ensure efficiency in data extraction. A good practice is to generate a list of all the critical information to be collected and logically group them by subjects. A table will hold a group of data on one subject (e.g., personal data) to avoid storing duplicate data and keeping tables in manageable sizes. Each field (column) within a table will hold only one small piece of data on the subject, and each row (line) contains a *unique* record of the subject; no two rows of data should be identical. Normally, one field or a set of fields in a table is designated to *uniquely* identify each row. This field is called a "primary key." For example, a student ID will be the "primary key" or *common* identifier in identifying

(continued on the back)

Keeping a Clean Local Database (continued)

each student record stored in a row while a class reference number is used at Mt. SAC as a “primary key” to identifying a specific class session of a course.

- Tip 1:** A database should be designed with such a structure that allows for easy entry, access, update, and building upon. If you plan to use a spreadsheet format such as Excel to collect your data, keeping it clean will present more challenges than using a database program such as Access. Whatever software you choose, keep in mind that you or someone else will probably want to sort the data. If the program you select can't do that, it is not a good database software choice.
- Tip 2:** Store and group the data you want to collect by subjects in different tables. Keep data in each field to a minimum (e.g., last name, first name, city, zip code).
- Tip 3:** Predetermine data type, format, length, value ranges in each field (column) and have a “primary key” field in each table that matches with the same field in the Mt. SAC Data Warehouse.
- Tip 4:** If codes are being used, descriptions of these codes need to be documented and stored as a “data dictionary” to provide reference for the assigned meaning.

► What data integrity and security mechanisms have been incorporated?

ACCURACY and CONSISTENCY are two fundamental principles of data integrity. Without them, data won't be valid and reliable to use. Setting up a routine procedure in checking data integrity is important and essential for “clean” data. Contact RIE for more information.

DATA SECURITY is another important issue that needs to be strongly emphasized. Installing an encryption¹ procedure to protect sensitive data in your computer is advised. Access to sensitive data should be restricted and no password sharing should be permitted. For protection, sensitive data should be hand-delivered instead of being electronically transmitted. It is also necessary to take precaution in protecting databases from accidental data loss. Routine backup and some recovery strategy should be planned to avoid such a disaster.

- Tip 1:** Set clear rules on accuracy and consistency for data entry. If you have established a rule that says all city names (example below) will be fully spelled out, then Line 2 of your city data must be corrected. If another rule says there will be no spaces in the course ID prefix, then Line 3 of your course ID data must be corrected. Remember, if you want to successfully sort your data elements, keeping data in common data formats and values will provide you with “clean” data results.

1	ID	LastName	FirstName	City	CourseID
2	123456789	Mountie	Joe	Rowland Hgts	ENGL1A
3	000123456	Mountie	Josie	Rowland Heights	MATH 50

Please Note: Remember to store your data rules in your data dictionary for others to follow.

- Tip 2:** Consider installing an encryption procedure to protect sensitive or restrictive data in your computer by contacting the IT Department. Do not electronically transmit sensitive data via e-mail. Instead, hand-deliver the data via CD or flash drive.

The Research and Institutional Effectiveness Office is here to serve you and provide assistance with your data needs. Contact us with your local database questions.

NOTE: Mt. San Antonio College is moving toward full implementation of the SCT BANNER system. The new information management system will have some capability of maintaining program specific databases. Individual programs that use their own local databases are encouraged to contact IT to set up/incorporate them in the BANNER system. BANNER will also have a reporting tool that allows designated users to select/generate ready-to-use reports. More information about the new BANNER system can be accessed through <http://banner.mtsac.edu/> from a college computer on campus.

¹ Encryption is a process which is applied to text messages or other important data, and alters it to make it humanly unreadable except by someone who knows how to decrypt it. ... *Source:* <http://cplus.about.com/od/glossar1/g/encryption.htm> as of October 10, 2008.