

Spreadsheet or Database?

We have spent some time discussing spreadsheets and their use. We will be moving on to databases in the next section, but it might be useful to briefly discuss when to use a spreadsheet and when to use a database. Spreadsheets are valuable for answering what-if questions—for instance, if you're looking at a spreadsheet for your budget, and you ask the question "what-if" I move to a different apartment and my rent is a \$100 more than it is now? To answer the question, you plug in the new value for your rent and the entire spreadsheet recalculates itself to show the impact on your budget. If that amount is too much, you can plug in different lower amounts and retest until you find one that is comfortable for your budget. These powerful 'what-if' questions are possible because formulas in the cells of spreadsheets can reference formulas in other cells.

Databases have numerous technical advantages over spreadsheets when you need to organize complex data with more than one user accessing the data. Spreadsheets in Excel and tables in Word are fine for managing simple lists like our example of tracking grades in one class. It gets more complicated when you need to track students in all of their classes. One student takes many classes and one class is taken by many students. Organizing this data gets very convoluted when we try to represent this data logically in a list. Lists work poorly when more than one department needs to have access to the students' information: grades, classes, addresses, etc.

One of the biggest advantages of databases over spreadsheets is minimizing data redundancy. Using spreadsheets, if the registration office needs the students' names and addresses and the financial aid office also needs this data, the same data would be kept in two different spreadsheets with each different spreadsheet reflecting the additional information important to each department. Think of the problems that arise when you need to change a student's address. How do we ensure the new address gets changed in both places? And if the change doesn't occur in both places, whose spreadsheet has the correct address? Even if the change does occur in both places, making the change doubles the work, doubles the disk space, and doubles the chance that human error will occur. This situation results in data integrity problems.

In a database, the data is kept in one place and stored only once. It is shared by all and when it is changed, everyone sees the same changes and it only has to be changed once. On the down side, databases cost money. Database software is complex, so not only do you have the cost of purchasing the database software, you also have the hardware costs, the cost of hiring people to manage it, programmers to access it, and the cost of training users. Even so, in a complex data environment, I think you will see how much these advantages far outweigh the disadvantages, which is why companies could not survive in this day and age without implementing database management software solutions.