Electronic spreadsheet models and modeling (ESMs) have constituted a very important area of organizational and end-user computing for well over a decade. Given the importance of ESM, a substantial amount of research has already been aimed at various related issues and a special-interest group SPRIG (Spreadsheet Productivity Research Interest Group, http://sprig.section.informs.org/) exists in the INFORMS organization. In this concept paper, I briefly review the existing literature. Then I concentrate in more detail on some new needs related to data warehousing and data mining that arise due to recent developments in policy capturing and related estimation methods.

Ragsdale’s (2001) book discusses applications to virtually every functional area of business and to the standard topics of management science. More specialized application research has appeared, such as budgeting (Troutt, Yew, & Zhang, 2001), simulation (Mumford & Schultz, 1991), simulation optimization (Nersesian, Troutt, & Weinroth, 2001; Troutt, Nersesian, & Weinroth, 2004), and production planning (Troutt, Tadisina, & Clinton, 1991). Solver engines are now readily available for linear and nonlinear programming, as well as genetic algorithms, which enable such applications as simulation optimization in particular. One can expect to see further growth in these capabilities to include other metaheuristic methods such as simulated annealing and Tabu search. The work in Troutt et al. (1991, 2001) seeks to further broaden the capabilities of ESM to a more macroview of general planning applications of ESM. Work on facilitating this view would be of interest and may be related to effective graphical user interfaces.

From the IT, organizational, and end-user issues perspectives, work has concentrated on quality and design issues, especially the detection and prevention of errors (Harvard Business Review, 1996; Panko, 1988). The paper by Cragg and King (1992) provides a survey, framework, and call for research in ESM. Despite having been written over a decade ago, it continues to provide an excellent guide and should be considered essential reading for ESM researchers. Though not intended to be a complete survey, the following articles are a sampling of some more-recent research. Chan (2000), Galletta and Hartzel (1997), Janvrin and Morrison (2000), Harrison, Harrison, Melrose, Wilson, and Vance (2002), Panko (1998), and Panko and Sprague (1998) continue work on understanding and preventing ESM errors and auditing. Chan and Storey (1996) and Floyd and Walls (1995) consider usage,
satisfaction, and management issues for the use of ESM in organizations. Classe (1992), studying accountants, reports on searches for ways to optimize how organizations manage the use of spreadsheet models. The Web site of Ray Panko (http://panko.cba.hawaii.edu/ssr/) is an excellent source of current research also. The new issues discussed below relate especially to the optimal organizational use of ESM.

Recent developments have introduced the need to consider the warehousing and potential data mining of organizational ESM files. Gray and Watson (1998) have observed that spreadsheet analyses are seldom warehoused. I will propose some reasons for this below. First, I discuss how this need arises from some work related to policy capturing and more-recent enhancements of that approach that can use past planning spreadsheet models as data.

Policy capturing refers to the modeling of past actual decisions with the aim of understanding and predicting outcomes in new similar situations. A sampling of articles in that area includes Boon and Sulsky (1997), Crosby and LeMay (1998), Deshpande and Schoderbek (1993), Dulebohn and Martocchio (1998), Graves and Karren (1992), Law and Wong (1998), Martocchio and Judge (1994), and York (1992). Generally, these policy-capturing studies have used regression analyses or its variants such as logistic regression. Dependent variables can be accept-reject, hire or do not hire, yes-no, or classifications with more categories. However, continuous dependent variables can also be used, such as pay-raise amounts or percents. Independent variables can be essentially any categorical or continuous variables hypothesized to influence or explain the decisions in question. An earlier related work, using linear programming instead of regression, was discussed in Troutt (1985).

A new direction, similar in spirit to policy capturing, was begun in Troutt (1995). This approach enables the fitting of optimization models to past decisions. Hence, it is also related to frontier estimation techniques such as Troutt, Gribbin, Shanker, and Zhang (2003), Troutt, Hu, Shanker, and Acar (2003), and Troutt et al. (1997). Applications have been discussed in Troutt, Gribbin, Shanker, and Zhang (2000), Troutt, Pang, and Hou (2004), Troutt, Tadisina, Sohn, and Brandyberry (2003), and Troutt et al. (1997). The approach can estimate costs and other parameters thought to influence plans based on past planning data, most often now contained in finalized spreadsheet models. Recently, application of the idea to estimating production-planning costs has been discussed in Troutt, Pang et al. (2004). However, the general approach might be applied to historical spreadsheet models for any situation for which an optimization model can be proposed. This development underscores the need for warehousing of ESMs and essentially amounts to the data mining of such models. In addition, such models can provide a window to the influences and data that planners actually used in the past. In fact, it can be argued that such models actually constitute knowledge, especially tacit knowledge, of subject-area experts that should become part of the knowledge base of the organization. As noted above, Gray and Watson (1998) observed that ESMs were seldom warehoused as of the time of their writing. So far, reported work in this area appears to be scant to nonexistent. Having looked at some advantages, there are a number of barriers to be considered.

What are the barriers related to warehousing of ESMs? First is the question of human factors and who owns a finalized ESM used in or for organizational decision
making. If an end user makes a recommendation based on ESM, does that ESM become property of the organization, or is it private? In the production-planning situation, plans are typically discussed among several analysts before finalization. Presumably, the executive in charge is entitled to keep and file such ESMs. But there may be a risk to both the executive and end users responsible for the ESM in case unfortunate decisions can be linked to errors in the ESM. If a strong policy requirement is made for this purpose, then use of appropriate ESMs may decline. Voluntary contribution of such ESMs to the knowledge base of the organization may lead to sporadic gaps in the data and reduce their usefulness. Second, on the technical side, the extended use of ESMs in a data warehouse reemphasizes considerably a number of concerns and needs already known. Clearly, standardization and documentation would take on new significance if ESMs were to be warehoused. Along with auditing, these steps would play a role similar to data cleansing in warehousing.

To summarize, the framework of research proposed by Cragg and King (1992) appears to leave a great deal yet to do in ESM. New methodology for mining of ESM files makes the possibility of their data warehousing attractive. Significant challenges exist, particularly in the human-factors issues. Technical issues should be expected to be less severe, but should make error detection and documentation even more important than before.

References


zations: Techniques and applications (pp. 21-31). Hershey, PA: Idea Group Publishing Co.