A long-standing problem with end-user computing is that tools such as simple databases and spreadsheets is that they are so easy to use that people with no background or training in computing can apply them to production problems. “Production” means applications on which an organization depends for mission-critical decisions or functions.

Gene Wirchenko, writing in RISKS 24.16 <http://tinyurl.com/khcwd> reported on a site that lists significant errors in spreadsheets: <http://www.eusprig.org/stories.htm>. The site is managed by the European Spreadsheet Risks Interest Group (EuSpRIG); their description reads, "These stories illustrate common problems that occur with the uncontrolled use of spreadsheets. We say how we think the problem might have been avoided. An obvious form of risk avoidance is simply to check your work before sending it out. For important spreadsheets, a second pair of eyes ('peer review') is even better. Where stakes are high, a thorough test and audit is a further defence." The group runs an annual conference that concentrates on quality assurance for spreadsheets.

One of the examples that comes to mind about spreadsheet errors occurred in March 2000. Mark Lutton reported in RISKS 20.84 <http://tinyurl.com/j6kls> on a week-long kerfuffle at MIT, when the grades of 22 students in a cell biology class were randomly altered. Initial suspicions focused on hacking, and the teacher, Harvey Lodish, told his class on 2 March 2000 that he had uncovered a cheating scandal. On March 10, the Boston Globe reported that in fact a teaching assistant had sorted the student-name column but not all the other ones, thus failing to carry all the data through the sort. Lutton suggested, "It seems to me that bound paper ledger books would be a much better tool for keeping grade records, at least for this teacher and his assistants." I commented in my INFOSEC Year in Review Database entry, “Some other ideas: (1) Enable the audit-trail feature (can create large files but does record all changes); (2) keep daily backups with version numbers so that a good version of the data can be located and used quickly.” <http://www.mekabay.com/iyir>

In a later issue of the RISKS Forum Digest (20.86), correspondents Tony Lima and John Pearson both pointed out that the fundamental problem was that the teaching team was using a spreadsheet to do a database’s job. Spreadsheets have no mechanism for ensuring record integrity, whereas even simple databases can protect against the kind of scrambling that occurred in this example.

Unfortunately, beginners rarely think about quality assurance in a systematic way. When I teach beginners how to use office products in the first-year non-majors’ computing course at Norwich University, I emphasize the importance of documenting and testing spreadsheets before relying on them. Professor Raymond R. Panko of the University of Hawai’i has studied the frequency of errors in spreadsheets in real-world applications. His paper, “What We Know About Spreadsheet Errors” <http://tinyurl.com/6748a> provides a meta-analysis of 13 studies of spreadsheet errors from 1987 through 2004. Prof Panko’s abstract is as follows:
“Although spreadsheet programs are used for small ‘scratchpad’ applications, they are also used to develop many large applications. In recent years, we have learned a good deal about the errors that people make when they develop spreadsheets. In general, errors seem to occur in a few percent of all cells, meaning that for large spreadsheets, the issue is how many errors there are, not whether an error exists. These error rates, although troubling, are in line with those in programming and other human cognitive domains. In programming, we have learned to follow strict development disciplines to eliminate most errors. Surveys of spreadsheet developers indicate that spreadsheet creation, in contrast, is informal, and few organizations have comprehensive policies for spreadsheet development. Although prescriptive articles have focused on such disciplines as modularization and having assumptions sections, these may be far less important than other innovations, especially cell-by-cell code inspection after the development phase.”

It makes sense to identify production spreadsheets in your organization and to review the quality assurance processes (if any) to ensure that errors don’t creep into operational decisions.

* * *


M. E. Kabay, PhD, CISSP-ISSMP is Program Director of the Master of Science in Information Assurance <http://www.msia.norwich.edu> at Norwich University in Northfield, VT. Mich can be reached by e-mail at <mailto:mkabay@norwich.edu>; Web site at <http://www.mekabay.com/index.htm>.

Copyright © 2006 M. E. Kabay. All rights reserved.

Permission is hereby granted to Network World to distribute this article at will, to post it without limit on any Web site, and to republish it in any way they see fit.