An Acknowledgment

As editor-in-chief of the Data Engineering Bulletin, I have been helped by the hard work of others. Mark Tuttle of Digital's Cambridge Research Lab defined the Latex style files that provide the clean and readable format of each issue. System support staff, at both Digital and Microsoft, have helped enable delivery of the Bulletin over the web. Both Digital and Microsoft have generously provided the infrastructure to support Bulletin distribution, as well as providing me with the editorial time.

I'd like to add to that list of acknowledgments.

- The style files used to generate the Bulletin work flawlessly only with a past version of Latex. I carefully maintain this version, but it is not readily available to associate editors or authors. The style file problem involves including an IEEE copyright notice, and arose after the style files were designed. Due to the efforts of **Do Hong Hai**, an author in this issue, this problem has been solved. Do Hong Hai modified one of the style files to permit the latest Latex system to successfully format the Bulletin. The result both simplifies my life (only the latest system is required) and makes it possible for associate editors and authors to easily check the formatting of the issue and the individual papers. This is the first issue produced with the new style files.

- The TCDE membership list is very large. In September, we experienced difficulty in sending email messages to very large distribution lists. As a result, the distribution list has been shifted to a web-based distribution list management tool at
`http://list.research.microsoft.com/scripts/lyris.pl?enter=debull`.
I would like to thank **Jeff Chirico** and **Wyman Chong** for providing the system support. The new email distribution tool makes it easier to update the distribution list, and TCDE members can remove themselves from the list if they so desire. As before, if your email address changes, you should contact me, mentioning both old and new email addresses.

The Current Issue

Heterogenous databases, with their syntactic and semantic differences, are an information legacy that will probably be with us always. Despite our efforts to define standards for information interchange, e.g., the current efforts with XML, bringing information together so that it can be sensibly queried remains a difficult and expensive task, usually requiring human assistence. This is the "data cleaning" problem. Data warehouses usually face this problem every day, in their role as repository for information derived from multiple sources within and across enterprises.

Data cleaning has not been a very active area of research. Perhaps the problem was too messy to deal with, too hard to attack with any generality. However, while research has paid little attention to data cleaning, industry has been forced to deal with this on a regular basis. This issue of the Bulletin addresses this long-standing and important problem. Sunita Sarawagi has herself worked in this important area, and has succeeded in soliciting papers from a mix of academia and industry, with papers from Europe and Asia, as well as the US. Thus the issue provides a very broad perspective on this important subject. Sunita has done a fine job as editor of the issue, and I want to thank her for her hard work.