



## **Data, Data Everywhere, How Do You Manage It?**

Gerald A. Wilson, Ph.D.

January, 2009

OmniSoft, Inc.

Carlsbad, CA

Schools have always had a large number of records about their students, but until the 21<sup>st</sup> century, these were mostly paper, not electronic records. The motivations for moving from paper to electronic records vary, but usually include saving file-cabinet space, making them easier to search and maintain, and reducing the cost of record keeping. In addition to the direct educational uses of this data (report cards, transcripts, health records), in recent years there has been a substantial growth in the number of state and federal reporting requirements also using these records. Schools have had to develop or purchase special Student Information Systems (SIS) and Data Information Systems (DIS) to have the means to comply with all the reporting requirements without further growth of the administrative staff. While the direct costs of purchasing and maintaining these systems impact the budget, it is the hidden costs that are the greatest burden.

The hidden costs for SIS and DIS systems are the teacher and administrator hours for entering, checking, and correcting the data, for administering the data archives, as well as the general supervision of maintaining these systems. In fact, over the past decade there has been a tremendous growth of both the quantity of data gathered and the number of different groups examining that data. Adding to these burdens are new legal requirements which are intended, with good reason, for protecting the information from exposure to un-authorized individuals and organizations. These new legal burdens are further complicated by the recent connections of SIS and DIS systems to the world-wide web to allow parents and students to get information from teachers more quickly and reliably. To understand the nature of these hidden costs and how they are impacting schools we need to start by examining the type of data being collected and maintained.

Students and teachers alike know the day-to-day classroom items that are maintained, such as course materials, test results, grades and student progress reports. Added to this are results from standardized tests, diagnostic tests, special workups and extra-curricular activities. Administrative data includes names, home addresses, social security numbers, attendance records, and health records (legally mandated vaccinations, disease reports, etc.). All this data is maintained by the SIS and DIS systems, which also need to have consistent data (e.g. a student's name and student id must be the same in both systems in order to find the correct).

Probably the trend with the greatest impact is that of everyone wanting to know everything immediately. Parents want to know the latest grades and any problems within hours. Teachers and administrators must keep records with regard to everything from vaccination records to attendance records to discipline actions. Various outside groups and agencies want to know progress, school ranking, special needs, standardized test results, and a variety of other statistics. In general, everyone wants to know everything as soon as it happens.

## **A Bit of History**

Teachers and administrators have been keeping extensive records of all things related to schools and students on paper records for decades. File cabinets of such records were in classrooms, school offices, and even warehouses. Industries were created to produce all types of education oriented record-keeping devices. Academic calendars (August to July instead of January to December), grade books, special information forms, etc. Putting this information into computer programs can save space, reduce loss, make it easier to organize and locate the information, make it easier to enter the information, and thus save space and reduce costs. But the information is only as good as the quality of the data; a fact that was recognized long before computers were used.

Information is recorded and maintained because it is of use and importance. Sometimes it is just providing an historical record, but most often the information will be needed in the future to make new decisions and confirm prior decisions. For example, student transcripts record accomplishments, show that prerequisites for milestones like graduation have been completed, and are used in decisions such as college or technical school entrance. It has always been vital that the data recorded be accurate and reliable. It makes no difference whether the information is maintained on paper or using computer tools, the principles are identical.

Paper records achieve accuracy by the diligence of the people creating and maintaining the records. Teachers make sure that grades are recorded properly. Students, in their own interest, double check to make sure their course grades are consistent with the scores from their tests, projects, etc. Administrators look for unusual reports, incomplete information, and inconsistencies as a further check of data quality. Administrators and teachers alike do their best to assure that records are maintained, can be found when needed, and are protected from unauthorized use and from loss. This process is a well understood one that has stood the test of time.

## **Moving Historical Approaches Forward**

With so much data required by so many people at such a rapid pace, the use of Student Information System (SIS) and Data Information System (DIS) tools has become almost essential. Both schools and parents are becoming familiar with names such as Aeries (Eagle Software), PowerSchool (Pearson School Systems), Zangle (C Innovation, Inc.) and a host of lesser known products are in use or in the process of being installed in hundreds of schools and school districts. Some schools have even developed their own information systems using popular database tools as the base. These tools can help a school greatly, but both initial entry of the data and maintenance of the data quality is a burden on an already overworked faculty and staff. But how is that computer-based tools intended to make life easier are adding to the burden of tasks to be completed? Even more importantly, how can we work to get this under control so that these SIS/DIS tools provide what was intended, namely make things easier, faster, more accurate, and less burdensome?

Moving information from paper files to computer-based tools does not change its nature, importance, or value. All that has been done is to change the medium by which it is maintained. The difficulty is that changing the medium also substantially changes the process by which the information is recorded, verified for accuracy, maintained, and protected from loss or unauthorized use, and the means by which it can be searched. From the point of view of the value and usefulness of the data, it is not important whether the new processes are easier or harder, faster or slower, just that they are different. There are many good and strong reasons why using computerized tools are beneficial, so that point is not in dispute. Our objectives remain that the data be as accurate, complete, protected, available, and well maintained as possible.

While the methods for accomplishing these objectives must be different because the medium is different, much can be gleaned by comparing the new methods for use with computer tools with those effectively used in the past for paper records. The most important myth to overcome is that just because the data is maintained in the computer does not make it correct. Making the information correct takes the care of the teachers and administrators who understand and can verify the data.

## Data Correctness Basics

When records are maintained on paper the key identifying information is usually a name, student and/or teacher, with additional information used to distinguish in case of ambiguity, e.g. two students with the same name. SIS and DIS systems, which are actually specialized database applications, depend on numbers to uniquely identify the records. These numbers are usually specially assigned student, teacher, course, etc. identifiers. The technical details are not important, but it is essential that these identifying numbers be used properly and consistently. With paper records, putting the wrong name on a report card is disastrous. With electronic records, putting the wrong student ID on a report card record is disastrous.

Understanding the importance of these electronic identifiers is critical to maintaining correct electronic data. For example, each of the following course names will be interpreted correctly when read on a schedule by a human:

Algebra I  
Beginning Algebra I  
First Year Algebra I

In a computer system, each of these would have a unique course ID associated, and these textual descriptions would be only for human ease of reading. Since all the course ID's would be the same, the textual descriptions would not be important to the SIS or DIS system.

On the other hand, seemingly insignificant changes in a course ID put into an electronic record might have a quite different result. For example:

MA01  
MA1  
MA001

would likely be interpreted by a human reader as the same course ID. The computer system (SIS or DIS) would treat each of these as unique, and thus three different courses. So when the principal adds John Smith to course MA1 while the counselor adds Sarah Jones to course MA01, they are assigned to different courses. This situation gets compounded when the teacher comes to complain that John Smith came to the class but wasn't on the roster, and the counselor adds John Smith to the class without realizing that there is already a record assigning John Smith to MA1, instead of MA01. In addition, John Smith's record would show more courses than are permitted, which will cause other error messages to be generated by the SIS.

This example illustrates a number of important points relating to data maintaining correct data in an SIS or DIS system. These are:

1. Identification fields – These are fields in the computer system data that must have only certain selected values or they are not correct, e.g. Student ID, Course Number, Teacher ID, etc. These are the most critical fields with regard to data errors, as they are core to the operation of the SIS/DIS system. Some SIS/DIS systems will prevent a user from entering invalid values if you are using the User Interface, but not all. Often large

- quantities of data (such as creating a new system, starting a new school year, or entering the set of courses and schedules for a new semester) are entered using a spreadsheet, in which case the SIS/DIS system rarely checks for proper/correct values even in the identification fields. Thus it is critical for the data to be carefully reviewed and corrected before it is used to load information into the SIS/DIS system.
2. Fixed format fields – These are fields in the computer system data that have a special format, but the content is varied, e.g. dates, times, cost, and percentages. Like the identification fields, the User Interface often guides proper data entry, but bulk entry of the data through a spreadsheet or other mechanism may allow incorrectly formatted data to be entered. Errors in this data will probably not impact other parts of the data system, so the impact will be more localized. However, specialized reports may be impacted, e.g. the report of students who took the SAT-I in December of 2008 will include/exclude students due to incorrect dates entered in their records.
  3. Free-form text fields – These include items such as student name, teacher name, course title, course description, etc. Errors tend to be localized to records for a single student, but errors are quite important. Even though these fields are “explanatory information”, it is often used for searching or reporting. Errors in some of these fields, such as student or teacher names, could cause substantial confusion in the use of the information from the system, and could prevent records from being located.
  4. Numeric fields – These fields are designed to contain only numbers. These may be integers or real numbers (numbers with decimal values), and may be restricted in the maximum or minimum values allowed. These fields might be used for test grades and scores, street addresses or room numbers. Errors in these fields can be difficult to detect, but no less important. Finding and correcting these errors can take substantial time and effort.

Regardless of the type of data field involved, making sure that the data is correct is essential. The most important rule to keep in mind is consistency. Consistency is vital for identification fields like student ID's and course ID's, but lack of consistency in other types of fields may show up in many different uses of the SIS/DIS systems. Another aspect of consistency is to stay consistent with the design rules and assumptions of the SIS/DIS system. Here are a few examples of what can happen if the data is inconsistent:

1. The person doing scheduling created two courses for Algebra I, C100 meeting M-F first period, and C101 meeting M,W,F first period. There was also an independent studies, C110 meeting T,T created. The intent was for students assigned to C110 to be also assigned to C101, and all other 9<sup>th</sup> grade students to be assigned to C100. When the data was actually entered into the SIS system, many students assigned to C110 were also assigned to C100, instead of C101. This resulted in T,T conflicts for students incorrectly assigned, but the error was not recognized until student schedules were printed and distributed. Had a test run of the student schedules been done and checked first, the errors would have been caught before final schedules were printed for students.
2. In paper-copy systems, administrators were always concerned with clearing out old records to make space for new records in filing cabinets. Although retrieving the older records became more difficult, it was important to the efficiency of handling current student records. Since everything tended to be stored in folders labeled with the students name or other identifier, the entire folder would be moved. With electronic SIS/DIS systems, records are usually contained in multiple computer “folders” (actually multiple database tables). Clearing the information in one “folder” does not generally clear corresponding information in other tables. For example, clearing the student ID record would probably not also clear the corresponding folder of report cards, or the folder of standardized test results, for that student. At the same time, once the student ID record is removed, it may be difficult to obtain the identifying information for the student associated with the report-card folder. Understanding the nature of how to handle such deletions in an SIS/DIS system is critical to maintaining an accurate system. In addition,

- because the amount of electronic space required for these records tends to be small, it is usually neither required nor advised to delete old student records.
3. Two students, one in ninth grade and one in twelfth grade, attended the same school. When schedules were being created, the scheduler selected the same student ID when selecting classes for these students, without double checking the student names. The result was that the twelfth grader was assigned a double schedule, and the ninth grader had no classes. Of course, it could easily have ended up with the ninth grader taking twelfth grade courses, and vice versa for the twelfth grader. Double checking of that the student ID's were for the desired student would have prevented this error, but only if both the name and grade level of each student was verified.

A natural question to ask is why errors such as these are not caught as the data is entered. There are many different answers to that simple question, several of which were discussed above. The overriding answer is, from our experience, multi-factored:

- a. People usually have a difficult time spotting errors when they have just entered the data. The usual remedies to this are to have a second person check the data after entry, or to have some time elapse before the data is checked by the same person.
- b. Take advantage of any built-in data consistency checks of the SIS/DIS system. Sometimes these are associated only with the User Interface, so that if bulk data entry by imports from a spreadsheet or other similar method are used the SIS/DIS error checking is not invoked. However, there are often error message logs which in which some types of data errors and warnings are reported. It is important to examine and understand these errors and warnings, and not just ignore them because they are confusing.
- c. Only part of the "picture" shows up in a specific User Screen or data record in an SIS/DIS. This is because student identification information is maintained separately from grade reports, from standardized test scores, from attendance records, etc. In paper records more information tended to be on a single piece of paper or form, and thus there were more chances for this information to be cross-checked as the students records were updated with new information. One remedy to this is to have someone take the time to cross-check information periodically to make sure that it is correct. Another is to generate special reports with all the information on them that can be checked by one or more people for errors and consistency.
- d. There is still a tendency of people to assume data from a computer system is correct, more so than when the same data appears on written forms. This is a mindset that must be overcome.
- e. It is more difficult to correct the data in an SIS/DIS. Sometimes this is due to the complexity of the SIS/DIS system, but most often, it is due to the fact that the person who identifies the error has no authorization to change that data in the SIS/DIS system, and thus must locate someone with the proper authorization. In a paper record, anyone with a pen will tend to make the correction, even if they should get permission before doing the change.

While these and other answers explain why the data may contain errors, the fact remains that time and effort must be allocated to locating and correcting the errors in order for the SIS/DIS system to fully replace paper records and provide the full value for which it was purchased.

## Conclusion

Getting the most from SIS & DIS systems requires attention to the details of data correctness. The correctness of data in school records has always been important, but achieving correctness often requires more time and attention with SIS & DIS systems than it might have taken for corresponding paper-based record systems. The time and effort required to make and keep the data correct in SIS/DIS systems is usually considered "hidden" costs. They are hidden in the

sense that teachers and administrators are rarely guided to plan for them, thus these activities come as a surprise.

Comparing the process of creating and maintaining paper records, as has been done for many years, versus the electronic records of SIS/DIS systems can help to reveal why these costs should be expected. In particular, that SIS/DIS systems can only fully replace paper records and achieve the full benefits possible if time and money are allocated in the budget on an ongoing basis to locating and correcting errors in the system data.

*Gerald A. Wilson, Ph.D. is the CEO of OmniSoft, Inc., a company specializing in network security and risk management with special attention to K-12 education. Dr. Wilson has more than 25 years of experience in a broad variety of networking systems design and analysis. He currently does technical consulting, project management, and lecturing.*